Information Processing and Management 54 (2018) 758–790

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

A survey towards an integration of big data analytics to big insights for value-creation

Mandeep Kaur Saggi*, Sushma Jain

Department of Computer Science, Thapar, University Patiala, India

A R T I C L E I N F O

Keywords:
Big data
Data analytics
Machine learning
Big data visualization
Decision-making
Smart agriculture
Smart city application
Value-creation
Value-discover
Value-realization

A B S T R A C T

Big Data Analytics (BDA) is increasingly becoming a trending practice that generates an enormous amount of data and provides a new opportunity that is helpful in relevant decision-making. The developments in Big Data Analytics provide a new paradigm and solutions for big data sources, storage, and advanced analytics. The BDA provide a nuanced view of big data development, and insights on how it can truly create value for firm and customer. This article presents a comprehensive, well-informed examination, and realistic analysis of deploying big data analytics successfully in companies. It provides an overview of the architecture of BDA including six components, namely: (i) data generation, (ii) data acquisition, (iii) data storage, (iv) advanced data analytics, (v) data visualization, and (vi) decision-making for value-creation. In this paper, seven V's characteristics of BDA namely Volume, Velocity, Variety, Valence, Veracity, Variability, and Value are explored. The various big data analytics tools, techniques and technologies have been described. Furthermore, it presents a methodical analysis for the usage of Big Data Analytics in various applications such as agriculture, healthcare, cyber security, and smart city. This paper also highlights the previous research, challenges, current status, and future directions of big data analytics for various application platforms. This overview highlights three issues, namely (i) concepts, characteristics and processing paradigms of Big Data Analytics; (ii) the state-of-the-art framework for decision-making in BDA for companies to insight value-creation; and (iii) the current challenges of Big Data Analytics as well as possible future directions.

1. Introduction and motivation

The notion of Big Data Analytics (BDA) is driven by underpinning new waves of innovation, analytic services with intelligence and stirring advances in technology over the last few decades. The emergence applications of BDA have prompted the attention of many academic researchers, industry practitioners, and government organizations. It is a technology-driven ecosystem, where better decision-making will help many organizations to extract knowledge from data in an interpretable and appropriate form.

Strawn (2012), described Big Data as “fourth paradigm of science”, whereas (Hagstrom, 2012) defined it as “new paradigm of knowledge assets”, or “the next frontier for innovation, competition, and productivity” (Manyika et al., 2011). Gantz and Reinsel, (2011) defined Big Data as “a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery, and analysis”. It was described an integrated approach to organize, process, analyze the six characteristics (namely volume, variety, velocity, veracity, valence, and value). BDA is used to generate action for delivering the insights, value, measuring performance, and establishing competitive advantages (Wamba,

* Corresponding author.
E-mail addresses: mandeepsaggi90@gmail.com (M.K. Saggi), sjain@thapar.edu (S. Jain).

https://doi.org/10.1016/j.ipm.2018.01.010
Received 20 December 2016; Received in revised form 26 January 2018; Accepted 30 January 2018
Available online 09 February 2018
0306-4573/

Akter, Edwards, Chopin, & Gnanzou, 2015). The paper by (De Mauro, Greco, & Grimaldi, 2016) defined that "Big Data is the information asset characterized by such a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value" .

The BDA, as a scientific topic of investigation, provides some significant and insightful readings that are discovered by various researchers. However, it is still needed to carryout the systematic review of innovative analytical methods, techniques, and tools for making insightful decisions in various domains. Indeed, it became a key component of decision-making processes in business (Hagel, 2015).

The big data and advanced data analytics techniques can be used for the development of analytical and computational models (Iqbal, Doctor, More, Mahmud, & Yousuf, 2017). There are still several research interest how to develop the infrastructure, apply various data mining and machine learning algorithms in different domains. The BDA is concerned with modern statistical and machine learning techniques to analyze huge amount of data (Suthaharan, 2014). The researchers suggested that Big Data Analytics and deep learning have the potential to provide new generation applications based on modeling and simulation (Chen & Lin, 2014; Tolk, 2015).

The traditional tools are not able to address the issues of scalability, adaptability, and usability, whereas such issues are critical to its success as they influence how big data is developed, managed and analyzed. The BDA is categorized by the requirement of advanced data acquisition, data storage, data management, data analysis, and visualization. To turn BDA into big insights for value-creation, there are great challenges in terms of data, process, analytical modeling and management for different applications. It should not be considered as synonymous with data collected through the internet as data can be originated from sources such as commercial transactions taking place in supermarkets, bank etc. Big Data can also be originated from sensors (satellite and GPS tracking data from mobile phones) and administrative data (education records, medical records, and tax records) (Eagle, Pentland, & Lazer, 2009).

The BDA helps in acquiring a deep understanding and useful insights of various sectors such as: agriculture, healthcare, cyber-physical system, smart cities and social media analytics etc. The enormous amount of information is needed to analyze it in an iterative way and time sensitive manner (Jukic, Sharma, Nestorov, & Jukic, 2015). By the use of advanced BDA tools such as NoSQL, BigQuery, Map Reduce, Hadoop, Flume, Mahout, Spark, WibiData, and Skytree, it provided an insight in desirable form to enhance the ability and decision-making process in various sectors such as business intelligence analytics (Chen, Chiang, & Storey, 2012), healthcare analytics (Archenaa & Anita, 2015), smart agriculture or farming analytics (Majumdar, Naraseeyappa, & Ankalaki, 2017; Wolfert, Ge, Verdouw, & Bogaardt, 2017), social media analytics (Vatrapu, Mukkamala, Hussain, & Flesch, 2016), smart cities (Khan, Anjum, Soomro, & Tahir, 2015), intelligent transport management (Fiosina, Fiosins, & Müller, 2013), financial and accounting (Sledgianowski, Gomaa, & Tan, 2017), financial risk management (Cerchiello & Giudici, 2016), energy management (Tu, He, Shuai, & Jiang, 2017), and future predictions (Waller & Fawcett, 2013).

The BDA is data-driven decision framework. This article is directed to comprehensively study the BDA to solve the challenges, gain insight, and to make informed decisions by using various data analytics approaches. This paper summarizes an extensive and systematic methodological review on various tools and technologies of BDA and also reports the research gaps for further investigation. In more detail, our review article addressed following research questions:

- RQ1: What are the most important seven characteristics of Big Data Analytics?
- RQ2: How to design BDA-DM framework?
- RQ3: What are the main tools, techniques, and technologies of Big Data Analytics?
- RQ4: What are the main application areas of Big Data Analytics?
- RQ5: What is the relation between value-creation and Big Data Analytics?
- RQ6: Which are the specific aspects of the data management, data transformation and utilization drive value for companies?
- RQ7: Can the value of data be monetized, tracked and considered for financial accounting?
- RQ8: What are the different challenges of each component of the BDA framework?

This article attempts to answer the above research questions (RQs). RQs will guide, centre our research work and clearly focus on specific topics to indicate our distinctive perception. However, this work leads to a new advancement for the conceptual framework of BDA.

The contributions of this research article are as follows:

- Categorize the current approaches and general requirements for various components of BDA architecture by demonstrating the open state-of-the-art frameworks and challenges.
- Summarize various existing tools, methods, and technologies in advanced BDA.
- Provide the summary of the key technology for value-creation applications, financial companies of BDA.
- Present the, future research directions relating BDA in new emerging technologies.

This paper is structured into eight sections. The Section 2 describes the relevant research methodology and summarizes the review studies. The Section 3 presents an ecosystem of Big Data Analytics and Decision-Making Framework (BDA-DMF). The Section 4 presents a big data management phase of the framework. The Section 5 presents Big Data Analytics techniques, technologies, tools, and its applications phase. In this section, we present a concise statement of different steps of data analytics framework. A brief review of different areas of application such as Agriculture, Healthcare, Cyber security and Smart City is also presented. The Section 6

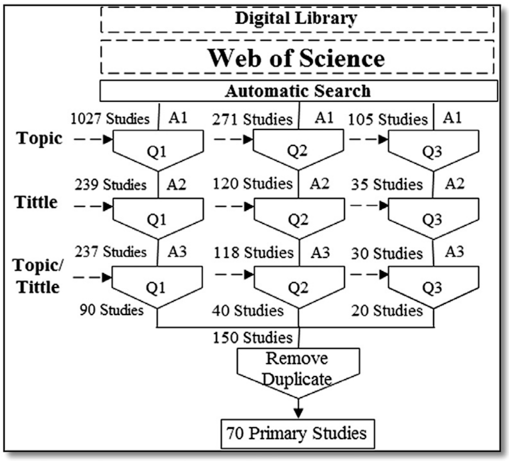

Fig. 1. Primary studies selection process.

covers a visualization phase of BDA framework. The Section 7 describes the value-creation need, benefits, and framework of BDA for financial and accounting companies. Section 8 describes the conclusion and future research directions in the area of Big Data Analytics.

2. A glimpse of big data research method: systematic mapping process

In this paper, the articles from Web of Science digital database are considered. To ensure thoroughness and consistency in our review, the guidelines presented by (Brereton, Kitchenham, Budgen, Turner, & Khalil, 2007) are followed and used the digital library databases (Springer, Science Direct, Google scholar, IEEE Xplore, ACM library).

The Web of Science database index contains several types of documents namely articles, reviews, proceeding papers, meeting abstract, editorial material, book review, and book chapters. Significant research publications have obtained from Web of Science on BDA, Big Data Analytics-Management (BDA-M) and Big Data Analytics-Machine Learning (BDA-ML) for a considerably large period of 20 years from (2000–2017).

2.1. Data inclusion and exclusion process

In this paper, 70 primary studies have been selected and analyzed through a process that formulates criteria for inclusion and exclusion articles for review. The selection process of primary studies is summarized in Fig. 1.

- Firstly, the selection process of primary studies is conducted based on different queries. The queries are executed topic-wise, title-wise and combination of both. The last filter is based on the abstract and full reading of the paper. If it is not relevant to the study, it is automatically excluded. On reading the abstract of results, the inclusion/exclusion criterion was applied at the end.
- Secondly, the papers reporting on theoretical, empirical and both qualitative or quantitative case studies have been considered.

Table 1
Examples of query extracting process.

Query Number	Topic/Title	Keywords
Q1	A1	TS = (“Big Data” AND “Big Data Analytics”)
Q2	A1	TS = (“Big Data Analytics” AND “Big Data” AND “Management “)
Q3	A1	TS = (“Big Data Analytics” AND “Big Data” AND “Machine Learning”)
Q1	A2	TI = (“Big Data” AND “Big Data Analytics”)
Q2	A2	TI = (“Management” AND “Big Data”) OR (“Management” AND “Big Data Analytics”)
Q3	A2	TI = (“Machine Learning” AND “Big Data”) OR (“Machine Learning” AND “Big Data Analytics”)
Q1	A3	TS = (“Big Data Analytics” AND “Big Data”) AND TI = (“Big Data Analytics” AND “Big Data”)
Q2	A3	(TS = (“Management” AND “Big Data ”) OR (“Management” AND “Big Data Analytics”)) AND (TI = (“Management ” AND “Big Data”) OR (“Management” AND “Big Data Analytics”))
Q3	A3	(TS = (“Machine Learning ” AND “Big Data Analytics”) OR (“Machine Learning ” AND “Big Data ”)) AND (TI = (“Machine Learning ” AND “Big Data Analytics”) OR (“Machine Learning” AND “Big Data”))

Table 2
Top 18 research areas of the existing Big Data Analytics contributions.

Research areas	Topic			Title			Topic and title		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Computer Science	483	127	63	125	50	13	124	49	13
Engineering	196	63	20	56	25	6	55	25	5
Business Economics	128	55	2	29	18	1	29	18	2
Telecommunications	78	18	7	32	11	3	32	11	2
Information Science Library Science	69	18	2	15	4	1	14	3	1
Operations Research Management Science	61	23	1	11	6	1	11	6	1
Science Technology Other Topics	35	9	5	9	5	1	9	5	1
Health Care Sciences Services	32	8	4	9	1	1	9	1	1
Automation Control Systems	10	6	2	2	3	1	2	3	1
Medical Informatics	24	6	5	2	1	1	2	1	1
Mathematics	18	1	1	3	1	2	3	1	1
Environmental Sciences Ecology	15	6	6	1	6	2	1	6	2
Neurosciences Neurology	13	1	1	3	1	1	3	1	1
Remote Sensing	11	1	1	3	2	1	3	2	1
Mathematical Computational Biology	10	2	2	1	1	1	1	1	1
Communication	8	1	1	2	1	1	2	1	1
Biotechnology Applied Microbiology	8	1	1	2	1	1	2	1	1
Agriculture	6	2	2	2	1	1	2	1	1

2.2. Study selection process and data analysis

The review article selection process is based on “Query Extracting Process”. The queries are numbered as Q1, Q2, and Q3 by using a combination of various keywords. Table 1 shows some examples of the executed queries.

In the first scenario, 1027 research articles are selected on the basis of their topic-wise execution of “Q1”, which includes 867 articles, 84 editorial material, 66 reviews, 26 proceeding papers, 3 book reviews, 7 book chapters, and 1 meeting abstract. On execution “Q2”, a total of 272 research articles are listed that contains 222 as articles, 30 reviews, 19 editorial material, 1 letter, 2 book chapters, and 7 proceeding papers. Furthermore, 105 research articles obtained on execution of “Q3”, among which, there are 87 articles, 11 reviews, 4 editorial material, 3 meeting, and 3 proceeding papers.

In the second scenario, the aforementioned queries are searched on a title-wise. As a result, 239 research articles are selected on execution of “Q1”, which contains 159 articles, 54 editorial material, 13 reviews, 2 proceeding papers, 3 book reviews, 1 book chapter, and 10 meeting abstract. Further, 120 research articles are listed on the execution of “Q2” that consists of 66 articles, 6 reviews, 43 editorial material, 1 letter, 4 meeting abstract, and 1 proceeding papers. The 35 research articles are listed on executing “Q3”, which includes 18 articles, 6 reviews, 6 editorial, 3 meeting, 1 correction and 1 proceeding papers. Finally, the topic-wise and title-wise queries are combined and executed simultaneously which results into 90, 40 and 20 papers on the execution of Q1, Q2, and Q3 respectively. After the manual cleaning of the data, 70 primary study papers were obtained.

Table 3
Top 16 Journals publishing articles of Big Data Analytics.

Journal	Topic			Title			Topic and title		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Big data	32	6	13	3	1	1	8	1	1
IEEE Access	18	2	3	8	1	1	3	2	1
IBM Journal of Research And Development	17	5	1	3	1	1	3	2	1
PLOS one	11	1	1	3	1	1	2	1	1
Decision Support Systems	11	6	1	2	1	1	4	2	1
Big Data Research	11	4	1	4	2	2	4	2	1
IEEE Network	9	2	1	4	1	1	5	1	1
Computer	9	1	1	5	1	1	1	1	1
Information System	4	1	1	1	1	1	1	1	1
Future Generation Computer Systems The International Journal of Grid Computing And Science	5	2	1	1	1	1	3	1	1
Communications of The Acm	5	1	1	3	1	1	2	1	1
Cluster Computing The Journal of Networks Software Tools And Applications	5	1	1	1	1	1	1	1	1
Expert Systems With Applications	4	1	1	1	1	1	1	1	1
Knowledge And Information Systems	3	1	2	1	1	1	1	1	1
IEEE Communication & Survey Tutorial	1	1	1	1	1	1	1	1	1

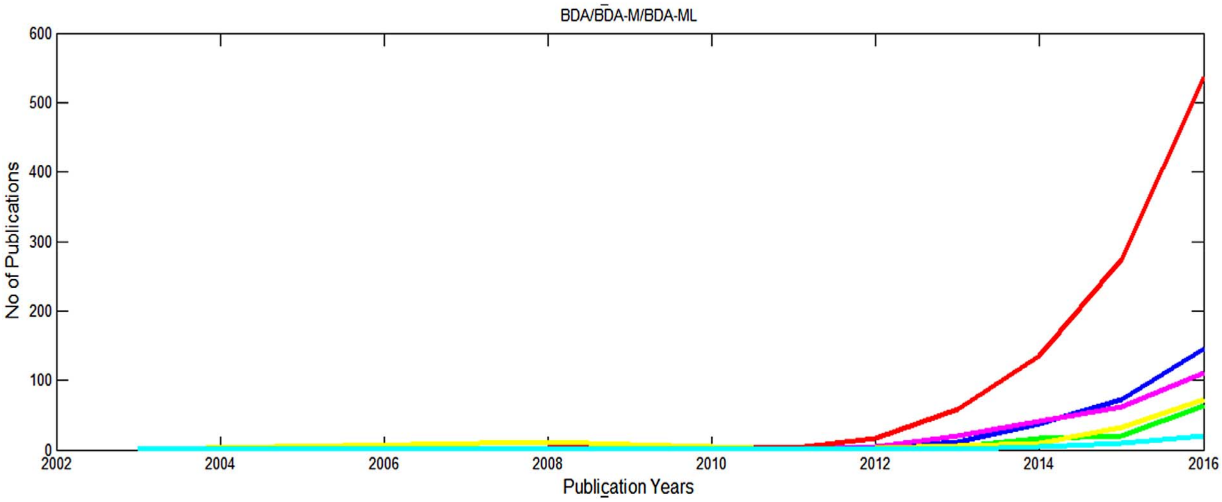

Fig. 2. Distribution of articles by year of publications in WoS.

2.3. Result related work

The summary is presented in Table 2, Table 3 and Fig. 2 and Fig. 3. The Table 2 indicates the number of publications by top 18 research area, the Table 3 indicates the list of the number of publications by top 16 journals, the Fig. 25 indicates the number of publications per year to justify trends of BDA (2000- 2016), and the Fig. 3 indicates the percentage of publications in the area of BDA. Further, it represents the topic-wise, title-wise, and both in topic-wise and title-wise percentage of publications in the area of BDA, BDA-M, BDA-ML.

