Big Data: Analytical Solutions, Research Challenges and Trends

[1,3] *N.M. Ali, ORCID: 0000-0002-3922-7136 <no3man_mohamed@himc.psu.edu.eg>*
[2] *B.A. Novikov, ORCID: 0000-0003-4657-0757 <borisnov@acm.org>*

[1] *Port Said University,*
Port Fuad, Port Said 42526, Egypt
[2] *National Research University Higher School of Economics,*
3, bldg. 1, ul. Kantemirovskaya, St. Petersburg, 194100, Russia
[3] *Saint Petersburg State University,*
7-9 Universitetskaya Emb., St Petersburg 199034, Russia

Abstract. The term Big Data refers to an extensive collections of digital data generating every second. Produced datasets come in structured, semi-structured, and unstructured formats throughout the world, which is difficult for the traditional database management systems to analyze. Recently, big data analytics emerges as an essential research area due to the popularity of the Internet and the advent of new Web technologies. This growing area of research represents a multi-disciplinary that attracts researchers from various research fields. Interested researchers are invited to design, develop, and implement several tools, technologies, architecture, and platforms for analyzing these large volumes of data. This paper begins with a brief introduction to big data and related concepts, including the main characteristics of big data, followed by discussions of the most significant open research challenges and emerging trends. Next, we review a study of big data analytics, the advantages of using big data solutions, and the preliminary assessments required before migrating from traditional solutions. Finally, we present a review of the recent main applications to obtain a broad perspective of big data analytics.

Keywords: Big Data; Big Data Analytics; Decision Making; Big Data Applications

Большие данные: аналитические решения, исследовательские задачи и тенденции

[1,3] *Н.М. Али, ORCID: 0000-0002-3922-7136 <no3man_mohamed@himc.psu.edu.eg>*
[2] *Б.А. Новиков, ORCID: 0000-0003-4657-0757 <borisnov@acm.org>*

[1] *Университет Порт-Саида,*
Египет, 42526, Порт-Саид, Порт-Фуад
[2] *Национальный исследовательский университет «Высшая школа экономики»,*
194100, Россия, Санкт-Петербург, ул. Кантемировская, д. 3, корп. 1
[3] *Санкт-Петербургский государственный университет,*
199034, Россия, Санкт-Петербург, Университетская набережная, д. 7–9

Аннотация. Термин «большие данные» относится к объемным коллекциям цифровых данных, генерируемых каждую секунду. Производимые наборы данных представлены в структурированном, полуструктурированном и неструктурированном форматах по всему миру, и

их трудно анализировать с применением традиционных систем управления базами данных. В последнее время аналитика больших данных становится важной областью исследований из-за популярности Интернета и появления новых веб-технологий. Эта растущая область исследований представляет собой междисциплинарную деятельность, которая привлекает исследователей из различных областей. Исследователи проектируют, разрабатывают и внедряют инструменты, технологии, архитектуры и платформв для анализа этих больших объемов данных. Эта статья начинается с краткого введения в проблематику большие данные и связанные с ними концепции, включая основные характеристики больших данных, после чего обсуждаются наиболее важные открытые исследовательские проблемы и возникающие тенденции. Далее приводится обзор исследований в области аналитики больших данных, обсуждаются преимущества использования решений для больших данных и обсуждаются виды оценок, требуемых перед переходом с традиционных решений. Наконец, представлен обзор основных существующих приложений, обеспечивающий общую панораму аналитики больших данных.

Ключевые слова: большие данные; аналитика больших данных; принятие решений; приложения больших данных

1. Introduction

The era of big data is now coming. This fact due to the popularity of the Internet and the advent of Web 2.0 technologies, also the increase of utilizing digital sensors, communications, computation, and storage that create massive collections of data [1]. Now, the volume of data available on the internet measured in exabytes (10^{18}) and zettabytes (10^{21}). Accordingly, expectations refer to that, in the next few years, the volume of data on the internet will exceed the storage capacity of living people's brains around the world [2].

Digital data generated every second throughout the world is producing in a structured, semi-structured, and unstructured format. This massive accumulation of generated data known as «Big Data». Moreover, the generation and adoption of specialized applications related to Social Media, Marketing, E-commerce, etc., provides extensive opportunities and challenges for researchers and practitioners. The erroneous volume of data generated by users using these platforms is the result of the integration between their experience and daily activities [3].

Recently, Big Data analytics has emerged as an important research area and intensively researched. Unfortunately, traditional data analytic techniques may not be able to handle such large quantities of data [4]. Such data consist of data sets that are difficult for legacy database management system to analyze [5]. Therefore, this emerging field has attracted researchers around the world to design, develop, and implement various tools, techniques, architecture, and platforms to analyze this growing volume of generated data [6-9].

Interested researchers are invited to handle the following challenges: how to design and develop a high-performance framework for efficiently analyzing big data; and how to design a suitable algorithm for mining and extracting useful information from big data [4]. To deeply discuss this issue, the structure of the paper is depicted below.

This section is an introductory section about the subjects and motivations for this paper. Section 2 describes topic foundations and the most significant aspects of big data, including the main characteristics of big data, the most significant research challenges, as well as the identification of emerging trends regarding big data. Section 3 presents a study of big data analytics and state the advantages of using big data solutions. Additionally, it involves asserting the necessary preliminary assessments that are required to perform successful migration to the new technologies. Next, Section 4 presents a review of the most significant fields that employ big data applications to obtain a broad perspective regarding big data analytics. Finally, in section 5, we provide some conclusions.

2. Big Data: Concepts, Characteristics, and Challenges

In this section, theoretical conceptualizations of big data and its characteristics are presented to reveal the challenges of tackling big data analytics in various fields of applications. Uncovering these challenges helps to determine, at a high level, essential functional and non-functional requirements that should put into consideration while the process of designing and developing big data analytics frameworks.

2.1 Concepts and Definitions

Proceeding from the fact concerning the existence of many types of modern digital technologies, that have permeated our daily lives like mobiles, sensors, and social media networks as a result of the expansion of using advanced digital artifacts. The proliferation of these technologies in everyday life enhances human-to-human, human-to-machine, and machine-to-machine interaction at unprecedented levels, resulting in vast amounts of data known as Big Data.

Several proposals appeared to describe this phenomenon and to give a definition of the term Big Data, which invented by Roger Magoulas from O'Reilly Media in 2005 [10].

James Manyika et al. [11] define Big Data as «datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze». According to this definition, there are no standard limits for considering dataset as big data (e.g., describe big data in terms of being larger than a certain number of terabytes). Also, we can notice that the volume of data is not the only factor in considering a dataset as big data. Therefore, it is significant to distinguish big data from massive data.

The analyst of Gartner [12] introduces a definition of big data, which considers one of the most comprehensive and widely use in this context, «Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation».

The challenge with Gartner's definition is that, in addition to state the main characteristics of big data, also, it focuses on describing how the benefit of big data can achieve and state the desired outcome. According to a clear understanding of these perspectives, organizations could determine whether they are using big data solutions or even if they have problems that need a big data solution, regarding the difficulties in scoping what is intended to design, developed and delivered, and what the result means to the organization.

In accommodation with Gartner's definition, David Loshin states in [13] that, «Big Data is fundamentally about applying innovative and cost-effective techniques for solving existing and future business problems whose resources requirements exceed the capabilities of traditional computing environments». Furthermore, Krish Krishnan [14] defines Big Data as «volumes of data available in varying degrees of complexity, generated at different velocities and varying degrees of ambiguity, that cannot be processed using traditional technologies, processing methods, algorithms, or any commercial off-the-shelf solutions».

Likewise, based on the distinctions between the capabilities of legacy database technologies and new data storage and processing techniques and tools (e.g., Hadoop clusters, Bloom filters, and R data analysis tools), Davis and Patterson [15] states that big data refers to «data too big to be handled and analyzed by traditional database protocols such as SQL». Also, Paul C. Zikopoulos et al. [16] says, «Big Data applies to information that can't be processed or analyzed using traditional processes or tools».

2.2 The Four V's Characteristics

For the most part, in the popularization of the big data concepts, the group of authors mentioned previously proceeding away from recognizing the size aspect of data only while the process of defining Big Data! Therefore, there are other meaningful characteristics of big data to be

considered in addition to the volume of data. The research community looks sticking with the attraction that appears in the common parts of presented definitions, which focus heavily on what referred to as the 3, 4 or even 9 V's as depicted in [17].

Although the definition of V's is ubiquitous, it should note that the origin of the concept is not entirely new, it provided by the analyst Doug Laney in a research note published by Meta Group (Now Gartner Group), from 2001 concerning «3-D Data Management» [18]. The author noted that the changing of economic conditions, affects the efforts done by companies as they struggle to standardize systems and fold redundant databases to enable greater operational, analytical and collaborative coherence, it also made this task more difficult. Also, he identifies e-commerce as the reason for raising data management challenges across three dimensions: Volumes, Velocity, and Variety. Finally, the author advised information technology organizations to assemble a variety of methods at their disposal to deal with each.