The Fig. 2 shows the no of publications of BDA (537 and 110), BDA-Management (146 and 73) and BDA-Machine Learning (63 and 20) on the basis of topic-wise and title-wise queries. The topic-wise (TS) queries Q1, Q2, Q3 are represented with Red, Blue, and Green coloured lines respectively, while the title-wise (TI) queries Q1, Q2, Q3 are represented as Magenta, Yellow and Cyan coloured lines respectively.

2.4. Past, present, and future of big data analytics

In 1997, the terminology of "big data" was firstly described by (Cox & Ellsworth, 1997). The visualization provides an interesting challenge for computer systems as the: - data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. The popular press McKinsey and Company leveraged resources to document a 5–6% increase in global productivity from data-driven analytics, over the non-big data-friendly company. International Data Corporation (IDC) found that the created and copied data volume in the world was 1.8 zettabytes (ZB). IBM indicates that every day 2.5 EB of data is created. CISCO predicted that, by 2020, 50 billion devices will be connected to networks and to the Internet. Fig. 4 shows the timeline of the Big Data processing paradigms and technologies (Nino & Illarramendi, 2015; Buyya, Calheiros, & Dastjerdi, 2016)

The Big data resources present a great opportunity for digital business models, and can be seen with Google, eBay, Amazon,

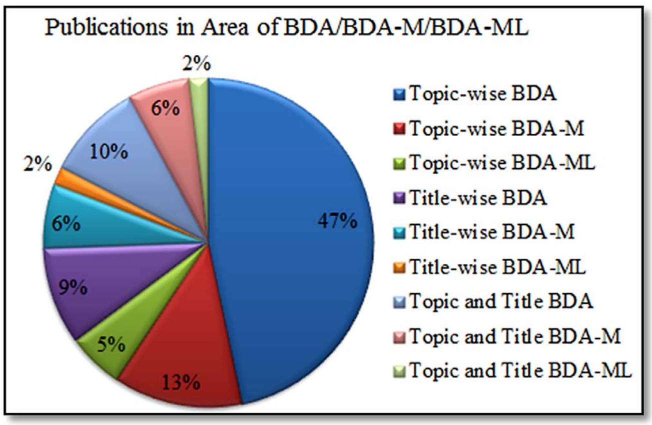

Fig. 3. Publications in the area of big data analytics, management, and machine learning.

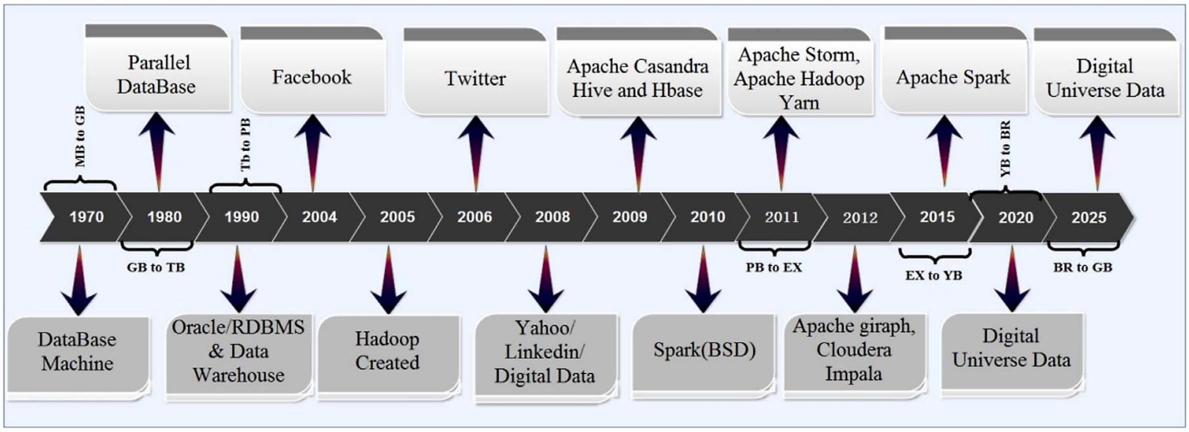

Fig. 4. The timeline of the big data processing and technologies.

Facebook and Netflix, Borders and many other businesses (Mithas & Lucas, 2010). In 1999, Apache software foundation (ASF) was established. For Big Data, batch processing was introduced in 2003 and 2004, Google popularized its papers on Google File System and Map Reduce. The first generation of Big Data was initiated in 2006 when Hadoop was born. Similarly, Apache Pig is originally developed and endorsed by Yahoo, Facebook began the development of an open-source tool for identical desire, Apache Hive. Yahoo! introduced Pig in 2008 and Facebook started Hive in 2009 (Casado & Younas, 2015).

The Second generation presented the Apache storm for real-time processing. It was started by Nathan Marz and released as open source by Twitter. Other companies like Cloudera or Linkedin presented interesting technologies such as Flume and Kafka. These open source developments have defined an ecosystem of Big data tools around Apache Hadoop, together with other components such as Apache Spark, Mahout (machine learning), Sqoop (data transferring between Hadoop and other systems), Oozie (job scheduling and monitoring on Hadoop), Zookeeper (distributed process configuration and coordination), and Apache Giraph (to process data stores as graphs) in 2014.

Since 1997, many characteristics added to Big Data. The first 3-V volume, variety, and velocity characteristics have been familiarized by (Gartner, 2011), the fourth V, Veracity has been included by Dwaine Snow in his blog named “Dwaine Snow's Thoughts on Databases and Data Management” in 2012. The first 3Vs: volume, velocity, variety (Chen & Zhang, 2014), 4Vs: volume, velocity, variety, and veracity (Abbasi, Sarker, & Chiang, 2016; Zikopoulos & Eaton, 2011) are described. Both variety and velocity are essentially working beside the veracity of the data. These V's decrease the capacity to cleanse the data before analyzing it and making useful insights. The 5V's are volume, velocity, variety, veracity, and value (Oracle, 2012), the fifth V introduced by (Gamble & Goble, 2011) refers to worthwhile and valuable data for business. The 7V's: volume, velocity, variety, veracity, value, variability, and visualization (Seddon & Currie, 2017). Variability and complexity are two other facts specifically for analytical areas.

***RQ1:** What are the most important seven characteristics of Big Data Analytics?*

Some of the technical challenges have been associated to different “V” characteristics, in particular “Volume” (support of very high data volumes), “Velocity” (fast analysis of data streams), “Variety” (support for diverse kinds of data), “Veracity” (support for high data quality), “Value” (the value of the insights and benefits), “Variability” (support for constantly changing), and “Valence” (support of connectivity in data).

The seven characteristics of BDA include some exploration of different steps and processes of data analytics. These seven aspects represent different difficulties in analyzing big data. Our major aim is to provide a comprehensive picture of each characteristic and also describes their challenges. These seven characteristics of BDA are shown in Table 4 and further explained as follows (Sivarajah, Kamal, Irani, & Weerakkody, 2017):

Currently, Big Data Analytics has become a trendy practice in business intelligence that consists massive amount of dataset and advanced analytic techniques. Villars, Olofson, and Eastwood (2011) stated that business and organizations can “extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and analysis”. Kambatla, Kollias, Kumar, and Grama (2014) presented a literature survey on Big Data Analytics. Assuncao, Calheiros, Bianchi, Netto, and Buyya (2015) stated that cloud computing plays a key role for Big Data because it can act as a business model to follow popular terms e.g. Analytics as a Service (AaaS) or Big Data as a Service (BDaaS). Zhang and Xiang (2015) discussed that BD integration, data quality issues, privacy and analytics can be used for effective business decision.

The paper by (Chen, Kazman, & Haziyev, 2016) introduced an architecture-centric approach, called Architecture-centric Agile Big data Analytics (AABA). Its purpose is to address technical and organizational challenges in big data system development and agile delivery of big data analytics for web-based systems. Fogelman-Soulié and Lu (2016) presented an application of Big Data Analytics in business (e.g. credit-card fraud detection). The framework developed in this study showed that how companies can store their Big Data in a data lake if they want to implement many Big Data projects.

The paper by (Ahmed et al., 2017) explored the recent advances and key requirements for managing BDA on the Internet of Things (IoT) environment. Bashir and Gill (2016) proposed an IoT big data analytics framework to overcome the challenges of storing

Table 4
The Seven V's characteristics of Big Data Analytics.

Name	Description	Examples	Challenges
Volume (Barnaghi, Sheth, & Henson, 2013)	Volume of big data is explained in terms of its size and exponential growth. Large-scale and the sheer volume of data is a big challenge. It is known as size. Applications: -Medical data, Social media	Scale of data: -Terabyte -Petabytes -Exabyte -Yotabyte	-Data Storage -Data acquisition -Processing of data -Performance -Cost
Variety (Chen et al., 2013)	It refers to the complexity of large data set which may be semi-structured, unstructured or structured. It is known as complexity. Applications: Weather data, DNA Sequencing, Biology	Different forms of data: -Text, documents -Images, voice, audio, video -Geo-spatial data -Network data -Sensors data	-Heterogeneity of data -Diverse -Dissimilar forms
Velocity (Sivarajah et al., 2017)	It is a high rate of data inflow with non-homogenous structure. It is known as speed. Applications: Financial market, ad agencies	Analysis of streaming data: -Batch processing -Real-time processing -Streaming processing	-Slow and expensive nature of data processing
Veracity (Vasarhelyi et al., 2015)	Veracity feature measures the accuracy of data and its potential use for analysis. It is known as quality.	Uncertainty of data: -increasingly complex data structure, -inconsistency in large data-sets	-Accuracy of data -Reliability of the data sources -Context within Analysis -inaccuracy, latency, subjectivity
Valence (Sivarajah et al.,2017)	It refers to the connectivity of big data in the form of graphs. It is known as Connectedness. Applications: Healthcare data	Measure of Connectivity: -Data Connectivity	-More complex data exploration algorithm. -Modeling and prediction of valence changes. -Group event detection. -Emergent behavior analysis
Value (Sivarajah et al., 2017)	Big Data = Data + Value? It is the heart of the data challenge. It extracts knowledgeable value from vast amounts of structured and unstructured data without loss, for end users. Applications: Business or industries	Seven V's: -Size -Complexity -Quality -Connectedness -Speed -Variations -Value (important)	-Increase revenue -Decrease operational costs -Serve Customers
Variability (Sivarajah et al., 2017)	It refers to data whose meaning is changed constantly and rapidly. It remains a constant challenge. Application: Stock market, finance data	Variation in data flow rates -Complexity	-Inconsistency of data -Peak-level computing Demand -Periodic peaks and Troughs

and analyzing a large amount of data originating from smart buildings. Rathore, Ahmad, and Paul (2016) proposed a smart city management system based on IoT that exploits big data and analytics. Sezer, Dogdu, Ozbayoglu, and Onal (2016) proposed an augmented framework that integrates semantic web technologies, big data, and IoT.

For the processing and analysis of Big Data, various recently used platforms are investigated for large amount of IoT generated data as follows: (i) enabling capability for storing & processing large amount of data (Apache Hadoop, 2011), (ii) enabling capability for advanced data analytics: extraction, transfer and loading (ETL) (1010data), (iii) enabling capability of big data IoT processing and analytics (SAP-Hana, 2013), (iv) enabling capability that support for Hadoop in order to big data processing and analysis (Cloudera, 2008), (v) enabling capability for parallel processing, analysis and security for unstructured data (HP-HAVEn, 2013), (vi) enabling capability for Hadoop based processing and analysis on large amount of data (Hortonworks, 2011), (vii) enabling capability for analytical database that combine massively parallel processing (MPP) petabyte scale volume data (Pivotal big data suite, 2016), (viii) enabling capability for data analyze and management problem solving up to 50 terabyte (Infobright, 2005), (ix) enabling capability for fast processing, analyzing, and predictive capabilities (MapReduce, 2008).

Further, the structures of the top primary studies are classified. The structure for classification is based on the method which was proposed by (Jabbour, 2013). The classification scheme includes six categories: - namely study, objective, focus, capabilities, benefits, and their results as shown in Table 5.

- *Study:* It consists conceptual, theoretical, empirical, literature review, and case study.
- *Objective:* Various objectives of BDA, related review, and research.
- *Focus:* Various researches focus on direction of BDA in different domains of application.
- *Capabilities:* It includes important data capabilities such as analytics, prediction, decision, and management.
- *Benefits:* Various benefits and impact of BDA.

Table 5
Top 20 Primary researches composed the sample.

Primary	Study	Objective	Focus	Capabilities	Benefits	Result
1. (Pääkkönen & Pakkala, 2015)	-Conceptual -Theoretical	Survey and use case on BDA applications	Referenced architecture on commercial product and services for BDA system	-Analytical	Commercial product & services for BDA system	A new perspective on research
2. (Oussous, Benjelloun, Lahcen, & Belfkih, 2017)	-Conceptual -Theoretical	Survey of BDA technologies and algorithms	Various tools and technologies	-Analytical -Decision	BDA opportunities, application, challenges and issues	A new perspective on research
3. (Liu, Li, Li, & Wu, 2016)	-Theoretical -Empirical -Case Study	Survey of empirical studies on data quality & data usage of BDA	Big data errors in spatial information science	-Decision	Reduced big data error in data collection processing and analysis	A new perspective on research
4. (Zhang & Xiang, 2015)	-Conceptual -Theoretical	Survey of BDA data quality	Data quality solutions for business organization	-Analytical	Increase the data privacy, security, quality issues	Consistent with previously published literature
5. (Mikalef, Pappas, Krogstie, & Giannakos, 2017)	-Conceptual -Theoretical -Literature Review	Literature survey on BDA	Resource base theory (RBT) and capabilities of BDA	-Analytical -Decision -Management	Theoretical framework on business value and competitive advantage for BDA application	Research in domain of knowledge to gain insights through analysis
6. (Yaqoob et al., 2016)	-Theoretical -Case study	Survey on BDA processing, technologies & organization case study	Usage in many multidisciplinary application	-Analytical -Predictive -Decision	Increase in productivity of industries/ companies and provide consumer density of the firm with BDA	Research in domain of knowledge to gain insights through analysis
7. (Zhou et al., 2016)	-Conceptual -Theoretical	Systematic review of BDA for smart energy management.	Industrial development of big data driven smart energy management	-Decision -Management	Energy Efficient big data- driven optimization & real-time monitoring and forecasting	Research in domain of knowledge to gain insights through analytics
8. (Tu et al., 2017)	-Conceptual -Theoretical	Survey on smart grid integration of BD management and BDA algorithm	Empirical studies on smart grid and energy big data analytics	-Analytical -Decision -Management	Stability & Reliability Utilization & Efficiency Better Customer Satisfaction	A new perspective on research
9. (Kshetri, 2016)	-Modeling -Literature review	Survey on role of big data in facilitating the access to financial services in china	Analytics in Financial companies	-Analytical -Decision -Management	The use of BDA helps to overcome the reducing information opacity and transaction costs	A new perspective on research
10. (Cerchiello & Giudici, 2016)	-Modeling -Literature review	Systemic risk model based on big data	Financial Risk management/ markets and tweets	-Predictive -Management	Understanding of Financial services	Replication to a different context or period
11. (Yang, Zhong, Liu, & Feng, 2014).	-Conceptual -Theoretical	Theoretical framework of financial data classification standard	Improve Big data storage mode.	-Analytical	Eliminate data noise and to remove data redundancy	Consistent with previously published literature
12. (Sun, Chen, & Yu, 2015)	-Modeling -Empirical	Generalized optimal wavelet decomposing algorithm for big financial data	Big Financial and financial analytics	-Analytical -Predictive -Decision -Management	FA provide better understand the viability, stability, and profitability of business/ market/beneficial decisions	Research in domain of knowledge to gain insights through analytics

(*continued on next page*)

Table 5 (*continued*)

Primary	Study	Objective	Focus	Capabilities	Benefits	Result
13. (Crawley & Wahlen, 2014)	-Literature Review	Survey on analytics in empirical/ achieving financial accounting	new analytics for testing hypotheses questionnaires analytics in accounting	-Analytical	Informed business professionals about financial accounting	Consistent with previously published literature
14. (Tian, Han, Wang, Lu, & Zhan, 2015)	-Empirical	System architecture for Big data analytics	Analysis on the critical latency analytics requirement in finance using BDA	-Management	Useful for the banking and financial organizations	Comparative research
15. (Edwards & Taborda, 2016)	-Theoretical	Review on domain of analytics, risk management, and knowledge management	To understand relationship between the knowledge data, techniques, and experience	-Analytics -Decision -Management	Understand the knowledge domain using data analytic capabilities	Replication to a different context or period
16. (Wu, Li, Cheng, & Lin, 2016b)	-Theoretical -Empirical	Healthcare-wearable technology optimize insights	Bring new opportunities for healthcare-wearable device providers	-Analytics	higher-quality firms, the optimal quality level for health and biomedical sector	A new perspective on research
17. (Cetin, Demirçiftçi, & Bilgihan, 2016)	-Theoretical	Review of revenue management challenges	Hotel revenue managers with KSAs required in managing inventory and prices	-Decision -Management	Helps to understand the revenue management challenges	Research in domain of knowledge to gain insights through analytics
18. (Addo-Tenkorang and Helo, 2016)	-Theoretical -Literature Review	Review on “big data,” its application and analysis of operations or supply-chain management	Big data applications attempting to identify and understand the challenges in industrial or supply chain	-Analytics -Decision -Management	The four main attributes or factors identified with “big data” Variety, Velocity, Volume, Veracity, and Value-adding	Comparative research
19. (Hazen, Skipper, Ezell, & Boone, 2016)	-Theoretical	Review of big data and predictive analytics	Focus on eight theory-driven impact of BDPA's on supply chain management	-Analytics -Management	Identifying development of BDA Prediction and modern competitive upon firm performance	A new perspective on research
20. (urRehman et al., 2016)	-Theoretical -Case Study	Review of the big data analytics process and popular relevant tools for value-creation	Proposed knowledge-driven based big data reduction framework for value creation	-Analytics -Management	It enables local knowledge availability, privacy preservation, and secure data sharing functions to build trust between customers and enterprises.	Research in domain of knowledge to gain insights through analytics

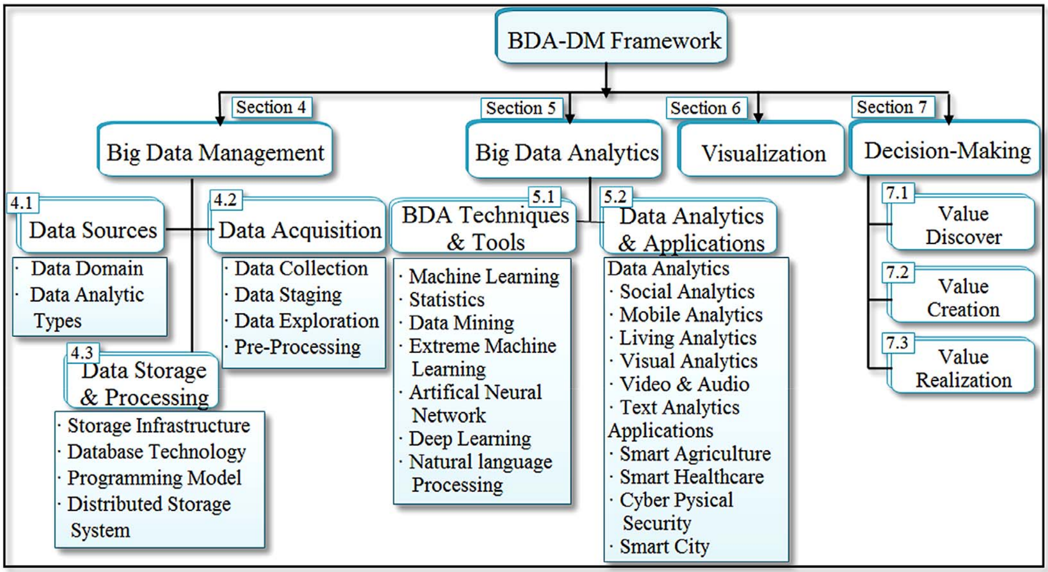

Fig. 5. Review structure of the paper.