Commonly, big data characterizes by four V's characteristics, as mentioned in Fig. 1: Volume, Velocity, Variety, and Veracity. Other researchers have built upon that trend to include additional V's such as Visualization or Validity, intended to capitalize on an apparent improvement to the definition. As follows, a brief discussion of the fundamental characteristics of big data.
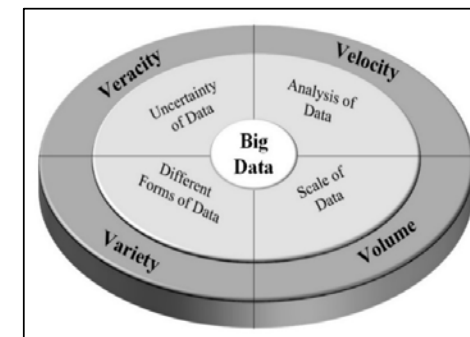

Fig. 1. Big Data Characteristics

- **Volume:** As the name implies, the size of data exceeds the capacity of traditional operational databases or data warehouses. In 2019, Hootsuite & We Are Social, published the Global Digital Statshot report regarding Internet Trends in Q3 [19]. The report displays the continuous growth of digital connectivity at an extraordinary rate around the world. Authors say that every day over the past year, almost 900,000 people came online for the first time.

  Also, many factors participate in increasing the volume of data like streaming data, storing different types of data from social networks, and other resources. Moreover, the Internet of Things (IoT), and scattered sensors all over the world in all devices that generate data every second represents a major grantor to the expanding digital universe [20].

  Consequence, International Data Corporation (IDC), expects that by 2025, the Global Datasphere will grow around 61% from 33 Zettabytes (ZB) in 2018 to 175 ZB. It noted that as much of the data residing in the cloud as in data centers [21].

  One of the primary goals is to make the volume of data useful for users and consumers and optimize future results. Nowadays, with decreasing storage costs, better storage solutions like Hadoop and the algorithms, the processing of large data sets, and creating meaning from all of the data are not a problem at all. Thus, companies are required to accommodate the new volumes by improving archiving and data importance strategies.

- **Velocity:** Denotes the speed of generating, storing, analyzing, and visualizing the data. It notes the high rate of data streaming into hosting platforms. Currently, the speed of data generation is almost unimaginable. For example, users upload more than 720,000 hours of new content per day on YouTube, which means 500 hours of fresh video per minute [22]. Moreover, there are 500 million tweets sent every day on averages of more than 20,000 tweets per minute [23]. Also, in 2014, the Facebook research center reported that over 4 new petabytes of data generate and run of 600,000 queries per day [24].

  This characteristic, in addition to the high rate of data generation, imposes an essential concern on data aging, and the lifetime of data. How long the data will be valuable is a big challenge that organizations have to cope with, regarding the high rate of data generation and use in real-time. The speed of data generation requires keeping up with processing tasks to meet the demand. Sometimes, the speed of applying the analysis of streaming data is critical [25].

- **Variety:** Big data refers to the large volumes of data generated in different formats. The complexity of Big data formats requires different approaches and techniques to store all raw data. Several different types of data differ in the way of creation and store. These types require various types of analysis to apply or use different tools. According to nature and characteristics, data categorized into three types.

  Structured Data: In the past, most created digital data was structured data, but today it constitutes around 10% of the total digital data. This type of data concerns all data which could be neatly fitted in columns and rows and stored into a spreadsheet or database. Systematic data refers to highly organized information, with relative simplicity in entering, storing, querying, and even analyzing, but a strict pre-definition of the field name and type is indispensable, besides having a relational key to be easily mapped into pre-designed fields. Relational data (e. g. SQL database), Meta-data (e. g. time and date creation), Library catalogs (e. g. date and author), Census records (income and employment), and Economic data (e. g. GDP) are various examples of this type.

  Semi-structured Data: This type refers to data that have some organizational properties, such format could help in the analysis process, also known as a self-describing structure. Unlike structured data, it couldn't establish in a rational database directly among the formal structure of data models associated with relational databases or other forms of spreadsheets. However, this format implies the involving of tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Special processing may help to neatly fit, and store data into a spreadsheet or database. Examples of semi-structured data exist like XML (e. g. stored personal data), JSON (e. g. script documents), and NoSQL (e. g. databases like MongoDB).

  Unstructured Data: This type refers to, as the name implies, non-systematic data. That may include every kind of data that carries an unknown form or structure. It may have an internal structure but does not conform neatly into a spreadsheet or database. Today more than 80% of the data that is generated by the organization is unstructured data. There exist many examples of unstructured data like sensor data, social media streams, images, videos, mobile data, text files, etc.