- *Result:* It shows the useful value insights as in result from these primary study articles.

The studies primarily focused on different applications areas of BDA. Various researchers proposed framework for Big Data and Product Lifecycle Management (BDA-PLM) (Zhang, Ren, Liu, Sakao, & Huisingh, 2017), the big scholarly data lifecycle (Assuncao et al., 2015; Khan, Liu, Shakil, & Alam, 2017), 3-As Data Quality-in-Use model for data quality characteristics for use in Big Data projects (Merino, Caballero, Rivas, Serrano, & Piattini, 2016), the Unified Technology Acceptance and Usage Theory (UTAUT) aligned with the idea of Big Data as a Service (Shin, 2016), novel conceptual basis Operational business intelligence (OpBI) systems designed with value-based business requirements (Hänel & Felden, 2015), unified and dynamic framework analysis of big data business values and managerial, operational, organizational changes led by data-driven approach (Sheng, Amankwah-Amoah, & Wang, 2017), conceptual model of the seven V′s of big data analytics to gain a deeper understanding of the strategies and practices of high-frequency trading (HFT) in financial markets (Seddon & Currie, 2017).

3. Big data analytics & decision-making framework (BDA-DMF)

The framework of Big Data Analytics and Decision-Making Framework (BDA-DMF) is shown in Fig. 5 to discover value in the business ecosystem. This figure indicates the big data management, big data analytics, data visualization, and decision-making for value-creation that are discussed in Sections 4, 5, 6, and 7 respectively.

RQ2: *How to design BDA-DM framework?*

Big data analytics is a data-intensive architecture that provides various technologies and platforms used in various phases such as data generation, data acquisition, data storage, advanced data analytics, visualization and decision-making for value-creation as shown in Fig. 7. It follows a top-down approach. It consists various techniques and technologies i.e. Hadoop, HBase, Cassandra,

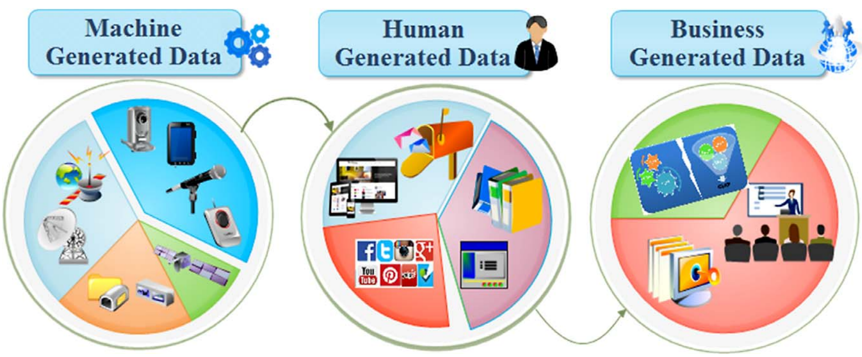

Fig. 6. Different types of data domain.

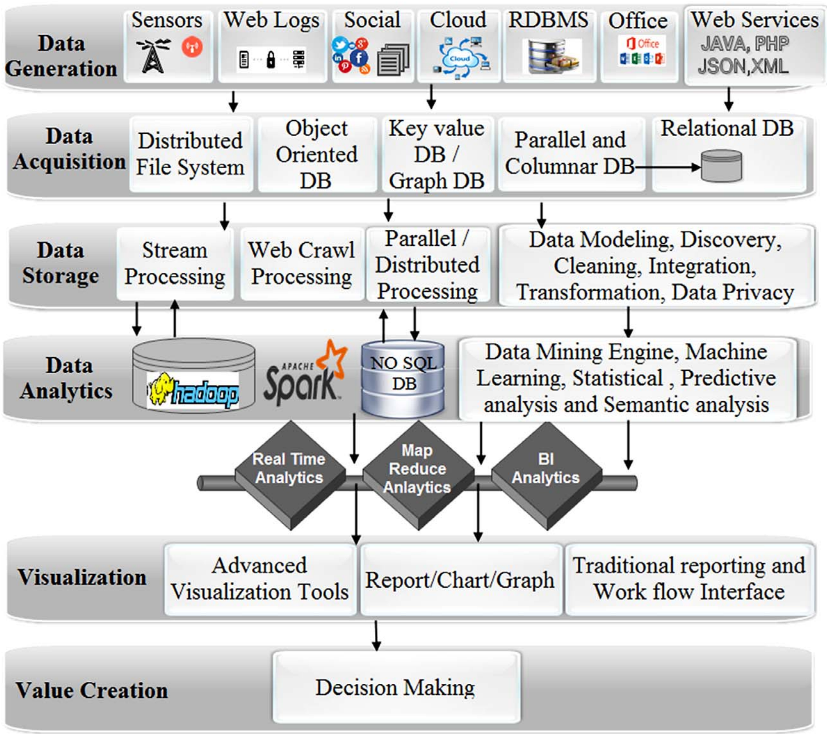

Fig. 7. Architecture of big data analytics.

MongoDB, NoSQL and so on. As limitation, these tools and techniques cannot solve the real word problems of data storing, data searching, data sharing, data visualization, and also real-time analysis.

4. Big data management

Big data management (BDM) provides an infrastructure to Big Data Analytics, where data management techniques, tools, and platforms including storage, pre-processing, processing and security can be applied (Bilal et al., 2016; Siddiqa et al., 2016). The components involved in BDM are described as:-

4.1. Data sources

Big data generation refers to generation of data from various relevant sources. It can be generated by humans, machines, business processes, and data techniques that are descriptive, predictive, and prescriptive.

4.1.1. Data domain

A flourishing domain of data is expressed by variety of descriptive terms such as:-structured, unstructured, machine and sensor-generated data, batch, and real-time processing data, biometric data, human-generated data, and business-generated data. The Fig. 6 shows the relevance for various generations of big data analytics domains

- ***Machine-Generated Data:*** The machine-generated data comes from several computer networks, sensors, satellite, audio, video streaming, mobile phone applications, and prediction of security breaches.
- ***Human-Generated Data:*** It can be collected by people, for example: identification details having their name, address, age, occupation, salary, qualification etc. Whereas, real streaming data can be generated by various files, documents, log files, research, emails, and social media websites such as Facebook, Twitter, YouTube, LinkedIn.
- ***Business-Generated Data:*** The volume of business data of all companies across worldwide is estimated to double every 1.2 years such as transactional data, corporate data, and government agencies data. When Business intelligence (BI) of BDA is discussed, it means: value (does the data contain any valuable information for my business needs?), visibility (focus of insight and foresight of a problem and an adequate solution associated with it) and verdict (potential for decision-makers based on problem, computational capacity and resources) within the business intelligent domain (Wu, Buyya, & Ramamohanarao, 2016a).

4.1.2. Data types

Following are the three types of analytics that organizations and industries can use to learn and get the insights to promote their business.

- ***Descriptive***: It is composed of various technologies and summaries of inferred data that represent current and previous happening process. Standard reporting, ad-hoc reporting, dashboards, querying, and drilling down are the various examples of descriptive analytics. It is defined as look into past in order to draw some inferences.-"What has happened?"
- ***Predictive:*** The predictive analytic modelings are root-cause analysis, Monte Carlo simulations, and data mining. It is sometimes used in real-time or in batch-time processes. Siegal (2010) illustrated that seven sequential objectives are organized by adopting these predictive analytics namely compete, grow, enforce, improve, satisfy, learn, and act. It predicts future trends.-"What could happen"?
- ***Prescriptive:*** This technique is applicable to future scenario and advises a solution or insightful actions from the predictions. Basu (2013) represented the five pillars of prescriptive analytics namely: hybrid data, integrated predictions and prescriptions, prescriptions and side effects, adaptive algorithm, and feedback mechanism.-"What should we do?"

4.2. Data acquisition

Here, data acquisition covers a broad spectrum of collecting, filtering and cleaning process of data ingesting in a data warehouse or any other databases. (Chen, Mao, & Liu, 2014) investigated that data acquisition supports heterogeneity due to a variety of devices.

4.2.1. Data collection

It is a process to acquire the unprocessed data from real-world environment, and develop it proficiently. Log files are widely used to expand data collection that is generated by multiple sources and all applications working on electronics devices such as extended log format (W3C), common log file format (NCSA) and IIS log format (Microsoft).

Sensors are another substitute that measures a physical quantity and transfers it in readable form by digital signals. There exist several types of sensors such as audible, sound, automotive, vibrate, electric current, weather, thermal, pressure transferred through wired or wireless networks. Web crawlers are generally used to collect data or applications from various website based processes such as (web search engines or web caches) (Castillo, 2005).

4.2.2. Data staging

Further, it is defined as a process for the collecting of wide variety of data sets along with noisy, redundant, and consistent data. It is divided into two alternative models namely: - the streaming processing models and batch-processing models. The streaming processing model analyzes the data as soon as possible to derive its results where the data arrives in continuous form at very fast speed. To support it, there are some open source systems that include Storm, S4, and Kafka (Hu, Wen, Chua, & Li, 2014).

In the batch-processing model, data is first stored and then analyzed. In this model, MapReduce (Dean & Ghemawat, 2008) has become the dominant platform. Fig. 8 shows (a) the data staging into two parts of data exploration and data pre-processing forms and (b) the predictive model.

- ***Data Exploration:*** There are two main aims of data exploration. Firstly, to determine and understand nature as well as characteristics of data. Secondly, to determine the data quality issues that can badly affect the model. Data exploration and data mining are widely used to discover new insights. For example: - data quality report (mean, mode, median, and range; standard deviation and percentiles; bar plots, histograms and box plot) and data quality issues (valid or invalid).
- ***Pre-Processing:*** To extract the meaningful information from the big data, it is necessary to clean, integrate and transform the data (Hu et al., 2014) through various tools namely Apache Hadoop, NoSQL, and MapReduce. Pre-processing is related to series of steps namely how to integrate data, how to transform data, how to select the right model for analysis and how to provide the results.

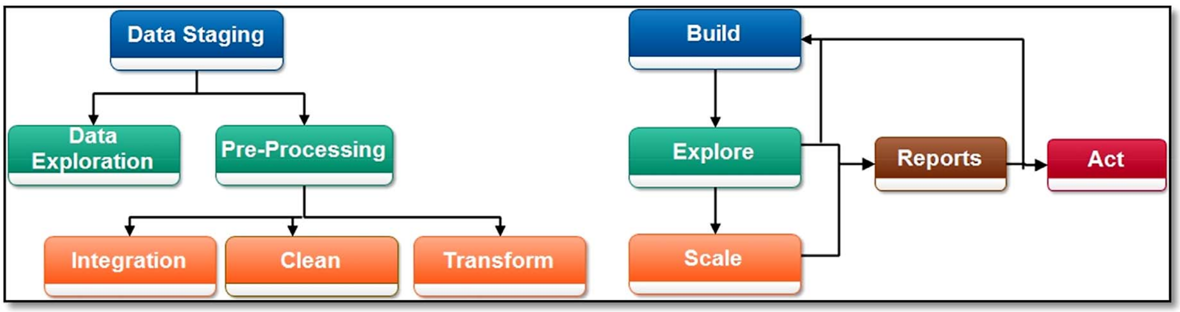

Fig. 8. (a) Data staging and (b) The predictive model.

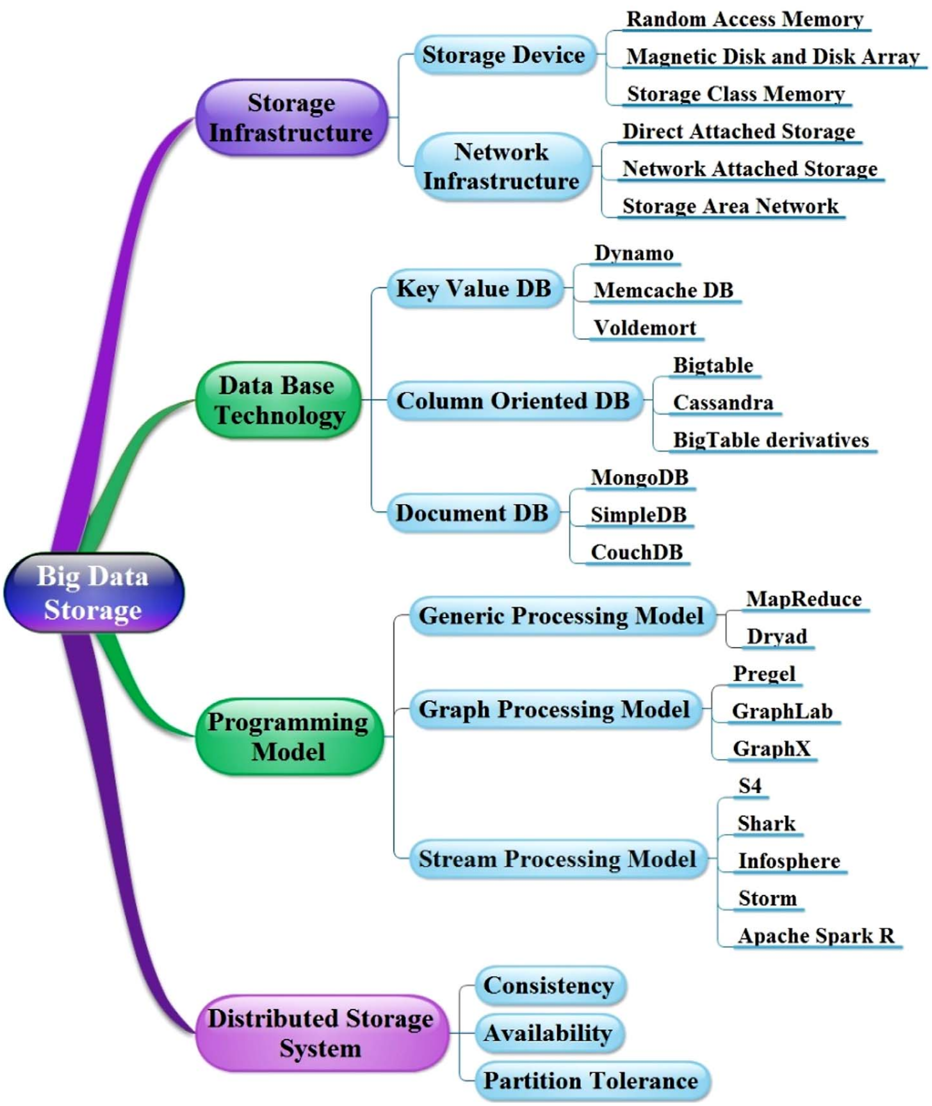

Fig. 9. The Platform of various big data storage.

-Cleaning: It is an essential goal of pre-processing which clean, address the data quality and format because of its messy nature. It enables us to discover imprecise, insufficient, or immoderate data that requires altering, removing and improving data quality.
-Integration: With the use of extract, transform and load (ETL) process, the data can be cleaned, well transformed and made it applicable to data mining and various online analytics.
-Transformation: The transformation of the raw data is to make it suitable for analyzing and getting data into shape such as integrating and packaging of data using some tools: ETL, DMT, and Pig. There are various actions that can be applied in the real-time format of data such as splitting of data, merging it, performing computations, connecting it with the outside data domain and spreading data to multiple destinations.

4.3. Data storage & processing

It is the process of managing data storage. It performs activities in parallel to optimize the storage process. Data clustering, replication, and indexing are adequate activities that are significant to accomplishing the storage phase in big data management (Siddiqa et al., 2016).

It refers to how numerous types of data can be stored in different forms after collecting them from different sources. There are various useful tools for big data storage namely Hbase, NoSQL, Gluster, HDFS, and GFS (Gandomi & Haider, 2015; Pole & Gera, 2016). (Cheptsov & Koller, 2015) introduced an innovative approach to parallelism data-centric based application on the message passing interface. The Fig. 9 describes the big data storage for various platforms (Hu et al., 2014; Chen, 2014).

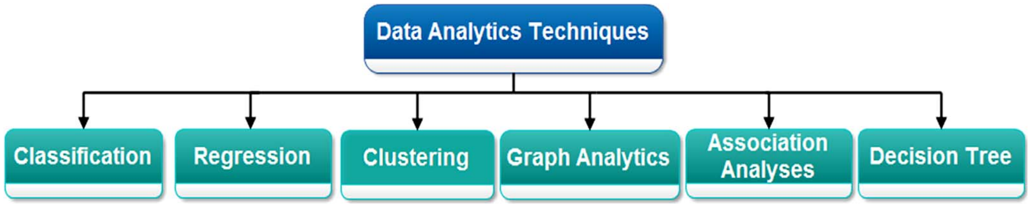

Fig. 10. Classification of different data analytics techniques.

5. Big data analytics

Advanced Big Data Analytic process refers to analyze heterogeneous data and mine insightful information through unknown patterns by applying various predictive algorithms, semantic analysis, statistical analysis methods, and technologies. Collection and transportation of big data share a common goal: - analyzing the data for insights and better application guidance (Li & Jain, 2013). (Fahad et al., 2014) described few efficient algorithms such as sampling, data condensation approaches, density-based approach, grid-based approach, divide and conquer, incremental learning and distributed computing. (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) presented the steps that composed with knowledge discovery in database process. They defined significant iterations such as selection of data, pre-processing of data, the transformation of data, data mining algorithms applying to enumerate patterns for proper interpretation of results and to ensure useful knowledge discovery from data.

Tsai, Lai, Chao, and &Vasilakos (2015) presented big data analytics of various infrastructures that are categorized in the manner (i) Processing or Computing: Hadoop, Nvidia CUDA, or Twitter storm, (ii) Storing: Titan or HDFS, and (iii) Analyzing: MLPACK or mahout. There are some other tools such as Whiteboard, R, MATLAB, octave refer for (kilobyte to low megabyte); Numpy, Scipy, Weka, Blas refer for (megabyte to low gigabyte); and Hive, Mahout, Harna, Giraph refer for (gigabytes to terabyte).

5.1. BDA techniques

The recent advancements in techniques and technologies have enabled many enterprises to handle big data efficiently. The data analytics techniques are machine learning, data mining, statistics, artificial neural network, extreme machine learning, natural language processing, and deep learning etc. The Fig. 11 shows the origin of BDA techniques. BDA has led to numerous technologies to perform an analytics. Overview of Big Data Analytics Machine learning tools are described in Appendix A.

***RQ3:** What are the main tools, techniques, and technologies of Big Data Analytics?*

5.1.1. Advanced machine learning

Advanced machine learning (ML) analytic is an umbrella action that defines the selection of analytical technique to build a model for evaluation of an efficient result. By tradition, machine-learning research is divided into two categories: logical representations and statistical ones. Initially, it selects an input data technique to build a predictive model and generate model output or validate. The Fig. 8(b) shows the predictive model for activity iterative process including build, explore, scale, report, and act.

The most common predictive analyzing techniques that are used for advanced data analytic such as classification, clustering, regression, association analyzes, graph analyzes and decision tree. The predictive data analytic applications are supervised ML and unsupervised ML algorithms. The supervised ML methods are self-learning models that represent relationship between a set of descriptive and a target feature based on historical examples. However, in supervised machine learning, the first one category is regression which includes linear regression, generalized linear model, ensemble methods, decision trees, neural networks. The Fig. 10 shows the classification of different analytic data techniques.

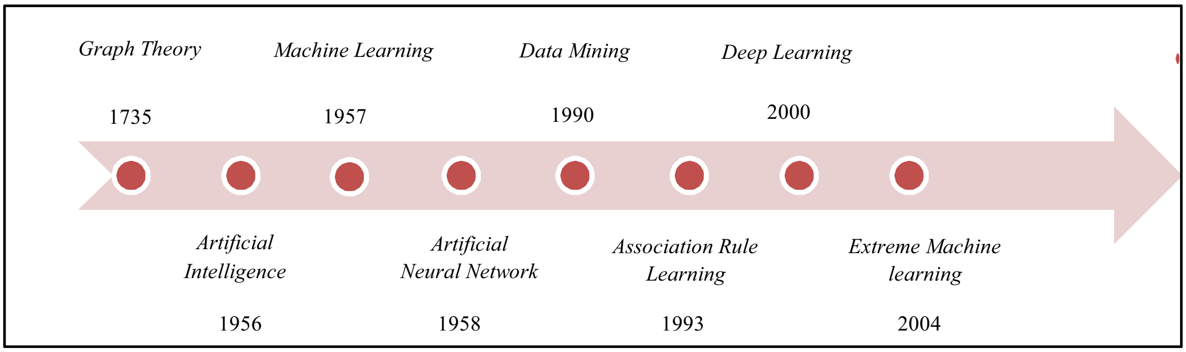

Fig. 11. Origin of Big data analytics techniques.

- ***Classification:*** To predict the categories of input data for e.g. weather attributes are sunny, windy, rainy etc.
- ***Regression:*** To predict numeric value e.g. price of stocks.
- ***Clustering:*** To organize similar items in-to groups e.g. grouping a company in senior, adults, and teenagers.
- ***Association Analyzes:*** To find interesting relationships between sets of variables.
- ***Graph Analyzes:*** To use graphic structure to find connections between entities.
- ***Decision Tree:*** To predict modeling insights of objective variables by learning simple decision rules inferred from the data features.

Further, it consists classification algorithms such as support vector machines, discriminate analysis, naive Bayes, and nearest neighbour. Unsupervised machine learning uses clustering techniques which include various models like as k-means clustering, k-medoids, fuzzy c-means, hierarchical, Gaussian mixture, neural networks, and hidden Markov model. There are used in various real-time applications such as medical diagnosis, stock trading, energy load forecasting, weather forecasting etc.

5.1.2. Advanced statistics

Advanced statistics analytics is primarily based on various tools and techniques for collecting, analyzing and visualizing the result from the large scale of data. It includes different domain of analytics that derives techniques from statistics and data-driven analysis that executes statistics algorithm. The statistical technique refers to clustered analytics, data mining and predictive modeling methods.

5.1.3. Advanced data mining

The BD mining is the most challenging technique as compared to traditional data mining such as pattern discover and extraction. Data mining depends on techniques such as data statistics, machine learning methods and pattern recognition (Chen & Zhang, 2014). Multiple linear regression and logistic regression are also commonly used in data mining, which includes various algorithms such as k-means clustering, association analysis, and decision trees. Overview of big data analytics techniques and their applications area shown in Table 6.

5.2. Big data analytics & applications

There are many techniques that can be used to analyze big data. This work presents various analytic technologies areas in which BDA is applicable as follows.

5.2.1. Social analytics

Social analytics is an important and growing analytics of real-time data analytics. It is categorized into social networks (e.g., Facebook and LinkedIn), blogs (e.g., Blogger and Word Press), micro blogs (e.g., Twitter and Tumblr), social news (e.g., Digg and Reddit), social bookmarking (e.g., Delicious and Stumble Upon), media sharing (e.g., Instagram and YouTube), wikis (e.g., Wikipedia and Wikihow), question-and-answer sites (e.g., Yahoo! Answers and Ask.com) and review sites (e.g., Yelp, Trip Advisor) general sites (Li, Chen, Wang, & Zhang, 2013) like Facebook, Instagram, Foursquare, Twitter, and Pinterest, which produce immense amounts unstructured form of data.

5.2.2. Mobile analytics

Personal mobile devices can be used as instruments to collect and monitor learning analytics towards self-regulation. It has discovered existing unknown meaningful patterns and knowledgeable data from a few dozen terabytes to numerous petabytes composed from mobile users at the network-level or the application-level (Yazti & Krishnaswamy, D. Z. 2014). There are some studies about mobile and ubiquitous learning analytics tools (Alsheikh, Niyato, Lin, Tan, & Han, 2016; Fulantelli, Taibi, & Arrigo, 2013) presented a scalable Apache Spark-based framework for deep learning in mobile big data analytics.

5.2.3. Living analytics

It is associated with the study of social and behavioral forms of individuals and societal groups. The domain of analytical social science is integrally using advances in storage and computing abilities to process readily in big data (Lazer et al., 2009). Several common challenges of living analytics with big data include high volume, high velocity, high dimensionality, sparse data, and a variety of diverse data sources and formats, etc.

5.2.4. Video and visual analytics

Video analytics is the research field that addresses the scalable and reliable analysis of video data. The visual analytics is described as "the science of analytical reasoning facilitated by interactive visual interfaces" and its general goal is to generate insight from data. It is a fascinating branch of big data investigation to provide analytical reasoning over collaborative visual interfaces.

5.2.5. Text analytics

It refers to techniques that can extract information from textual data. It contains statistical, computational linguistics, and machine learning (Gandomi & Haider, 2015). Text analytics assist businesses to adapt large volumes of human-generated text into meaningful insights, which supports evidence-based decision-making. Broadly speaking, summarization follows two approaches: the

Table 6
Overview of Big Data Analytics techniques and their application area.

Name	Review papers /Title	Reference	Application area	Reference
Machine learning	Strategies and principles of distributed machine learning on big data.	(Xing, Ho, Xie, & Wei, 2016)	-Analyzing social networks Interpreting texts, images, and videos	(Airoldi, Blei, Fienberg, & Xing, 2008), (Chandola, Banerjee, & Kumar, 2009), (Lee & Xing, 2012), (Zhao & Xing, 2014)
	Machine Learning on Big Data: Opportunities and Challenges.	(Zhou, Pan, Wang, & Vasilakos, 2017)	-Identifying disease and treatment paths	
	A survey of machine learning for big data processing.	(Qiu, Wu, Ding, Xu, & Feng, 2016)	-Tracking anomalous activity for cyber-security	
Extreme machine learning	Trends in extreme learning machines: A review.	(Huang, Huang, Song, & You, 2015)	-Computer vision	(He et al., 2014)
	Extreme learning machines: a survey.	(Huang, Wang, & Lan, 2011)	-Image processing	(An & Bhanu, 2012)
	Extreme learning machine: algorithm, theory and applications.	(Ding, Zhao, Zhang, Xu, & Nie, 2015)	-System modeling and prediction medical/biomedical application	(Tian & Mao, 2010)
				(You, Lei, Zhu, Xia, & Wang, 2013)
			-Time series analysis	(Butcher, Verstraeten, Schrauwen, Day, & Haycock, 2013)
Artificial Neural Network	Artificial neural network learning: A comparative review.	(Sovilj, Sorjamaa, Yu, Miche, & &Severin, 2010)	-Chemical engineering	(Zhang, 2000)
			-Cancer prediction	(Himmelblau, 2000)
	Artificial neural networks in business: Two decades of research.	(Neocleous & Schizas, 2002)	-Disease Prediction	(Agrawal & Agrawal, 2015)
	Artificial neural networks and its applications.	(Tkáč & Verner, 2016).	-Agriculture	(Weng, Huang, & Han, 2016) (Francik et al., 2016)
	Neural networks for classification: a survey.	(Jha, 2007).		
Data Mining	Educational data mining: A survey and a data mining-based analysis of recent works.	(Peña-Ayala, 2014)	-Educational data mining	(Chaturvedi & Ezeife, 2012)
			-Business & Management	(Baker & Yacef, 2009)
	Data mining techniques in social media: A survey.	(Injadat, Salo, & Nassif, 2016)	-Medical and Health	(Moss, Corsar, & Piper, 2012)
			-Social Networks	(Alowibdi, Buy, Philip, & Stenneth, 2014)
	Application of data mining techniques in customer relationship management: A literature review and classification.	(Ngai, Xiu, & Chau, 2009)	-Wind energy systems	(Soman, Zareipour, Malik, & Mandal, 2010)
			-Biomedicine	(Phillips & Buchanan, 2001)
	Data mining techniques and applications – A decade review from 2000 to 2011.	(Liao, Chu, & Hsiao, 2012)	-Finance	(Vavpetic, Novak, Grcar, Mozetic, & Lavrac, 2013)
	Data mining and wind power prediction: A literature review.	(Colak, Sagiroglu, & Yesilbudak, 2012)		
Deep Learning	Deep learning for visual understanding: A review.	(Guo et al., 2016)	-Image classification	(Krizhevsky, Sutskever, & Hinton, 2012)
			-Object detection	(Hoffman et al., 2014)
	Deep learning applications and challenges in big data analytics.	(Najafabadi et al., 2015)	-Image retrieval	(Liu, Guo, Wu, & Lew, 2015)
			-Semantic segmentation	(Dong, Chen, Yan, & Yuille, 2014)
	A survey of deep neural network architectures and their applications.	(Liu et al., 2017)	-Human pose estimation	(Ouyang, Chu, & Wang, 2014)
			-Speech recognition	(Bengio, 2013)
Natural Language Processing	Machine learning and natural language processing.	(Marquez, 2000)	-Spelling and grammar checking	(Zuker & Sankoff, 1984)
	A tutorial on techniques and applications for natural language processing.	(Hayes & Carbonell, 1983)	-information retrieval	(Saidi, Maddouri, & Nguifo, 2010)

extractive approach and the abstractive approach. In extractive approach, a summary is produced from the original text units. In comparison, abstractive approach contains extracting semantic information from the text.

5.2.6. Audio analytics

Audio analytics analyze and extract data from unstructured audio data such as human spoken language and it is referred to as speech analytics (Gandomi & Haider, 2015). The benefits that can be achieved are summarized while using these techniques for specific application areas of storage, pre-processing, and analysis etc.

RQ4: *What are the main application areas of Big Data Analytics?*

The various data analytics applications such as smart agriculture, smart healthcare, cyber-physical security, and smart cities are briefly described as.

5.2.7. Smart agriculture

As the technology rapidly spread in few decades, big data analytics is the key to fostering a new revolution in agriculture. It has evolved technology to solve real-world problems based on historical data, machine-generated data and real-time streaming data. Agricultural IoT generates a large volume of agricultural information (Lee, Hwang, & Yoe, 2013). Agriculture firms are adopting big data technologies with a promise to gain insights from the large amounts of heterogeneous data, to solve the problem of real-time, manage data incompleteness, and lack of prior knowledge, and capturing a variety of data in a complex form.

Smart agriculture is beneficial 'use case' in big IoT data analytics. Sensors are the actors in the smart agriculture 'use case'. These are installed in fields to obtain data on the moisture level of soil, trunk diameter of plants, micro-climate condition, and humidity level, as well as to forecast weather. It passes through an IoT gateway and the internet to reach the analytics layer (Marjani et al., 2017). The analytics layer processes the data obtained from the sensor network to issue commands. Automatic climate control of harvesting, timely controlled irrigation and humidity control for fungus prevention are the examples of actions performed on the basis of big data analytics (Gubbi, Buyya, Marusic, & Palaniswami, 2013).

Kshetri, (2014) presented a case study of agriculture to get various benefits, opportunities, and threats by implementing BDA and suggested about soil status to farmers, extreme variations in the weather patterns, new ways of planting, topography. It also provides information regarding variable market condition. Jiang, Chen, Dong, and Wang (2013) predicted the difficulties in sensors for storage and analysis by applying a large amount of data. So putting forward a distributed storage based on DSM architecture and combined with agriculture PaaS platform to provide service.

Xie, Zhang, Sun, and Hao (2015) proposed a big data processing technology to obtain a hierarchy of agricultural information system from the following aspects: gathering, storing, analyzing, and visualization of agricultural big data. This paper described that how to deal with the flood of agricultural data from the view of the big data technologies using Map Reduce Tool. The BDA provides a new insight to give advance decision support to improve yield productivity, and avoid unnecessary costs related to chemical fertilizers and pesticides. Bendre, Thool, and Thool (2016) presented the different sources of big data and types in precision agriculture, ICT-based e-Agriculture. Finally, they discussed rainfall prediction application using supervised and unsupervised method for data processing and forecasting.

5.2.8. Smart healthcare

Big data analytics is an emerging revolution in healthcare and medical research for Research and Development (R&D), treatment, testing, and diagnosis for health management. As the healthcare associations are expanding day by day, because of increment in the quantity of patients, there is an expansion in medications to be utilized for their restorative treatment. In this way, it makes challenges in storing, processing and analyzing. Hence, the demand of BDA is relevant in this field also. The Health-care organizations are critically applying 'wearable real-time sensors' to analyze the current condition of the patient and treat them according to their correct diagnosis and provide medical treatment.

Therefore, during diagnosis and treatment, there is a vast collection of data such as: - structured and unstructured data, self-monitoring health data, real-time sensor devices, images, videos, various reports, and documents. Presently, there are different healthcare systems such as: - health-care management, innovation drug discovery, face recognition, verification of signatures, fingerprint, and iris. The Fig. 12 shows the process of analyzing unstructured data in health organization (Wang, Kung, & Byrd, 2016).

The Big Data in health-care maintains an information regarding patient such as case history, physician notes, Lab reports, X-ray reports, diet rule, list of doctors, and nurses in a specific hospital, national health register data, medicine and surgical instruments expiry date identification based on RFID data. These organizations are further depending on BD technology to collect data from a patient to get more insights into care and treatment.

Moreover, data-analytics creates a dedicated Center for Health Analytics and insights to address the increasing demand from hospitals, clinics, and health professionals across the world. The new big data health-care platforms: - CHESS (Batarseh & Latif, 2016), EHR, LIMS, MQIC, CMS (Ward, Marsolo, & Froehle, 2014). The Big Data analytic is used to analyze health insurance claims and leverage big data to detect fraud, waste, and error (Srinivasan & Arunasalam, 2013). Dolin, Rogers, and Jaffe (2015) presented two case studies to predict asthma in clinical document architecture (CDA) by using BDA approach.