- **Veracity:** Refers to the reliability of the data. Many parts of data consider Not useful such parts involve noise, biases, anomalies, and abnormal data. Wrong data will lead to misleading results and incorrect decisions. So, organizations must decide before beginning the analysis, whether data meaningful or not. Accordingly, ensure the trustworthiness of data sources, in addition to the correctness of data, are mandatory. Today's developers become involves and more willing to invest in the effort to clean up the data at the source.

  Consequently, the term Big data exceed the reference to datasets only, and expand to cover space problems, technologies, and opportunities to enhance business value, so it considers

a general term. Specifically, the realization of the tremendous business value of data is the main reason for using big data. However, the expectation that the adjective big will fade over time still considered meaningful, as the explicit meaning of the data will be intuitively expanding to include all data types [26].

2.3 Big Data Challenges and Trends

Big Data is a general term for large and complex datasets where traditional techniques and applications for data processing become inadequate. The main objective of processing big data is to help organizations make appropriate decisions on various important issues. Therefore, before moving on to use big data solutions, organizations require to understand the nature and risks involves [27].

Accordingly, we can identify the main challenges of Big Data that should be taken into consideration while handling a big data solution [28-30]: Understanding of Big Data, Data Storage, Data Quality, Data Integration, Data Processing, Data Privacy and Security, Data Visualization, and Data Scalability. Within these challenges, some promoted by the characteristics of big data, others are through the current analytical models and methods, and others are due to boundaries of existing data processing systems [31]. In this section, a brief discussion regarding all the challenges listed in the above presented as follows.

2.3.1 Understanding of Big Data

Big data is basically about applying innovative and cost-effective techniques and technologies for solving existing and future business problems. These solutions handle difficulties that require resources and capabilities beyond the boundaries of traditional computing environments. Organizations require best understand the big data basics like what big data is, what its benefits are, what infrastructure is needed, etc.

Lack of a clear and sufficient understanding of the value that big data can offer an organization leads to the failure of the big data adoption project. The clear understanding and considering the market conditions help to avoid waste lots of time and resources on things they don't even know how to use.

Among the challenges listed before, poor understanding of what big data means, and how to use it could be considered the root of these challenges. It is very significant to differentiate between big data and any other digital trend. The power of big data helps to transform business and boost its efficiency. Organizations must invest in careful planning, come up with a decent strategy, and well-organized architecture. To achieve positive results, and to realize significant effective big data management and reduce the cost of upgrading the future.

Many evaluation criteria need to pass successfully before proceeding with the process of making decisions regarding the integration of big data solutions as a part of an enterprise information management architecture. There exist many variables that are relevant to the evaluation process. These variables include the characterization of the term big data, the examination of the reasons of inadequacy the traditional data management framework with the growing data variability with the evaluation of owned technologies, etc.

2.3.2 Data Storage

The storage and management of the massive data generated through various devices are vital challenges in Big Data. Most enterprise focus settles on analytics issues, but they will never get there without having an efficient, long-term data storage solution to provide a stable foundation. Data requires a place to stay in, if an organization plan on keeping enormous amounts of data, there is a necessity for invest in storage infrastructure [32].

Several approaches proposed to deal with this problem [33], but the trade-off between cost and effectiveness still a significant challenge. One of the current solutions is to take advantage of another company's infrastructure to save data by using cloud hosting and cloud storage.

On the other hand, concerning a data management perspective, the limitations of existing techniques regarding the high rate of data generation is another challenge. Various techniques proposed to solve this problem that involves the activities of data clustering, data replication, and data indexing [34, 35]. However, these activities impose a development challenge to improve its effectivity and performance [36-38].

2.3.3 Data Quality

Regardless of the size of the data, the need for data quality still urgent. The influence of data quality on achieved business value from big data settled data professionals under stress, which is a fundamental property of data that determines its reliability for making decisions. Preservation of the quality, integrity, and relevance of data represents a trickier challenging task. With the lack of satisfactory quality and relevance, the data processed will be useless [39].

Depending on the type of analysis designed, certain data needs to be collected and managed in a particular way, to handle the new challenges. The collected data must be valuable for the analysis to achieve the correct results. The next step involves applying the appropriate techniques to the gathered data for assuring its quality and relevance.

Organizations look for realizing the maximum values of their data assets. The task of data quality assurance requires to deliver results like trusted analytics, operational reporting, self-service functionality, business monitoring, and governance for taking decisions.

Danette McGilvray proposed many dimensions to describe the quality of big data that represent characteristics or aspects of data quality [40], which depicted in Fig. 2 as follows: Data specifications, Data integrity fundamentals, Duplication, Accuracy, Consistency & Synchronization, Timeliness & Availability, Ease of use & Maintainability, Data coverage, Presentation quality, Perception, Relevance & Trust, Data decay, and Transactability. These dimensions present a method for measuring and managing the quality of data and information. Each one needs various tools, techniques, and processes to measure it. That results in different levels of time, money, and human resources to complete the evaluations.
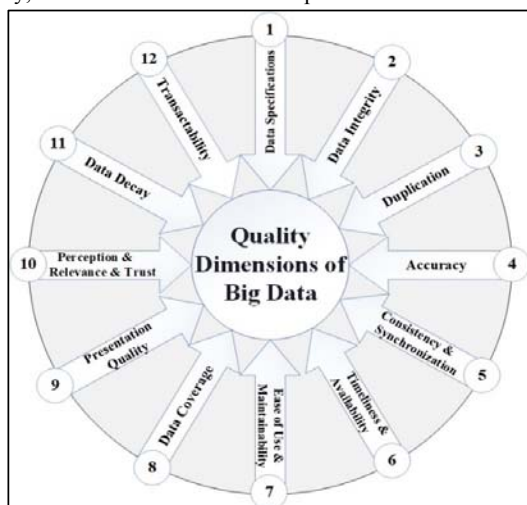

Fig. 2. Quality Dimensions of Big Data

There is a possibility to adapt traditional data quality techniques like virtualization to the new paradigms of modern data management. Adjustments and optimizations make data quality tasks (e. g. Standardization, deduplication, matching, profiling & monitoring, and customer data) are relevant to big data [41].

2.3.4 Data Integration

The problem of data integration appears when organizations realize there is a need to analyze data that comes from diverse sources in a variety of different formats. The "variety" characteristic of big data makes data integration a great challenge since differences between several data structures became much more significant and matching them is problematic. Additionally, the exponential growth of volumes and velocities, driving both systems and processes to their limits [42].

Frequently, data warehouses use data integration techniques and combine multiple data sources, to consolidate operational system data, and to enhance reporting or analytical needs. An example of integrating data of various types may be that eCommerce companies need to analyze data from website logs, call-centers, e-mails, scans of competitor's websites, and social media. All listed information can be integrated and made available to support decision-makers [28].

The combining of data from diverse external sources is an additional complication imposed by big data on the data integration process, that due to the limited control of organizations over data standards at the source. The integration process of big data involves additional challenges. Such challenges need to be considered as a confusing diversity of big data management technologies that mean risk in the selection, synchronizing data across various sources, the lack of expertise, moving data into a big data platform, extracting useful information from big data, etc. [43].

2.3.5 Data Processing

Data processing is the method to be applied after the data storage to extract useful information. The primary goals of data processing are to get a significant relationship between the various fields of data collection, and the extraction of valuable analysis. The complexity and scalability of big data make data processing tasks more complicate and imposes another challenge [44].

The traditional paradigm regarding performing operations on the consolidated datasets becomes inappropriate. Also, the combination of all related data in a whole database may cost a lot of time and require extra investments in infrastructure. The shortages of traditional techniques appear during the time it takes to analyze through a single set of datasets, while the speed of processing a query in big data is significant demand [45].

Generally, the data processing tasks employ two main techniques, classification, and prediction. Classification is a data mining technique used to divide datasets into different categories and groups. On the other hand, the prediction is responsible for making decisions based on the analysis results of past transactions. Regarding big data, in the case of manipulating massive volumes of data collected from various sources with different formats, the data processing task becomes tricky.

Therefore, a shift in paradigm includes application parallelization to simultaneously process against multiple chunks of data and divide-and-conquer are natural computational techniques for handling big data problems. This approach moves the processing to the place where the data stored by distributing the query and update requests across distributed servers rather than attempting to process against one combination of a dataset.

Several techniques introduced in the context of big data processing for storing the unstructured data in distributed databases, like HBase, Apache Cassandra, or SimpleDB , and for classifying data, techniques like MapReduce in Hadoop introduced [46]. Also, to enhance query processing, optimization techniques introduced like HiveQL, SCOPE, etc. [47].

2.3.6 Data Privacy and Security

The enormous volumes of big data may threaten to overwhelm the organization's ability to preserve its privacy, as analysis results provide complete information about activities and processes. Consequently, giving a low priority to the security of big data, and turning it off until later stages of big data adoption projects, is not always a smart move, "Security first" is an indispensable requirement [48].