5.2.9. Cyber-physical systems

The organization and government protect their sensitive information by using computer security networks. The Big Data is used to collect, organize, and store the data. An information technology is applied by cyber defenders to protect their data efficiently, detect

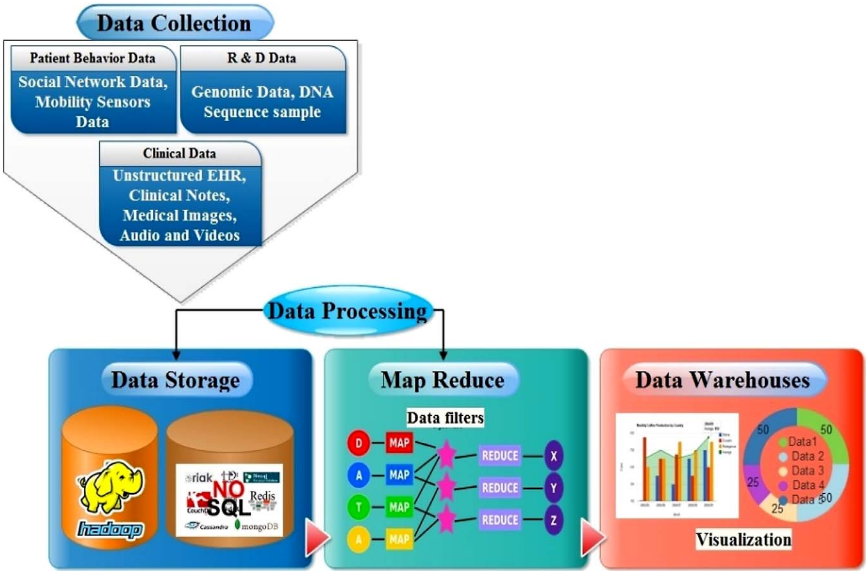

Fig. 12. Big data analytics in healthcare sector.

all malware, and cyber attackers. Developers must synchronize and make hardware component compatible with software applications using computer networks, wired or wireless sensors and different operating system, data formats and analytic system.

Hence, BDA plays a critical role to overcome the serious issues about security, privacy and thus authenticate various organizations to access data, gain a complete insight of business. The emergence of cyber-physical systems can be used for production, transportation, logistics, and other sectors to bring new challenges for simulation and planning, for monitoring, control, and interaction with machinery or data usage applications (Becker, 2016).

5.2.10. Smart cities

Smart city is a wide concept, which takes into account not only the physical structure but also human and social aspects. It utilizes several technologies to expand the performance of health, transportation, energy, education, and water services leading to an advanced level of comfort of their citizens. The application of BDA is effective for data storage and processing to generate information for diverse environments such as Smart grid environment (SGE) (Zhou, Fu, & Yang, 2016), Smart City (Ortiz-Rangel, M. 2015; Strohbach, Ziekow, Gazis, & Akiva, 2015). Smart healthcare is used to predict or diagnose the early disease (Demirkan, 2013; Roy, Pallapa, & Das, 2007).

Mostly the survey papers on Big Data Analytics have focused on discussing the opportunities, challenges, and architecture. Whereas, the concepts, architecture, challenges, and new future directions of Big Data Analytics are being presented herewith. So, this study provides useful insights through the integration of various technologies used in the application of Big Data Analytics. Data sources and application areas of big data analytics for development are shown in Table 7.

6. Visualization

A big data visualization method is concerned with the design of a graphical representation in the form of a table, images, diagrams, and spontaneous display ways to understand the data. Visual analytics has potentially brought the new federation of data mining and machine learning tools. Visual perception, design, data quality, missing data, end-user visual analytics are future trends of visualization (Becker, 2016). There are various well-known visualization analytical tools such as Dive, Rattle, FlockDB, Orange (Pole & Gera, 2016), Flare, Amcharts, and Protovis. Recently, different companies such as Amazon, Twitter, Apple, Facebook, and Google are searching visualization tools for solutions that can provide useful insights from various business aspects (Simon, 2014). The Fig. 13 shows the evolution of visualization methodology (Song, 2014). These tools and methods are appearing in form of charts, graphs, histogram, box plots, excel spread-sheets, heat maps, geographical maps etc. Interpretation is tackled with the presentation and visualization of inferences drawn in a comprehensible manner. Two main mechanisms are often used to interpret big data: - visualization and modeling. The use of big data has significant implications for modeling and theory development from a statistical-scientific point of view (Ramannavar & Sidnal, 2016).

7. BDA & decision-making framework for value-creation

In Big Data Analytics and Decision-Making Framework (BDA-DMF) for value-creation model, the framework is presented by which BDA can create value for financial & accounting companies. Framework for BDA and business insights for financial accounting

Table 7
Data sources and application areas of big data analytics for development.

Analytics	Data type	Medium	Application area	References
Social analytics	-Movie revenues	-Websites, blogs	-Sentiment analysis of social data	(Asur, & Huberman, 2010)
Mobile analytics	-Call detail Records	-Cell phones	-Social Network Analysis, Population Mobility Patterns, Transportation System -Planning, Awareness Campaigns, Mobile App. -Usage Patterns, Mobile data, traffic Analysis	(Laurila et al., 2012)
Living analytics	-Tweets and Comments -Text -Personal Health Data	-Social Media Sites -The Internet -Wearable's	-Social Network Analysis, Sentiment Analysis -Cultural Changes, Policy Effectiveness -Healthcare	(Technical report, 2014)
Visual analytics	-Images -Climate Variables, Temperature, Pollutant Levels	-Sensors -Camera	-Weather Forecasting, Pollution Control, Urban Planning	(Centro de, 2015)
Video analytics	-Anonymous Audience Data -Multimedia, Images	-Intel's Audience Impression Metrics (AIM) Suite, -Camera	-Market Research, -Public Security System -Automated security and surveillance Systems	(Balkan & Kholod, 2015) (Xu, Mei, Hu, & Liu, 2016)
Text analytics	-Text Data	-Social network feeds, email, blogs, online forum, survey responses,corporate documents, news and call-center logs.	-Stock market based, Sentiment analysis -financial news	(Gandomi & Haider, 2015) (Chung, 2014)
Audio analytics	-Voice (audio data)	-Human spoken Language	-Speech analytics, Customer call-center, Healthcare, Interactive Voice Response	(Hirschberg, Hjalmarsson, & Elhadad, 2010)
Smart agriculture	-Sensors, Text Data, Images, Audio, Video	-Documents, sensors device, GPS -Website	-Watershed management analysis -Crop modeling, Irrigation Water Management, Irrigation Scheduling	(Hu, Cai, & DuPont, 2015) (Wolfert et al., 2017)
Smart health	-Health Related Databases	-Wearable devices, sensor data, machine generated	-Electronic Health Record Analysis, clinical decision support, disease surveillance,	(Raghupathi, & Raghupathi, 2014)
Cyber-physical System	-Expert Databases	-Sensors, controller, networked manufacturing system	-CPS based Industry 4.0 Systems -CPS for TES Systems	(Lee, Bagheri, & Kao, 2014) (Lee, Jin, & Liu, 2017)
Smart cities	-Databases	-Smart phones, Computer, -Environmental sensors, -Cameras -Geographical Positioning Systems	-Transportation, Healthcare, Power grid, Smart education, Energy	(Al Nuaimi, Al Neyadi, Mohamed, & Al-Jaroodi, 2015)

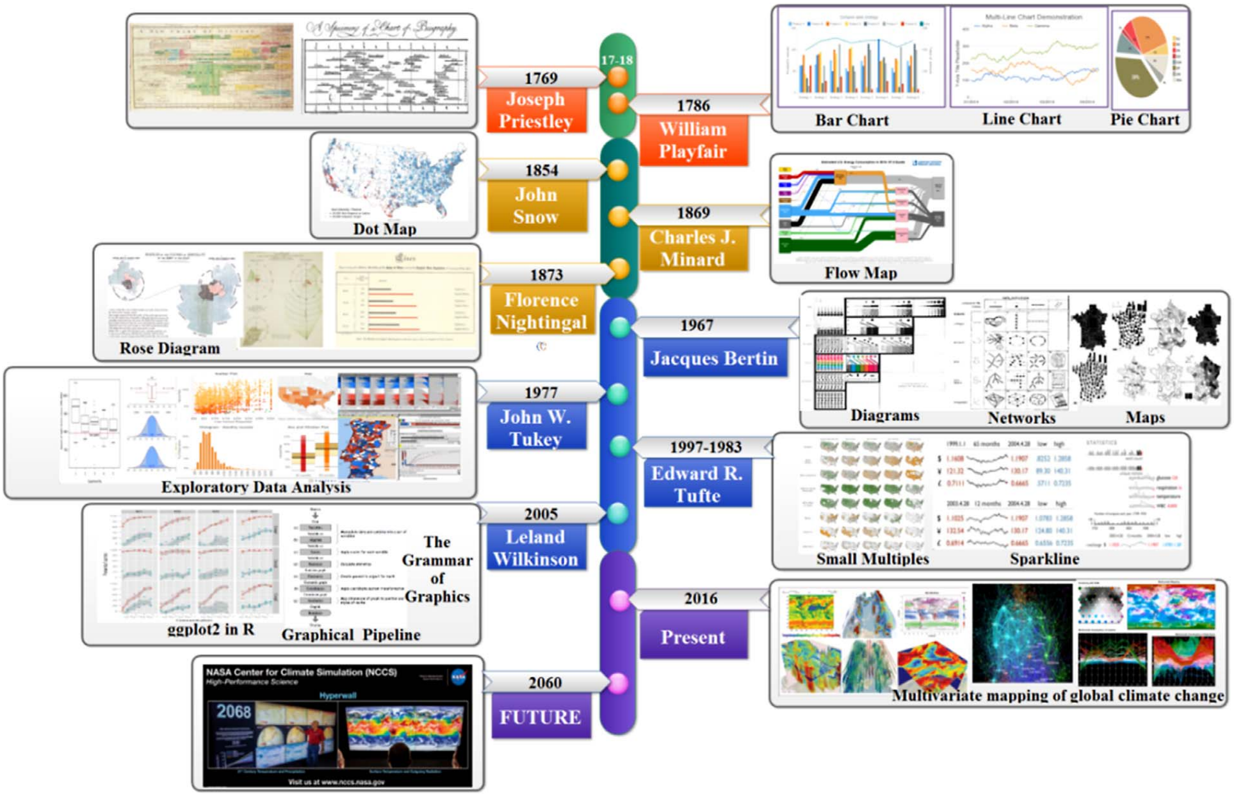

Fig. 13. The evolution of visualization methodology.

is shown in Fig. 14. This figure indicates the three-phase method by which big data analytics can create value for customers and firms:-

- *Phase1 Value Discover:* (i) Big Data Sources, (ii) Big Data Processing,
- *Phase 2 Value Creation:* (iii) Big Data Analytics, (iv) Big Data Analytics Capabilities,

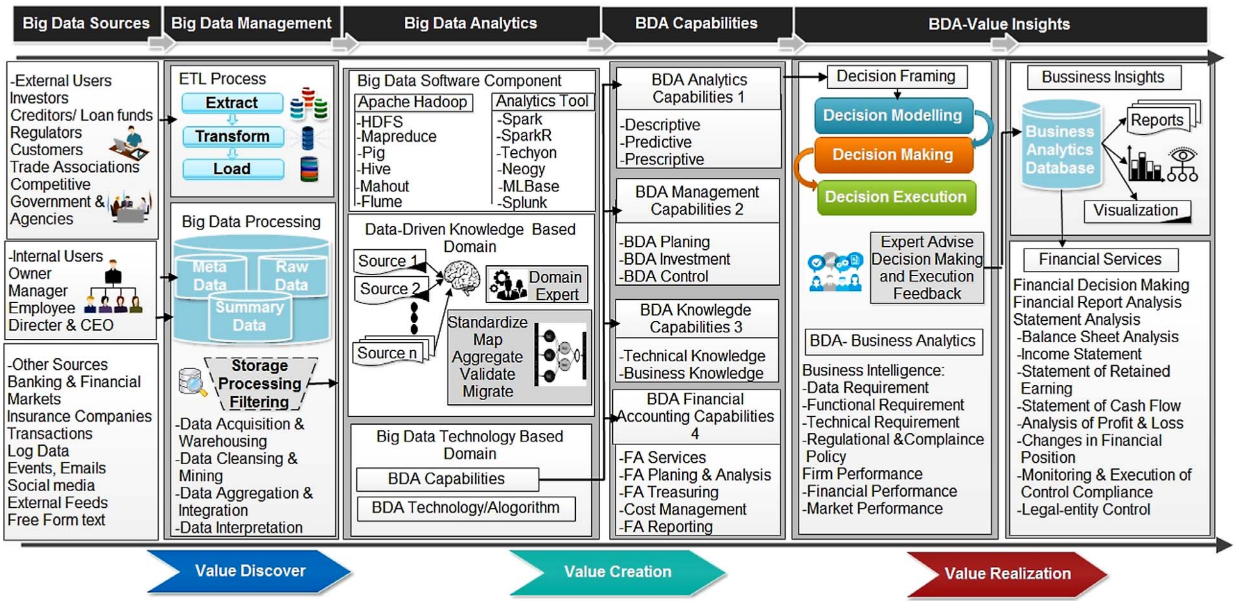

Fig. 14. Framework for big data analytics and business insights for financial accounting.

- *Phase 3 Value Realization:* (v) Big Data Analytics-Value Insights.

7.1. Value discover

In the first phase, BDA can create new insights that improve business-driven decision-making. For example, BDA can show that how firms can improve customer satisfaction and the specific features of the service experience. The growing importance of big data as a company asset is driving the development of new ways to value data assets. In the past, customer databases were considered as an important asset for firms (Srivastava, Tasadduq, & Fahey, 1998). For example, these databases could be used to create stronger relationships with customers, achieve higher loyalty, and create more efficient and effective (cross)-selling techniques. Big data holds tremendous potential for financial services firms to develop new and innovative solutions that result in a significant business value.

The BDA mainly focuses on three major values to discover by the implementation of big data technology, for example, minimize hardware costs, check the value of big data before committing significant company resources, and reduce processing costs (Leavitt, 2013). It requires congruence between business objectives, the big data storage and analytics approach. (Serrato & Ramirez, 2017) discussed three challenges for managers and decision-makers in order to take advantage of BDA. The first is to think critically about analytics techniques and the analyzes based on such data, second is to identify opportunities for creating value using BD, and the third one is to estimate the value created while using BD to address an opportunity.

Stage 1 Data Sources: The creation of value discovers, facilitate ideas or insight to better decision-making for the big data model. The process began with the data input from various sources of data. For example, external users are investors, creditors, regulation, customer, and competitor etc and internal users are the owner, manager, and employee etc. Initially, pre-processing of the data conducted to clean and transform the data into meaningful Big Data. It results in the creation of knowledge for big data discovery.

Stage 2 Big Data Management: On assessing the value of BDA to organizations, the key benefits are recognized as timely access to decision-making information, greater transparency, scalability, and better change management. Big Data Management (BDM) systems are of great value that can monitor and report the exact information a user wishes to analyze. Cloudera is an example of an “Analytics infrastructure”, which provides a Hadoop-based platform for execution of data analytics jobs in an enterprise environment. 10Gen provides “Operational infrastructure” for enterprise-grade databases and management based on MongoDB technology.

7.2. Value creation

In the second phase, the value-creation benefit of BDA is the development of more effective marketing campaigns by selecting the right customer. However, the classical definitions of marketing by (Armstrong, Adam, Denize, & Kotler, 2014) highlighted that marketing should focus on creating superior value for customers (through high quality, attractive brand propositions, and striving for an appropriate relationship), and that firms can capture value from customers in return of value creation. (Verhoef, Kooge, & Walk, 2016) defined value-to-firm and customer to firm's metrics.

RQ5: *What is the relation between value-creation and Big Data Analytics?*

The value-creation is a major sustainability factor for companies, in addition to profit maximization, customer retention, business goals and revenue generation. The adoption of Internet of Things (IoT), big data, and cloud computing technologies by companies/organizations has led to better value-creation at the customer and enterprise ends (Haile & Altmann, 2016). Such as enterprise applications are designed to collect direct customer feedback and information from internal business operations (Verhoef et al., 2016).

Adopting big data analytics as a firm-level innovation aims to achieve firm heterogeneity and hence affords higher value and contribute directly to the overall value-creation performance of banking firms, financial services, supply chain management, and IT companies etc. McKinsey & Co. added ‘Value’ as the fourth ‘V’ to define big data (Chen et al., 2014). Value refers to the worth of hidden insights inside big data. Value represents the transactional, strategic, and informational benefits of big data. Moreover, it represents the extent to which big data generates economically worthy insights and benefits through extraction and transformation (Wamba et al., 2015). Big data has the potential to transform almost every aspect of business from research and development to sales and marketing, and supply-chain management that provide new opportunities for growth.

Stage 3 Big Data Analytics: The process of data collection, processing and analysis for BDA is expected to play a key role in financial & accounting sector. Big Data has become a large pool of unstructured and structured data that can be captured, communicated, aggregated, stored, and analyzed which is now treated as a part of every sector and function of the global economy.

RQ6: *Which are the specific aspects of the data management, data transformation, and utilization drive value for companies?*

In today's digital landscape, data is more readily available and easily gathered, than ever before. With the ability to track most customer interactions, transactions across devices and across channels, companies are looking at, and leveraging, their data in new innovative ways. Dealing with a large amount of information from different data sources is a concern of Big Data management. The issues like how to store, integrate and process data in an effective and efficient way have been pointed out by (Brereton et al., 2007). For storing and processing large datasets companies can use traditional parallel database systems, Apache Hadoop technologies, key-value data stores (Hbase, NoSQL databases) etc. Apache Spark, SparkR, Techyon, MLbase, mahout and Splunk are a relatively new type of analytical tools, which is becoming more and more popular mostly among web companies today.

The process of BDA framework helps to determine the right business model. *(i) Raw data:* The companies which generate a rich pool of raw data can sell it with little investment, *(ii) Processed data:* Processed data comes from multiple sources that are stored, managed and analyzed for others to consume, *(iii) Insights:* Use of data science, predictive modeling, machine learning, and analytics help perform complex correlations on data and gain business insights, *(iv) Presentation*: The ability to present the data, insight and analytic models to key business partners, helps them to build scalable new business ventures.

Related technologies include data management (uncertainty, query processing under near real-time constraints, information extraction), programming models, machine learning, statistical methods, systems architectures, and information visualization. A highly scalable platform support technologies named as Big Data Management System (BDMS) (Borkar, Carey, & Li, 2012). High Performance Computing Cluster Platform (HPCC) enables data integration on a scale which is not previously available and it answers in real-time to millions of users.

Stage 4 Big Data Analytics Capabilities: In addition, BDA capabilities are the key factors throughout the whole process of big data-driven decision-making for useful insights. Big data assets and analytics capabilities are the important building blocks underlying big data value-creation. These capabilities include:

(i) *BDA types:* Descriptive shows “What happened in-to business”, Predictive shows “What is likely to happen in the future”, and Prescriptive shows “So what? Now What?”.
(ii) *BDA Management capabilities:* BDA performs planning, investment, control for planning, budgeting, and forecasting, analyze the performance management, cost management, and internal controls.
(iii) *BDA knowledge capabilities:* Technically business knowledge is needed for the financial & accounting end to end process and understand the technical software knowledge. These big data services are needed to customize for the requirement of individual customers. Although, it is imperative for the accounting service provider to first understand the end to end process of each individual business and then offer a solution.
(iv) *BDA Financial Accounting Capabilities*: Financial services provide planning & analysis, treasuring, cost management, and reporting data about company's financial performance, its current cash flow.

Big Analytics is about turning information into knowledge by using a combination of existing and new approaches. Increasing attention is being paid to big data analytics to extract information, knowledge, and wisdom from big data. However, as model-driven decision-making is still important when failure mechanisms cannot be achieved. Bayesian reliability is a good way to combine model-driven approaches and knowledge-driven approaches. Big data and data science are only useful for insight if they can be turned into an action and if actions are carefully defined and evaluated.

However, it is determined what actions should be taken based on the insight gain. The Fig. 15 shows the various types of decision-making data analytics models (Tavana, 2014).

- *Data Mining* – It extracts the knowledgeable data from organizational data by employing a broad variety of statistical approaches, data visualization, and the pattern recognition approaches.
- *Predictive Modeling* – It is used to recognize organization threats and prospects by developing patterns found in historical and transactional data.
- *Simulation Modeling* – It is used to analyze, compare and scenarios prior to implementation by exploiting application of simulation in enterprise and organizational context.
- *Optimization Modeling* – It is used to define the best decision by means of various optimization models.
- *Prescriptive Methods* – It is used to generate better solutions by exploiting the joint application of predictive models and optimization technology.
- *Business Intelligence*: It is used to increase and sustain a competitive edge by utilizing the modern techniques in data mining, analysis, and performance manage.

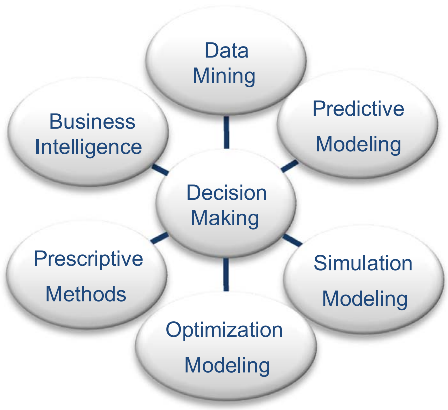

Fig. 15. Various types of decision making data analytics.

7.3. Value realization

The third phase for value realization benefits is the development of big data analytics technology-based solutions to the customer. Value realization from Big Data often requires a strategic transformation. Financial accounting analytics service (FAAS) can support the collection and processing of large amounts of raw accounting data and organize to allow an accurate assessment for identification of an appropriate course of actions. Financial accounting (or financial accountancy) is the field of accounting concern with the summary, analysis, and reporting of financial transactions pertaining to a business. It involves the preparation of financial statements available for public consumption and is a process of analyzing company's asset, liabilities and equality.

For instance, (Price water house Coopers, M. 2015) outlined recommendations for analyzing Big data related to technical competencies in audit, tax, risk management, and consulting. Association to Advance Collegiate Schools of Business International (AACSB) emphasized the importance of integrating Big Data and business analytics into the accounting curriculum. It includes the creation, sharing, and reporting of data, as well as data mining and analytics (AACSB, 2013). The facilitation of better decision-making is identified as one of the greatest benefits of big data. In a survey by Tata Consulting Services, 80% of businesses found that implementing big data initiatives had improved their decision making (Tata Consultancy Services, 2013).

The accounting competencies, "enable accountants to integrate management and analytical methods, supported by technology, to assist an enterprise to formulate and execute its strategy successfully" (Lawson et al., 2013). These competencies are typically taught essential courses such as principles, intermediate, and advanced financial accounting, management and cost accounting, accounting information systems (AIS), auditing, and taxation. It includes outcomes such as knowledge of spread-sheet modeling, use of technology to access Big Data for financial analyzes, use of communication technologies, such as interactive data visualization, knowledge of information systems design and Big Data, including the hardware and software that enable them to run, and related issues, such as computer security and business continuity.

Stage 5 Big Data Analytic-Value Insights: To effectively deliver an analytics capability, business customers need to be assisted with decision-framing, decision-making, and decision-execution. The decision framing is achieved by a clear meaning of the purpose of the decision need and the environment in which the need will be met (through a decision).

(i) *Decision-modeling:* In case of Decision modeling, it refers to modeling technique which is used to design and monitor value improvement roadmaps and analyzes the resulting costs, benefits and risks. In the process of decision modeling, it presents various finding, criteria, scenarios, options, and recommendations.
(ii) *Decision-making:* The decision-making method has two phases: first where the decision model is placed into the business context (the "big picture" model of the business system), and second where the decision model is used for decision-making. For example, decision makers who make the decisions and advisors are "decision coaches" who guide and facilitate the decision-making process.
(iii) *Decision-execution:* It includes user view reports, through a dashboard that provides data as support for business decisions. Such as policies and procedure, system and workflows, training and development, and reports and analytics.
(iv) *Big data business analytics:* It refers to business intelligence, and firm performance as shown in Fig. 14.
(v) *Financial Decision Making/ Value Insights*: The chief financial officer (CFO) who takes responsibility for obtaining and analyzing data and sharing insights with business teams. It creates a need for not just financial reporting, but also wider reporting to external and internal stakeholders. It became far easier for CFOs and their finance teams to evaluate opportunities based on potential, cost reduction, and revenue growth. Finally, financial accounting services present insights that can be achieved by these firms such as:
   - A balance sheet that shows company's position at a certain period.
   - An income statement that details expenses and revenues during a period
   - A statement of cash flows that shows how cash levels have changed.
   - A statement of changes in stockholders' equity that shows how equity has changed.

RQ7: *Can the value of data be monetized, tracked and considered for financial accounting?*

The increased demand for high-level analytic skills creates important opportunities for accountants and finance professionals. The management of big data for accountants and finance professionals means more than 'game-changing' opportunities. New solutions can facilitate the discovery of new business opportunities (including data monetization), challenges, and help the CFO to expand his or her role as trusted advisor to the chief executive officer (CEO) in areas such as strategy development and the identification of the best areas for future investment. It shifts the focus of decision-making from IT (how do we process and store the data) to the CFO and the business units (how do we make the best use of the insights generated from the data). An integrated platform is to gather, cleanse, query, analyze, and visualize data can help to monetize data within and beyond the enterprise. Data-oriented enterprises discover value quickly and better manage information.

Vasarhelyi, Kogan, and Tuttle (2015) provided an overview of big data in accounting focusing on the sources, usages, and challenges of big data in accounting and auditing. (Fanning & Drogt, 2014) pointed out that many firms use big data, e.g. social media data to analyze the target firm's customers and markets in order to make more informed merger and acquisition decisions. In relation to the financial function and management accounting information provision, (Bhimani & Willcocks, 2014) recognized the potential benefits of big data and its complexities in order to deliver potentials. The financial services sector is using big data to improve risk management, detect fraud and track consumer behavior, in order to keep up with compliance standards and increase customer satisfaction and revenues. In accounting and finance area, big data is regarded as an information source that can affect and predict

Table 8
The strategic, operational, business, analytics, financial insights and value driver benefits for BDA.

Perspective	Key interest	Big data Impacts	Research benefits	Value insights
Organisation & strategic	Organisational change and strategic decision making	Data-driven decision, Ecosystem change, Management process	Structure organisation & strategic actions, Business intelligence	Understanding Customer need and Analysis of 360-degree information
Operational analytics	Improving existing operations, efficiency and analytics	Availability data on channels, reduced cost and risk, new innovations	Production, supply of chain and innovation in sensor/ web/social media	Operational innovation in product process, big data supply chain and management
Business & marketing	Consumer & Market effectiveness, Business efficiency	Online evaluation, Better understanding of consumers and market, As fuel to keep investing and building Employees, Customers, stockholders & suppliers	Social influence, Social media/ sensor, Advertising, Consumer behavior and sentiment	User Satisfaction, Business innovation and efficiency, new strategy evaluation, better relationship with customer, Better product and services
Analytics key insights	Advanced analytic techniques for various application	Challenges in processing and analysis of big data management, analysis, and advanced algorithm	Data acquisition, processing, modeling	To monetize existing data, data quality, Speed to insights, Better decision making,
Financial insights	Accounting and finance, Auditing International business	Firm performance, Greater transparency and participation, Better understanding of foreign market, To find and exploit		
New data sources, better financial performance	Financial statement analysis, Investment decisions, To create new revenue streams	Profit & Loss, Profit maximization with short- term focus, Detect fraud & track consumer behavior, Accounting quality, Challenges and risks of big data use in public sector,		

firm's financial performance (Balakrishnan, Qiu, & Srinivasan, 2010). Especially social media has great effects on stock performance (Schniederjans, Cao, & Schniederjans, 2013). Textual data is effectively used to detect financial fraud in reports and improve the financial statements for audit efficiency (Cao, Chychyla, & Stewart, 2015). Although the opportunities and the potential benefits of big data are significant for accounting and financial domain.

The Table 8 shows the strategic benefits, operational benefits, business benefits, analytics key benefits, financial insights and value driver benefits for big data analytics (Sheng et al., 2017). The developed insights and model can create value for firms in five prospective ways:

- *Organization & Strategic:* Customer and firms also need to devote attention to organizing BDA.
- *Operational Analytics:* It provides operational features to run real-time, interactive workloads that ingest and store data.
- *Business & Marketing:* Creating value with BDA making smarter business-driven and marketing decision.
- *Analytics Key Insights:* Analytics focusing on gaining insights and models to improve decision making
- *Financial Insights:* Decision support system for financial and accounting companies.

The key benefits of BDA are to understand customers better, to improve products and services, to improve the management of existing data, to create new revenue streams, to monetize existing data, to become leaner improve internal efficiencies, to find and exploit new data sources, for better management of governance, risk and compliance, to improve the detection and prevention of fraud.

The BDA provides a powerful analytical tool to digital business and marketing managers for understanding the market and customer need in terms of product or services and providing them better decision-making. One of the most important goals of decision-making data analytics is to emerge a forum between academic researchers, policy-makers, and practitioners that concerns with the growth of new approaches to explore and solve organizational issues.

The value insights using BDA help external auditors to better identify financial reporting, fraud and operational business risks, and tailor their approach to deliver a more relevant audit. Audit data analytics methods can be used in audit planning and for improving auditors' knowledge about the transactions and balances underlying the financial statements. Internal auditors act as strategic advisors by analyzing the data to produce actionable information. It can also help to improve accounting data integrity and quality supporting: More accurate and efficient reporting, increased confidence in reporting, Better business and risk insights, Enhanced efficiency and quality of processes and controls, financial statement analysis. Although, financial services firms are focusing their analytics efforts on new opportunities to enhance revenue by mining customer data.

The financial sector worldwide should use big data wisely to overcome the potential global financial crisis and thus, reap the benefits of data analytics in rapidly changing scenario hence, satisfying their customers in the best possible way. Big data will be one of the transformational opportunities of the 21st century. Whether it transforms the finance and accountancy professions for better or for worse depends on how they choose to respond to the challenges it creates.

RQ8: *What are the different challenges of each component of BDA framework?*

Challenge 1- Data Challenges: It is a group of the challenges related to the characteristics of the data. Data Types: Volume (the era of size), Velocity (the era of streaming data), Variety (the era of unstructured data), Veracity (the era of data pollution that needs cleansing), Variability (the era of rapidly changing in data), Valence (the era of connectivity in data), Value (the era of cost associated with data). Data complexity includes the complex types, structure, and pattern of data. Computational complexity refers to processing, analysis, and computation of big data. System complexity is used for high data acquisition throughput, operational efficiency, energy consumption, and low energy consumption (Jin, Wah, Cheng, & Wang, 2015).
Challenge 2- Data Collection: To capture the useful insights, smart filters are required to process the data in an innovative, robust, and intelligent way. Moreover, it will also discard the useless information such as inconsistency or imprecision in data (Sivarajah et al., 2017). For the latter, the most important challenges are rapid data growth, data speed, data variety, and data security prospects.
Challenge 3- Data Storage: BD has transformed the way of acquiring and storing the massive amount of data on bounteous scale keeping in mind the challenge of clustering and storage limitation, high volume of data, and delayed process execution. Data storage challenges are security, integrity, real-time intelligence, reliability, high-efficiency low consumption, and high concurrency. To overcome these challenges, (Buza and Nagy & Nanopoulos 2014) introduced Storage-Optimizing Hierarchical Agglomerative Clustering (SOHAC). Moreover, K-means model is to cope with volume requirements, efficient clustering for large datasets, it is applied in the MR by (Zhao, Ma, & He, 2009). Various researcher introduced data storage replication techniques such as job data scheduling using Bee Colony (JDS-BC) (Taheri, Lee, Zomaya, & Siegel, 2013). In BDA storage, it is challenging to acquire optimized query execution, and indexing process that improves the performance etc.
Challenge 4- Data Pre-processing: The cleaning process is required for further data analysis to deal with various challenges, such as missing data in mobile environments (Zhao & Ng, 2012). Hu et al. (2014) is another data cleaning process for the enhancement resulting from the accurate analysis.
Challenge 5- Data Processing/ Analysis: It is divided into two foremost aim of processing which includes (i) to gain valuable information into the relationship between features and (ii) to develop effective data mining and machine learning approaches that can correctly predict upcoming observations. To overcome these challenges various classification techniques have been proposed such as large-scale machine learning algorithms implemented by use of Hadoop/ Map reduce technologies. Hbase, Cassandra or SimpleDB tools are used for storage purpose of unstructured data.

Challenges 6- Management: It deals with the data privacy, data security, data governance, lack of understanding, cost/operational expenditures, data sharing and analyzing data. Data warehouses store massive amounts of sensitive data such as financial transactions, medical procedures, insurance claims, diagnosis codes, personal data, etc.

In big data analytics, there are some more challenges such as: data integration, data transmission, data scalability, data security and privacy, data processing, data storage, data quality, lack of standardization, and data visualization.

8. Conclusions and research directions

In this article, a theoretically sound framework for value creation by using Big Data Analytics is presented. The objective of this article is to bridge the gap by focusing on value-discover, value-creation, and value-realization by using big data management, processing and advanced analytics. Because, there are numerous challenges for traditional analytics in terms of scalability, adaptability, and usability, presenting new opportunities for inspiring enterprises to adopt BDA for it for decision-making. The answers to RQs used in our paper are as follows:

By considering the first research question, BDA explored the most important seven characteristics namely Volume, Velocity, Variety, Valence, Veracity, Variability, and Value. A large research is being done in defining BDA in terms of Vs for data challenges. BDA has been prospected to raise the economic returns by gaining deeper insights from mountains of existing data.

Our response to second research question has provided an overview of the architecture of BDA-DM framework including six components, namely (i) data generation, (ii) data acquisition, (iii) data storage, (iv) advanced data analytics, (v) data visualization, and (vi) decision-making for value creation.

Our third research question has presented the detailed information of BDA tools, techniques and technologies. This field has received much attention due to its wide application as multi-purpose tools, borrowing techniques from Natural Language Processing (NLP), Data Mining (DM), Machine Learning (ML), Deep Learning (DL) etc. Currently, benchmarking software technologies such as e.g. (Hadoop/Map-Reduce based processing frameworks), NoSQL databases, graph data-bases and analytical frameworks have been developed for BDA.

Our fourth research question has addressed various solutions and helps to select the right combination of different BDA technologies according to their technical needs and requirements. To extract knowledge from Big Data, various models, programs, software's, hardware's and technologies have been designed and proposed in the domain of BDA. Some BDA applications have been discussed to support decision-making in various fields such as healthcare, agriculture, cyber-physical security, smart city and also described various analytics fields such as living analytics, social analytics, mobile analytics, living analytics, video and audio analytics and how they can help to improve performance, usage, and its applicability. Still there are opportunities for new innovations and optimizations in all domains.

Our response to fifth research question has provided the relation between value-creation and Big Data Analytics. It is becoming a key driver of value creation in modern enterprises, where enterprise applications are designed to collect direct customer feedback and information from internal business operations.

Our sixth research question has been about the specific aspects of the data management, transformation and utilization drive value for companies. It is observed that big data management process can collect big data from multiple direct and indirect sources, then data pre-processing and integration operations performed to improve the quality of big data. The statistical methods, machine learning-based, and data mining models are evaluated by using test data and can be deployed in real-time applications like prediction accuracies for financial companies.

Finally, in response to the seventh research question, we have answered how the value of data can be monetized, tracked and considered for financial accounting with the help of using big data analytics as a tool for relevant decision-making. It can also be improved by applying big data analytics and management process (including steps like data storage, ETL process, elaboration and analysis through advanced technologies and new techniques such as data visualization). A survey is presented on how the idea of value-creation is understood and managed by Big Data-based financial companies. It is concerned with the measurement of the results of business activities and with the preparation and use of financial reports such as the balance sheet, income statement, and the analysis of financial reports. The concept regarding the collection, processing, and reporting of financial information in a technology-based business is presented.

We have addressed significant and various different challenges of each component which are beyond the data types such as data collection, data storage, data pre-processing which includes:- inauthentic data collection, data quality and validation, data cleaning, feature engineering, high dimensionality, data reduction, data representation, distributed data sources, data sampling. Data analysis includes data inconsistency and incompleteness, scalability, timeliness, data security, data management scalability of algorithm and data visualization.