Recently, the number of cyberattacks has a significant increase, and their sophistication is expanding, seeking to traverse corporate firewalls, extract critical business information, or gradually deplete individual financial accounts while working entirely under the radar. Monitoring for cybersecurity events and ensure the privacy and security of data lies in the organization that is responsible for keeping the data (e. g. Cloud providers) from various sources and a wide variety of massive streaming datasets [49].

Therefore, the necessity to quickly capture and integrate threats into a model is inevitable, which enables for identification of known attack patterns as well as the discovery of the new emerged patterns as the attacks become more complicated [50]. Hence, the new security intelligence solutions are required to combine big data with advanced analytics, to link security events across multiple data sources, and provide early detection of suspicious activities. As follows, a list of the most ferocious security challenges that big data involve [51-54]:

1. the access and usage of data by unauthorized persons;
2. absence of security audits;
3. encryption protection problems;
4. possibility to extract sensitive information;
5. high speed in the development of NoSQL databases and lack of security focus;
6. data source difficulties;
7. the generation of fake data.

2.3.7 Data Visualization

The presentation of data in a readable manner and making it understandable for users represents a difficult task. Data visualization task includes the representation of essential information and knowledge efficiently and intensively by employing various types of infographics and other visual formats. Regarding the characteristics of big data (e. g. Large volume, variety, and velocity of the information), visualization becomes a big challenge [55].

Frequently, the use of data visualization gives the ability to summarize and review large volumes of data into a format that is helpful to human consumption, which may provide the ability to move to an additional level of detail upon request. Visualization of big data aims to provide decision-makers and other business users with insights.

Several solutions for data visualization are available like Tableau, Microsoft Power BI, Sisense, QlikView, etc., and the selection of the right and the appropriate tool indicates a bit tricky. Also, choosing the most proper data visualization technique for application from a wide variety of popular methods and techniques (e. g. Symbol maps, tree map, line charts, pie charts, heat maps, bar charts, scatter plot, map chart, parallel coordinate plot, etc.) appears to be a more complicated task than it seems [56].

2.3.8 Data Scalability

The organizations may be hindered by the limitations of their owned infrastructure for data acquisition for analytics while dealing with enormous volumes of big data. Although scalability is not a unique challenge in big data solutions, the ability to grow represents a crucial feature of any big data solution. Regarding the high complexity of algorithms, software scalability has always been a problem [57].

The scalability meaning in big data refers to the ability to grow with the rapid increase of data volumes, which require a change in the storage process, management, and analytical techniques. Therefore, if the technical infrastructure of an organization designed around a traditional data warehouse information flow, it may hinder its ability to handle big data problems and capability to manage the hesitated volume of data in real-time.

Traditionally, the research efforts for solving scalability problems concentrate on parallelizing the computation, with less resolution on storage distribution. Commonly, scaling approaches classify into two main classes: vertical and horizontal scaling [4, 58].

Vertical scaling, also known as scaling-up, aims to improve the performance of the system, it could achieve by enhancing the capability of processing platforms by adding additional computing power (e. g. RAM, CPU, etc.) to accommodate further data volumes. An example regarding this type of scaling platform is High-Performance Computing (HPC) clustering. This approach hinders by its high cost and complexity in terms of maintenance, also by the restrictions imposes by the platform upper limits (e. g. Maximum capacity of RAM, number of CPUs, etc.).

On the other side, horizontal scaling, also knowns as scaling-out, could achieve by adding more machines interconnected over a network. It employs a divide-and-conquer approach (e. g. Apache Hadoop). That helps in distribute the workload and generate parallel processing over multiple independent computing machines. Accelerate data processing by adding more computing machines as much as needed to improve the overall system performance. The ability to run and maintaining different computing machines with various operating systems impose additional challenges toward managing and maintenance these instances [51, 59].

In big data, the volume and variety of data can differ dynamically in response to potentially variable user demand. Hence, the process of up-and-down scaling according to the request for computational resources represents an important characteristic of big data solutions. That is due to the difficulties concerning allocation and de-allocation of resources in real-time, as they have an impact on the overall system performance.

3. Big Data Analytics

Today, enterprises seek to discover facts they didn't know before by searching for massive amounts of highly detailed data. Analysis of large datasets unfolds and improves business values. However, the growing volumes of data, increase the difficulties with management and manipulation. This section demonstrates the concepts and basics of big data analytics and the benefits of exploiting available assets of big data regarding Business Intelligence (BI) [60].