Towards this end, we have reviewed not only the latest research papers but also relevant research articles that have been published over the last few decades. This paper discussed the concept of Big Data Analytics, BDA-Management, and BDA-Machine learning articles. The Table 9 shows the list of emerging research technologies & their domain of interest. Following are possible directions for future research in this area:

- Previously, till date machine learning in big data analytics focused on the 3 V's (volume, velocity, and variety), but no one worked on the remaining four aspects of big data (veracity, valence, variability, and value).
- It can also be explored the other challenges to overcome data inconsistency, incompleteness, scalability, timeliness, and security.
- Protection and security is the most important aspect for making the right decision by using Big Data and Analytics technology.

Table 9
list of emerging research technologies & their domain of interest.

Future technologies	Field of interest	Domain
Cloud computing	-Virtualization -Software-defined networking -Resource Management	-Big Data as a Services (BDAAS) -Data Analytics as a Service (DAAS)
Artificial intelligence	-Optimization -Neural networks -Big data mining -IoT mining	-Help to make intelligent devices
Bio-inspired computing	-Immune systems -Linder Mayer systems -Membrane computers	-Insect Monitoring Drones
Quantum Computing	-Electronic computing -Optical computing, -Quantum clock	-Quantum Computational Models -Quantum Machine Learning Models/ Artificial Intelligence -Dynamic Quantum Clustering -Improved Optimization, -Better search, -Increased Security, -Pattern Recognition, chemical modeling
Quantum cryptography	-Public-key encryption -Signature schemes	-Helps in performing cryptographic tasks. -Image processing -Advanced Security
Edge Computing	-Fog computing -Mobile edge computing -Cloudlet	-Facilitates the users by bringing computation down towards the edge of the network.
Big Data Analytics & Machine learning	-Classification & Clustering -Predictive modeling -Real-time analytics -Deep learning	-Financial & Accounting -Agriculture IoT -Smart Transportation -Smart Healthcare -Smart City -Smart Grid -Smart Inventory System -Smart Parking

- There have been very few researchers making comparative researches and systematically discussing the advantages and disadvantages of financial risk.
- There is a need for initiatives that will address the integration with machine learning for big data processing.
- There is a need for empirical research on a comprehensive basis for studying qualitative and quantitative approach.

In future, we expect to improve the research in two aspects (1) by providing a more specialized review of frameworks and programming models/algorithms for BDA to extend the information about the features, advantages, and limitations, (2) by considering papers from other scientific databases.

Appendix A. Overview of machine learning tool for big data analytics

Tool name	Specified uses	Language	Algorithm	Advantages	References
Apache Spark	Apache Spark is a powerful open source processing engine built around speed, ease of use, and sophisticated analytics. It was originally developed at UC Berkeley in 2009.	Scala Java Python R	-Classification Linear models Navi Bayes Decision tree -Regression Least Square Decision tree -Clustering Kmeans Mixture models Hierarchical	Faster batch processing, flexible and powerful, iterative processing, Sophisticated Analytics, Real Time Stream Processing	(Apache Spark, 2014)

			-Collaborative filtering		
SparkR	SparkR is an R package that provides a light-weight frontend to use Apache Spark from R. In Spark 2.0.2, SparkR provides a distributed data frame implementation that supports operations like selection, filtering, aggregation etc.	R Java	-Generalized Linear Model -Accelerated Failure Time-Survival -Regression Model -Naive Bayes Mode -KMeans Model	Fast Scalable Expressive Numerical Packages Interactive	(Apache SparkR, 2017)
Rhadoop	RHadoop is a collection of three R packages that allow users to manage and analyze data with Hadoop.	R Hadoop	-Linear regression -Clustering -Classification	Revolutionary analytics	(Prajapati, 2013)
Apache Mahout	Mahout is an open source machine learning library from Apache.	Java Scala Maven Hadoop Python R	-Clustering -Classification -Collaborating filtering -Regression -Frequent pattern mining -Genetic algorithm	High performance ML applications Scalable	(Apache Mahout, 2011)
H_2O	H2O is an open source, in-memory, distributed, fast, and scalable machine learning and predictive analytics platform that allows you to build machine learning models on big data and provides easy productionalization of those models in an enterprise environment.	Java R Python Scala on Hadoop/ Yarn	-Regression -Classification -Clustering -Deep learning	Math platform Easy to use Better predictions	(H20, 2011)
R	R is an open-source software programming language for statistical analysis and graphics R is extremely extensible. With huge developer base, thousands of R packages are available to provide variety of functionalities.	Java C, C++ Fortan Python Perl Ruby Julia	-Collaborative filtering -Classification -Clustering -Regression	Numerical Packages Interactive	(R,1993)
MOA	Massive Online Analysis is the most popular open source framework for data stream mining, with a very active growing community (blog). MOA is an open-source framework software that allows to build and run experiments of machine learning or data mining on evolving data streams.	Java Scala	-Classification -Regression -Clustering -Outlier detection -Frequent pattern mining	Easy extensible for data feeding Algorithms Evaluation measures	(Apache MOA, 2011)
Vowpal Wabbit	Vowpal Wabbit is an open source fast out-of-core learning system library and program developed originally at Yahoo! Research, and currently at Microsoft Research. It was started and is led by John Langford.	Python C# Java	-Classification (both binary and multi-class) -Regression -Active learning -Multiple learning algorithms -Multiple	Fast Online learning Scalable solutions	(Vowal Wabbit, 2014)

			optimization Algorithms		
Giraph	Giraph is an open-source project and external contributions are extremely appreciated. It is iterative graph processing system built for high scalability.	Java Maven Hadoop	-Graph analytics -Clustering -Collaborative filtering	Fast Scalable graph processing	(Apache Giraph, 2012)
RHipe	R and Hadoop Integrated Programming Environment is an Rpackage that provides an environment for the R programmer to compute with very large data sets across a Hadoop cluster.	Java R Hadoop Maven Ant	-Statistical analysis -Data modeling	Statistical computing	(Rhipe, 2003)

References

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), 1–33.

ACM, Digital Library. http://www.acm.org/dl/ Accessed 30 Dec 2017.

Addo-Tenkorang, R., & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering, 101*, 528–543.

Agrawal, S., & Agrawal, J. (2015). Neural network techniques for cancer prediction: a survey. *Procedia Computer Science, 60*, 769–774.

Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., et al. (2017). The role of big data analytics in Internet of Things. *Computer Networks, 129*, 459–471.

Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research, 9*, 1981–2014.

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), 25.

Alowibdi, J. S., Buy, U. A., Philip, S. Y., & Stenneth, L. (2014). Detecting deception in online social networks. *Proceedings of IEEE/ACM international conference in advances in social networks analysis and mining (ASONAM)* (pp. 383–390). .

Alsheikh, M. A., Niyato, D., Lin, S., Tan, H. P., & Han, Z. (2016). Mobile big data analytics using deep learning and apache spark. *IEEE Network, 30*(3), 22–29.

An, L., & Bhanu, B. (2012). Image super-resolution by extreme learning machine. *Proceedings of the 19th IEEE international conference on image processing* (pp. 2209–2212). IEEE.

Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science, 50*, 408–413.

Armstrong, G., Adam, S., Denize, S., & Kotler, P. (2014). *Principles of marketing.* Australia: Pearson.

Association to Advance Collegiate Schools of Business International (AACSB). (2013). *Eligibility procedures and accreditation standards for accounting accreditation.* Tampa, FL: AACSB http://www.aacsb.edu/accreditation/standards/2013-business Accessed 12 Nov 2016.

Assuncao, 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.

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings of the international conference on IEEE/WIC/ACM web intelligence and intelligent agent technology (WI-IAT). 1. Proceedings of the international conference on IEEE/WIC/ACM web intelligence and intelligent agent technology (WI-IAT)* (pp. 492–499). IEEE.

Apache Spark, 2014. https://spark.apache.org/ Accessed 20 Dec 2017.

Apache SparkR. https://docs.databricks.com/spark/latest/sparkr/overview.html/ Accessed 20 Dec 2017.

Apache MapReduce, 2008. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html/ Accessed 20 Dec 2017.

Apache Mahout, 2011. https://mahout.apache.org/ Accessed 20 Dec 2017.

Apache MOA, 2011. https://en.wikipedia.org/wiki/Massive_Online_Analysis/ Accessed 20 Dec 2017.

Apache Giraph. http://giraph.apache.org/ Accessed 20 Dec 2017.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Balakrishnan, R., Qiu, X. Y., & Srinivasan, P. (2010). On the predictive ability of narrative disclosures in annual reports. *European Journal of Operational Research, 202*(3), 789–801.

Balkan, S., & Kholod, T. (2015). Video analytics in market research. *Information Systems Management, 32*(3), 192–199.

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.

Bashir, M. R., & Gill, A. Q. (2016). Towards an IoT Big Data Analytics Framework: Smart Buildings Systems. *Proceedings of the IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS), 2016 IEEE 18th international conference on high performance computing and communications* (pp. 1325–1332). .

Basu, A. (2013). Five pillars of prescriptive analytics success. *Analytics magazine,* 8–12 http://analytics-magazine.org/executive-edge-five-pillars-of-prescriptive-analytics-success/. Accessed 20 Dec 2017.

Batarseh, F. A., & Latif, E. A. (2016). Assessing the quality of service using big data analytics: With application to healthcare. *Big Data Research, 4*, 13–24.

Becker, T. (2016). *Big data usage. New horizons for a data-driven economy*. Cham: Springer143–165 (Chapter 8).

Bendre, M. R., Thool, R. C., & Thool, V. R. (2016). Big data in precision agriculture through ICT: Rainfall prediction using neural network approach. *Proceedings of the international congress on information and communication technology* (pp. 165–175). Springer.

Bengio, Y. (2013). Deep learning of representations: Looking forward. *Proceedings of international conference on statistical language and speech processing* (pp. 1–37). Springer.

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

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., et al. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics, 30*(3), 500–521.

Bloor, R. (2011). Big data analytics – this time it's personal. *The Bloor Group*. https://www.1010data.com/media/1073/1010data_bloorgroup_whitepaper_big_data_analytics.pdf/ Accessed 14 April 2016.

Borkar, V. R., Carey, M. J., & Li, C. (2012). Big data platforms: What's next? *XRDS: Crossroads, The ACM Magazine for Students, 19*(1), 44–49.

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software, 80*(4), 571–583.

Butcher, J. B., Verstraeten, D., Schrauwen, B., Day, C. R., & Haycock, P. W. (2013). Reservoir computing and extreme learning machines for non-linear time-series data analysis. *Neural Networks, 38*, 76–89.

Buyya, R., Calheiros, R. N., & Dastjerdi, A. V. (2016). *Big Data: principles and paradigms.* Morgan Kaufmann.
Buza, K., Nagy, G. I., & Nanopoulos, A. (2014). Storage-optimizing clustering algorithms for high-dimensional tick data. *Expert Systems with Applications, 41*(9), 4148–4157.
Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons, 29*(2), 423–429.
Casado, R., & Younas, M. (2015). Emerging trends and technologies in big data processing. *Concurrency and Computation: Practice and Experience, 27*(8), 2078–2091.
Castillo, C. (2005). *Effective web crawling. Acmsigir forum, 39*, Centro de Operações Rio: ACM55–56. http://centrodeoperacoes.rio/ Accessed 05 October 2016.
Cerchiello, P., & Giudici, P. (2016). Big data analysis for financial risk management. *Journal of Big Data, 3*(1), 1–12.
Cetin, G., Demirçiftçi, T., & Bilgihan, A. (2016). Meeting revenue management challenges: Knowledge, skills and abilities. *International Journal of Hospitality Management, 57*, 132–142.
Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR), 41*(3), 1–15.
Chaturvedi, R., & Ezeife, C. I. (2012). Data mining techniques for design of its student models. *Proceedings of the 5th international conference on educational data mining* (pp. 218–219). .
Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 275*, 314–347.
Chen, H. M., Kazman, R., & Haziyev, S. (2016). Agile big data analytics for web-based systems: An architecture-centric approach. *IEEE Transactions on Big Data, 2*(3), 234–248.
Chen, H., Chiang, R. H., & Storey, V. C. (2012). 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., et al. (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.
Chen, X. W., & Lin, X. (2014). Big data deep learning: Challenges and perspectives. *IEEE access, 2*, 514–525.
Cheptsov, A., & Koller, B. (2015). Leveraging high-performance computing infrastructures to web data analytic applications by means of message-passing interface. *Proceedings of modeling and processing for next-generation big-data technologies*. *4*. *Proceedings of modeling and processing for next-generation big-data technologies* (pp. 167–185).
Chung, W. (2014). BizPro: Extracting and categorizing business intelligence factors from textual news articles. *International Journal of Information Management, 34*(2), 272–284.
Colak, I., Sagiroglu, S., & Yesilbudak, M. (2012). Data mining and wind power prediction: A literature review. *Renewable Energy, 46*, 241–247.
Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings of the 8th conference on Visualization'97*IEEE 235-ff.
Crawley, M., & Wahlen, J. (2014). Analytics in empirical/archival financial accounting research. *Business Horizons, 57*(5), 583–593.
De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review, 65*(3), 122–135.
Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM, 51*(1), 107–113.
Demirkan, H. (2013). A smart healthcare systems framework. *It Professional, 15*(5), 38–45.
Ding, S., Zhao, H., Zhang, Y., Xu, X., & Nie, R. (2015). Extreme learning machine: Algorithm, theory and applications. *Artificial Intelligence Review, 44*(1), 103–115.
Dolin, R. H., Rogers, B., & Jaffe, C. (2015). Health level seven interoperability strategy: Big data, incrementally structured. *Methods of information in medicine, 54*(1), 75–82.
Dong, J., Chen, Q., Yan, S., & Yuille, A. (2014). Towards unified object detection and semantic segmentation. *Proceedings of European conference on computer vision*. *8693*. *Proceedings of European conference on computer vision* (pp. 299–314). Springer.
Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences, 106*(36), 15274–15278.
Edwards, J. S., & Taborda, E. R. (2016). Using knowledge management to give context to Analytics and big data and reduce strategic risk. *Procedia Computer Science, 99*, 36–49.
Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., et al. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing, 2*(3), 267–279.
Fanning, K., & Drogt, E. (2014). Big data: New opportunities for M&A. *Journal of Corporate Accounting & Finance, 25*(2), 27–34.
Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37.
Fiosina, J., Fiosins, M., & Müller, J. P. (2013). Big data processing and mining for next generation intelligent transportation systems. *Journal Teknologi, 63*(3), 21–38.
Fogelman-Soulié, F., & Lu, W. (2016). Implementing big data analytics projects in business. *Proceedings of big data analysis: new algorithms for a new society* (pp. 141–158). .
Francik, S., Ślipek, Z., Frączek, J., & Knapczyk, A. (2016). Present trends in research on application of artificial neural networks in agricultural engineering. *Agricultural Engineering, 20*(4), 15–25.
Fulantelli, G., Taibi, D., & Arrigo, M. (2013). A semantic approach to mobile learning analytics. *Proceedings of the first international conference on technological ecosystem for enhancing multiculturality* (pp. 287–292). ACM.
Gamble, M., & Goble, C. (2011). Quality, trust, and utility of scientific data on the web: Towards a joint model. *Proceedings of the 3rd international web science conference* (pp. 1–15). ACM.
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. (2011). Extracting value from chaos. *IDC iview, 1142*, 1–12. https://www.emcgrandprix.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf/ Accessed 12 Feb 2016.
Gartner, Inc., (2011). Pattern-based strategy: getting value from big data. http://www.gartner.com/it/page.jsp?id=1731916 Accessed 12 Feb 2016.
Google Scholar, Digital library. http://www.scholar.google.co.in/ Accessed 20 Dec 2017.
Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems, 29*(7), 1645–1660.
Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing, 187*, 27–48.
Hadoop, A. (2011). Apache hadoop. http://omeroztat.com/docs/Hadoop.pdf/ Accessed 12 Nov 2017.
Hagel, J. (2015). Bringing analytics to life. *Journal of Accountancy, 219*, 24–25.
Hagstrom, M. (2012). High-performance analytics fuels innovation and inclusive growth: Use big data, hyper connectivity and speed to intelligence to get true value in the digital economy. *Journal of Advanced Analytics, 2*, 3–4.
Haile, N., & Altmann, J. (2016). Value creation in software service platforms. *Futur Generation Computer Systems, 55*, 495–509.
Hänel, T., & Felden, C. (2015). Linking operational business intelligence with value-based business requirements. *Proceedings of conference on e-Business, e-Services and e-Society*. *9373*. *Proceedings of conference on e-Business, e-Services and e-Society* (pp. 147–159). Springer.
Hayes, P. J., & Carbonell, J. G. (1983). A tutorial on techniques and applications for natural language processing, 1–14. http://repository.cmu.edu/cgi/viewcontent.cgi?article=2483&context=compsci/ Accessed 12 Nov 2017.
Hazen, B. T., Skipper, J. B., Ezell, J. D., & Boone, C. A. (2016). Big Data and predictive analytics for supply chain sustainability y: A theory-driven research agenda. *Computers & Industrial Engineering, 101*, 592–598.
He, B., Xu, D., Nian, R., van Heeswijk, M., Yu, Q., Miche, Y., et al. (2014). Fast face recognition via sparse coding and extreme learning machine. *Cognitive Computation, 6*(2), 264–277.
Himmelblau, D. M. (2000). Applications of artificial neural networks in chemical engineering. *Korean journal of chemical engineering, 17*(4), 373–392.
Hirschberg, J., Hjalmarsson, A., & Elhadad, N. (2010). You're as Sick as You Sound": Using computational approaches for modeling speaker state to gauge illness and recovery. *Proceedings of advances in speech recognition* (pp. 305–322). .
Hoffman, J., Guadarrama, S., Tzeng, E. S., Hu, R., Donahue, J., Girshick, R., et al. (2014). LSDA: Large scale detection through adaptation. *Proceedings of advances in neural information processing systems* (pp. 3536–3544). .
Hortonworks, (2011). http://hortonworks.com/products/hortonworks-sandbox/ Accessed 12 Nov 2017.