3.1 Topic Conceptualization

Big data analytics refers to the advanced analytic techniques to apply on big data sets [61]. An example of analytical tasks may include searching for specific data, and patterns, data retrieval, and organization, etc. Therefore, the application of superior analytical techniques comprises the entire processes and tools required for knowledge extraction [62]. These processes incorporate multiple tasks that start with data acquisition and extraction, followed by the transformation of data. Next, preparing and loading data for analysis that includes the employment of appropriate tools and techniques for getting desirable results. Finally, the delivery of realized results to support decision-makers [58].

Generally, data analysis involves the examination of sets of data to uncover hidden patterns, correlations, and other insights as well as rendering conclusions. These tasks could classify into three main areas, Statistical analysis, modeling, and predictive analysis [63]:

1. *Statistical Analysis*: Regarding business intelligence, it involves the collection and examination of each sample of data in a set of elements from which it can draw. Statistical analysis strictly related to hypothesis testing. The main goal of statistical analysis is to recognize and predict future trends.

2. _Modeling_: Refers to the processes used to identify, describe, and analyze the requirements of data, also to describe the overall behavior of a system, it plays a crucial role in the growth of any business. These processes include the use of mathematical equations or some logical language. The main objective of data modeling is to support business processes and give answers to business questions more easily and quickly.

3. _Predictive Analysis_: Represent a form of advanced analytics that concerned with guessing how an individual, group, or data object will behave and forecasting trends based on historical data or the recognized behaviors of similar individuals or groups. It includes several types of algorithms like recommenders, classifiers, and clustering.

Big data analytics are different from small and traditional data analytics. From this perspective, Joshua Eckroth [6] defines big data in adaption with a definition from Philip Russom [61] as follows: "A data analysis task may be described as big data if the data to be processed have such high volume or velocity that more than one commodity machine is required to store and/or process the data".

Otherwise, from the perspective of Business Intelligence and the advantages of data analytics, Rick F. van der Lans [64] says, "Big data is revolutionizing the world of business intelligence and analytics". Thus, big data analytics revolves around two items, big data, and analytics, as well as how the collaboration and the combination between the two items could perform to create one of the most intellectual trends in business intelligence today [61].

3.2 Advantages of Big Data Analytics

Despite the difficulties and challenges involves in the process of building and developing big data analytics platforms due to the complex nature of big data, enterprises working quickly to build analytical solutions for big data. This situation is because of the great opportunity it offers to upgrade from the traditional methods of information extraction into new dimensions. Decisions made without data-driven answers will likely fail, so organizations must build their systems that support data-driven decision making [59].

Traditionally, organizations recognized that the insights of owned data could extensively benefit their business performance. The importance of big data does not rely only on the volume of data a company has but how a company utilizes the collected data. Hence, big data analytics represents a competitive advantage for businesses. Achieved benefits and values utilizing big data analytics solutions could be determined depending on the way enterprises use the data [65].

Enterprises are now facing challenges to create new business actions based on the benefits brought by available types of analysis. Efficient analysis of owned and collected data helps the business to find answers which will improve the organizational value [66]. Moreover, companies that use comprehensive big data analytics solutions realize the benefits, obtaining even more insights that drive intelligent decision-making. These insights represent new means of making business by leveraging new types of analytics across different types of data. Some benefits of using big data analytics solutions include [67-69]:

- the identification of the root causes of failure and issues in real-time;
- cost-saving;
- reduce processing time;
- enable the development of new products and services;
- a better understanding of market conditions;
- reassess risk portfolios quickly;
- full understanding of the potential of data-driven marketing;
- improve customer engagement and increase customer loyalty;
- personalize the customer experience;

- generate customer offers based on their buying habits;
- add value to online and offline customer interactions;
- controlling the reputation online.

Therefore, we can say that the significant contribution of big data to the organization revolves around four main elements, the essence of these items reflects on adding value to the organization by Increasing revenues, lowering costs, increasing productivity, and reducing risk.

3.3 Preliminary Assessments

Before starting to build a new big data solution, enterprises are required to perform primary assessments. The Migration from traditional analytic solutions to big data may cost a lot of time and require extra investments in infrastructure. So, it is significant to know the actual state of the business and evaluate the needs of this step. The trade-off between costs and potential benefits is essential to select the best path and decide whether to continue with traditional schemes or to transform into the new one. Without proper organizational evaluation and preparedness, neither strategy is likely to succeed [70].