HP-HAVEn, (2013). https://en.wikipedia.org/wiki/HP_Autonomy/ Accessed 12 Nov 2017.
Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access, 2*, 652–687.
Hu, Y., Cai, X., & DuPont, B. (2015). Design of a web-based application of the coupled multi-agent system model and environmental model for watershed management analysis using Hadoop. *Environmental Modeling & Software, 70*, 149–162.
Huang, G. B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: A survey. *International Journal of Machine Learning and Cybernetics, 2*(2), 107–122.
Huang, G., Huang, G. B., Song, S., & You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks, 61*, 32–48.
H2O. Big Data, 2011. https://www.h2o.ai/ Accessed 12 Dec 2017.
IEEE Xplore. Digital Library. http://www.ieeexplore.ieee.org/ Accessed 30 Dec 2017.
Infobright, (2005). http://www.infobright.com/ Accessed 20 Dec 2017.
Injadat, M., Salo, F., & Nassif, A. B. (2016). Data mining techniques in social media: A survey. *Neurocomputing, 214*, 654–670.
Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2017). Big data analytics and computational intelligence for cyber–physical systems: Recent trends and state of the art applications. *Future Generation Computer Systems,* 1–13.
Jabbour, C. J. C. (2013). Environmental training in organisations: From a literature review to a framework for future research. *Resources, Conservation and Recycling, 74*, 144–155.
Jha, G. K. (2007). Artificial neural networks and its applications. *IARI, New Delhi,* 1–9 http://bioinformatics.iasri.res.in/BAMAST/Book.html/EbookNew/web/book/module3/ANNanditsApplications.pdf. Accessed 12 March 2016.
Jiang, S., Chen, T. E., Dong, J., & Wang, C. (2013). Applications and implementation of decomposition storage model (DSM) in paas of agricultural. *Proceedings of international conference on computer and computing technologies in agriculture*. *420*. *Proceedings of international conference on computer and computing technologies in agriculture* (pp. 34–41). Springer.
Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research, 2*(2), 59–64.
Jukic, N., Sharma, A., Nestorov, S., & Jukic, B. (2015). Augmenting data warehouses with Big Data. *Information Systems Management, 32*(3), 200–209.
Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing, 74*(7), 2561–2573.
Khan, S., Liu, X., Shakil, K. A., & Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing & Management, 53*(4), 923–944.
Khan, Z., Anjum, A., Soomro, K., & Tahir, M. A. (2015). Towards cloud based big data analytics for smart future cities. *Journal of Cloud Computing, 4*(1), 2.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Proceedings of advances in neural information processing systems* (pp. 1097–1105). .
Kshetri, N. (2014). The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society, 1*(2), 1–20.
Kshetri, N. (2016). Big data's role in expanding access to financial services in China. *International Journal of Information Management, 36*(3), 297–308.
Laurila, J. K., Gatica-Perez, D., Aad, I., Bornet, O., Do, T. M. T., Dousse, O., et al. (2012). The mobile data challenge: Big data for mobile computing research. *Proceedings of pervasive computing* (pp. 1–8). .
Lawson, R. A., Blocher, E. J., Brewer, P. C., Cokins, G., Sorensen, J. E., Stout, D. E., et al. (2013). Focusing accounting curricula on students' long-run careers: Recommendations for an integrated competency-based framework for accounting education. *Issues in Accounting Education, 29*(2), 295–317.
Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Life in the network: The coming age of computational social science. *Science (New York, NY), 323*(5915), 721.
Leavitt, N. (2013). Bringing big analytics to the masses. *Computer, 46*(1), 20–23.
Lee, J., Bagheri, B., & Kao, H. A. (2014). Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. *Proceeding of international conference on industrial informatics (INDIN)* (pp. 1–6). .
Lee, J., Jin, C., & Liu, Z. (2017). Predictive big data analytics and cyber physical systems for TES systems. *Advances in Through-life Engineering Services* (pp. 97–112). Cham: Springer (Chapter 7).
Lee, M., Hwang, J., & Yoe, H. (2013). Agricultural production system based on IOT. *Proceedings of IEEE 16th international conference on computational science and engineering (CSE)* (pp. 833–837). IEEE.
Lee, S., & Xing, E. P. (2012). Leveraging input and output structures for joint mapping of epistatic and marginal eQTLs. *Bioinformatics, 28*(12), i137–i146.
Li, B., & Jain, R. (2013). *Survey of recent research progress and issues in big data.* USA: Washington University in St. Louis1–13.
Li, Y., Chen, W., Wang, Y., & Zhang, Z. L. (2013). Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. *Proceedings of the 56th ACM international conference on Web search and data mining* (pp. 657–666). ACM.
Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications, 39*(12), 11303–11311.
Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing, 115*, 134–142.
Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing, 234*, 11–26.
Liu, Y., Guo, Y., Wu, S., & Lew, M. S. (2015). Deep index for accurate and efficient image retrieval. *Proceedings of the 5th ACM on international conference on multimedia retrieval* (pp. 43–50). ACM.
Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: Application of big data. *Journal of Big Data, 4*(1), 20.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburg, C., et al. (2011). *Big data: The next frontier for innovation, competition and productivity*.
Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., et al. (2017). Big IoT data analytics: Architecture, opportunities, and open research challenges. *IEEE Access, 5*, 5247–5261.
Marquez, L. (2000). Machine learning and natural language processing. *Proceedings of complementary documentation for the conference "Aprendizaje automático aplicado al procesamiento del lenguaje natural*Soria Technical Report,.
Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A data quality in use model for big data. *Future Generation Computer Systems, 63*, 123–130.
Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management,* 1–32.
Mithas, S., & Lucas, H. C. (2010). What is your digital business strategy? *IT professional, 12*(6), 4–6.
Moss, L., Corsar, D., & Piper, I. (2012). A linked data approach to assessing medical data. *Proceedings of IEEE 25th international symposium on computer-based medical systems (CBMS), 2012 IEEE* (pp. 1–4). .
Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), 1–21.
Neocleous, C., & Schizas, C. (2002). Artificial neural network learning: A comparative review, Methods and Applications of Artificial Intelligence. *Hellenic Conference on Artificial Intelligence 2308* (pp. 300–313). Berlin, Heidelberg: Springer.
Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications, 36*(2), 2592–2602.
Nino, M., & Illarramendi, A. (2015). Entendiendo el big data: Antecedentes, Origen Y Desarrollo Posterior. *DYNA New Technologies, 2*(1), 1–8.
Oracle. (2012). *Big data for the enterprise.* Redwood Shores: Oracle http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf/ Accessed 30 Dec 2017.
Ortiz-Rangel, M., Rueda-Vasquez, L., Duarte-Gualdron, C., Petit, J., & Ordóñez-Plata, G. (2015). Towards a smart city: Design of a domestic smart grid. *Proceedings of innovative smart grid technologies Latin America (ISGT LATAM)* (pp. 863–868). .
Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2017). Big data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences,* 1–18.
Ouyang, W., Chu, X., & Wang, X. (2014). Multi-source deep learning for human pose estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2329–2336). IEEE.

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.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications, 41*(4), 1432–1462.

Phillips, J., & Buchanan, B. G. (2001). Ontology-guided knowledge discovery in databases. *Proceedings of the 1st international conference on Knowledge capture* (pp. 123–130). ACM.

Pivotal big data suite, (2016). http://pivotal.io/big-data/pivotal-big-data-suite/ Accessed 30 Dec 2017.

Pole, G., & Gera, P. (2016). A recent study of emerging tools and technologies boosting big data analytics. *Proceedings of innovations in computer science and engineering* (pp. 29–36). .

PricewaterhouseCoopers, (2015). Data driven: What students need to succeed in a rapidly changing business world. White Paper. https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf/ Accessed 23 May 2016.

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing, 2016*(1), 1–16.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems, 2*(1), 3.

Ramannavar, M., & Sidnal, N. S. (2016). Big Data and analytics—A Journey through basic concepts to research issues. *Proceedings of the international conference on soft computing systems*. *398*. *Proceedings of the international conference on soft computing systems* (pp. 291–306). Springer.

Rathore, M. M., Ahmad, A., & Paul, A. (2016). IoT-based smart city development using big data analytical approach. *Proceedings of IEEE international conference on automatica (ICA-ACCA),* (pp. 1–8). ACM.

Roy, N., Pallapa, G., & Das, S. K. (2007). A middleware framework for ambiguous context mediation in smart healthcare application. *Proceedings of third IEEE international conference on wireless and mobile computing, networking and communications, 2007. WiMOB 2007.*IEEE 72-72.

Rudolf, M., Paradies, M., Bornhövd, C., & Lehner, W. (2013). The graph story of the SAP HANA database. *In BTW, 13*, 403–420.

Rhipe, 2003. https://www.rhipe.com/ Accessed 30 Dec 2017.

Saidi, R., Maddouri, M., & Nguifo, E. M. (2010). Protein sequences classification by means of feature extraction with substitution matrices. *BMC bioinformatics, 11*(1), 175.

Schniederjans, D., Cao, E., & Schniederjans, M. (2013). Enhancing financial performance with social media: An impression management perspective. *Decision Support Systems, 55*(4), 911–918.

Science Direct, Digital library. www.sciencedirect.com/ Accessed 30 Dec 2017.

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.

Serrato, M., & Ramirez, J. (2017). *The strategic business value of big data. Big Data Management*. Cham: Springer47–70 (Chapter 3).

Sezer, OmerBerat, Dogdu, Erdogan, Ozbayoglu, Murat, & Onal, Aras (2016). An extended IoT framework with semantics, big data, and analytics. *Proceedings of IEEE international conference big data (Big Data)* (pp. 1849–1856). IEEE.

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics, 191*, 97–112.

Shin, D. H. (2016). Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy, 40*(9), 837–854.

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., et al. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications, 71*, 151–166.

Siegal, E. (2010). *Seven reasons you need predictive analytics today.* San Francisco, CA: Prediction Impact Inc683–1146 (415).

Simon, P. (2014). *The visual organization: data visualization, big Data, and the quest for better decisions.* Hoboken: Wiley.

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.

Sledgianowski, D., Gomaa, M., & Tan, C. (2017). Toward integration of Big Data, technology and information systems competencies into the accounting curriculum. *Journal of Accounting Education, 38*, 81–93.

Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. *Proceedings of North American power symposium (NAPS)* (pp. 1–8). .

Song, H. (2014). Introduction to data visualization: history, concept, methods. https://www.slideshare.net/SookyoungSong/hci-tutorial0212?qid=d736cd0f-b704-4b4f-82ea-7dead9b04e9bandv=andb=andfrom_search=9/ Accessed 24 June 2017.

Sovilj, D., Sorjamaa, A., Yu, Q., Miche, Y., & Severin, E. (2010). OPELM and OPKNN in long-term prediction of time series using projected input data. *Neurocomputing, 73*(10–12), 1976–1986.

Springer, Digital library. http://www.springerlink.com/ Accessed 30 Dec 2017.

Srinivasan, U., & Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs. *IT Professional, 15*(6), 21–28.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1998). Market-based assets and shareholder value: A framework for analysis. *The Journal of Marketing, 62*(1), 2–18.

Strawn, G. O. (2012). Scientific research: How many paradigms. *Educause Review, 47*(3), 1–26.

Strohbach, M., Ziekow, H., Gazis, V., & Akiva, N. (2015). Towards a big data analytics framework for IoT and smart city applications. *Modeling and processing for next-generation big-data technologies* (pp. 257–282). .

Sun, E. W., Chen, Y. T., & Yu, M. T. (2015). Generalized optimal wavelet decomposing algorithm for big financial data. *International Journal of Production Economics, 165*, 194–214.

Suthaharan, S. (2014). Big data classification: Problems and challenges in network intrusion prediction with machine learning. *ACM SIGMETRICS Performance Evaluation Review, 41*(4), 70–73.

Taheri, J., Lee, Y. C., Zomaya, A. Y., & Siegel, H. J. (2013). A Bee Colony based optimization approach for simultaneous job scheduling and data replication in grid environments. *Computers & Operations Research, 40*(6), 1564–1578.

Tata Consultancy Services, (2013). The emerging big returns on big data: A TCS 2013 Global Trend Study. http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf/ Accessed 03 April 2017.

Tavana: Editorial decision analytics, 2014, 1:1. http://www.decisionanalyticsjournal.com/1/1/1/ Accessed 20 Sep 2016.

Tian, H. X., & Mao, Z. Z. (2010). An ensemble ELM based on modified AdaBoost. RT algorithm for predicting the temperature of molten steel in ladle furnace. *IEEE Transactions on Automation Science and Engineering, 7*(1), 73–80.

Tian, X., Han, R., Wang, L., Lu, G., & Zhan, J. (2015). Latency critical Big data computing in finance. *The Journal of Finance and Data Science, 1*(1), 33–41.

Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing, 38*, 788–804.

Tolk, A. (2015). The next generation of modeling & simulation: Integrating big data and deep learning. *Proceedings of the conference on summer computer simulation, society for computer simulation international* (pp. 1–8). ACM.

Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data, 2*(1), 1–21.

Tu, C., He, X., Shuai, Z., & Jiang, F. (2017). Big data issues in smart grid–A review. *Renewable and Sustainable Energy Reviews, 79*, 1099–1107.

urRehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management, 36*(6), 917–928.

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

Vatrapu, R., Mukkamala, R. R., Hussain, A., & Flesch, B. (2016). Social set analysis: A set theoretical approach to big data analytics. *IEEE Access, 4*, 2542–2571.

Vavpetic, A., Novak, P. K., Grcar, M., Mozetic, I., & Lavrac, N. (2013). Semantic data mining of financial news articles. *Proceedings of international conference on discovery science* (pp. 294–307). Springer.

Verhoef, P. C., Kooge, E., & Walk, N. (2016). *Creating value with big data analytics: making smarter marketing decisions.* Routledge http://aksitha.com/Big%20Data%20Technologies/Big%20Data/Creating%20Value%20with%20Big%20Data%20Analytics%20Making%20Smarter%20Marketing%20Decisions.pdf. Accessed 12 March 2017.

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC, 14*, 1–14 http://www.tracemyflows.com/uploads/big_data/idc_amd_big_data_whitepaper.pdf/ Accessed 12 March 2016.
Vowal Wabbit, Big Data Software. https://en.wikipedia.org/wiki/Vowpal_Wabbit/ Accessed 20 Nov 2017.
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.
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.
Wang, Y., Kung, L., & Byrd, T. A. (2016). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change, 126*, 3–13.
Ward, M. J., Marsolo, K. A., & Froehle, C. M. (2014). Applications of business analytics in healthcare. *Business horizons, 57*(5), 571–582.
Weng, C. H., Huang, T. C. K., & Han, R. P. (2016). Disease prediction with different types of neural network classifiers. *Telematics and Informatics, 33*(2), 277–292.
Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in smart farming–A review. *Agricultural Systems, 153*, 69–80.
World Bank. (2014). *Central America: big data in action for development.* Washington, DC: World Bank https://openknowledge.worldbank.org/handle/10986/21325/ Accessed 04 March 2017.
Wu, C., Buyya, R., & Ramamohanarao, K. (2016a). *Big Data analytics= machine learning+ cloud computing* https://arxiv.org/abs/1601.03115
Wu, J., Li, H., Cheng, S., & Lin, Z. (2016b). The promising future of healthcare services: When big data analytics meets wearable technology. *Information & Management, 53*(8), 1020–1033.
Xie, N. F., Zhang, X. F., Sun, W., & Hao, X. N. (2015). *Research on big data technology-based agricultural information system. International conference on computer information systems and industrial applications*. Atlantis Press388–390.
Xing, E. P., Ho, Q., Xie, P., & Wei, D. (2016). Strategies and principles of distributed machine learning on big data. *Engineering, 2*(2), 179–195.
Xu, Z., Mei, L., Hu, C., & Liu, Y. (2016). The big data analytics and applications of the surveillance system using video structured description technology. *Cluster Computing, 19*(3), 1283–1292.
Yang, S., Zhong, Y., Liu, R., & Feng, Z. (2014). Framework formation of financial data classification standard in the era of the big data. *Procedia Computer Science, 30*, 88–96.
Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., et al. (2016). Big data: From beginning to future. *International Journal of Information Management, 36*(6), 1231–1247.
Yazti, D. Z., & Krishnaswamy, S. (2014). Mobile big data analytics: Research, practice, and opportunities. *Proceedings of IEEE 15th international conference on mobile data management (MDM)*. *1. Proceedings of IEEE 15th international conference on mobile data management (MDM)* (pp. 1–2). IEEE.
You, Z. H., Lei, Y. K., Zhu, L., Xia, J. F., & Wang, B. (2013). Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. *BMC Bioinformatics, 14*, 1–11.
Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(4), 451–462 Part C (Applications and Reviews).
Zhang, X., & Xiang, S. (2015). *Data quality, analytics, and privacy in big data. Big data in complex systems*. Cham: Springer393–418.
Zhang, Y., Ren, S., Liu, Y., Sakao, T., & Huisingh, D. (2017). A framework for Big Data driven product lifecycle management. *Journal of Cleaner Production, 159*, 229–240.
Zhao, B., & Xing, E. P. (2014). Quasi real-time summarization for consumer videos. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2513–2520). IEEE.
Zhao, W., Ma, H., & He, Q. (2009). Parallel k-means clustering based on mapreduce. *IEEE international conference on cloud computing* (pp. 674–679). IEEE.
Zhao, Z., & Ng, W. (2012). A model-based approach for RFID data stream cleansing. *Proceedings of the 21st ACM international conference on information and knowledge management* (pp. 862–871). ACM.
Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews, 56*, 215–225.
Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on Big Data: Opportunities and challenges. *Neurocomputing, 237*, 350–361.
Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: analytics for enterprise class hadoop and streaming data.* McGraw-Hill Osborne Media.
Zuker, M., & Sankoff, D. (1984). RNA secondary structures and their prediction. *Bulletin of mathematical biology, 46*(4), 591–621.