The big data approach is more appropriate to address or resolve business problems that meet one or more of the following criteria as depicted in the Fig. 3 [71]: Data throttling, computation-restricted throttling, large data volumes, significant data variety, and benefits from data parallelization. These evaluation criteria can be used to assess the degree of relevance of business problems toward big data technology if there exists a correlation with business problems, whose solutions fit big data analysis applications.
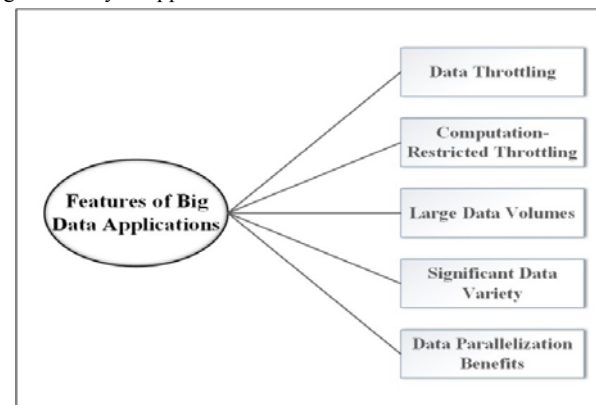

Fig. 3. Evaluation Criteria to Build a Big Data Solution

4. Big Data Applications

Concerning the continuous growth in the field of big data analytics and its substantial impact on human life, big data applications introduce cutting-edge opportunities in every aspect of our daily life. The primary goal of big data applications is to help companies make more informative-business decisions by analyzing large volumes of data, given the tremendous competition where we are living overall the world [72].

The applications of big data positively change our life for the better and smoother as well. We are utilizing big data solutions to improve our efficiency and productivity. This opportunity encourages researchers and technology providers to develop complicated platforms, frameworks, and algorithms to struggle with the challenging of big data. In this section, the most significant fields where use big data applications will discuss as follows.

4.1 Education

Education is the backbone of any country; it represents the engine of growth and prosperity in the lives of nations. Successive civilizations reflect the level of scientific and civilized development throughout the ages. Based on the great importance of education, it was necessary to continuously develop educational systems around the world to raise the efficiency of the educational process and the quality of the service provided [73].

The main characteristics of education development appear clearly in the expansion of using e-learning systems by offering the courses online [74]. These systems offer interactive online courses that involve carrying out assignments by learners as well as submitting results [75]. Data generated by using these educational systems are personal information, academic progress, attendance, student status, student activities, and interests, etc. Additionally, other types of data offered by systems to administrators like financial status, course plans, staff details, organization details, etc.

The responsibility to ensure the quality of education depends entirely on the government and educational institutions. Many of them are planning to implement smart education systems to improve e-learning systems in their countries. Since the availability of large quantities of data generated from several types of educational frameworks, the necessity to analyze such data to obtain insights is growing. Analysis of these types of data requires special techniques for efficient processing that are more suitable for big data techniques [76].

Big data solutions can produce fantastic results and offer innovative data-driven approaches for student learning. In many countries, using big data applications are common in schools and colleges. But developing countries are also gaining new technologies. Using these technologies enables storing, managing, and analyzing large datasets with maintaining security [77]. Also, it provides relevant data on class activities that are recorded using high-definition cameras and video clips that help evaluate student facial expressions and can track their movements as well as make decisions for organizations [78].

4.2 Agriculture

Starting with the significant role of agriculture that plays in the development of the social economy of any country, this has been a growing interest in developing big data solutions in this industry. Generally, agriculture is based mainly on geographical and climatic conditions. The main factors on which crop production depends are climate, temperature, precipitation, agriculture, fertilizers, pesticides, etc.

Nowadays, due to the advent of advanced technology in the area of agriculture, the procedures of crop production have changed, and it is possible to control greenhouses and cultivate crops by managing the temperature, humidity, sunlight, etc. The most recent greenhouses are provided with the latest sensors devices to determine the quality of the soil condition that represents an essential factor for crop yield [79].

In agriculture, big data is playing an influential role in improving the performance of the firms. The goal is to optimize crop efficiency by minimizing the firm's loss and increasing the generation of necessary food grains. The amount of collected data from sensors during the process of planting crops and running simulations to measure how plants react to various changes in conditions are pretty huge [80]. So, the processing of these data using big data analytics techniques allows it to discover the optimal environmental conditions for specific gene types [81].

The use of big data analytics helps the producers in overall the processes to decide on the crops, fertilizers to use, pesticides, etc., as well as from harvesting to distribution process of agricultural products like paddy, wheat, vegetables to increase the benefits. Other advantages may include the automation of the watering system of the firm and enabling the firm's owners to use the same land for several purposes throughout the year without any interval [82, 83].

4.3 Healthcare

Big data analytics have already affected patient care and pharmaceutical manufacturers [84]. Traditionally, the use of big data solutions in healthcare manufacturing has been much delayed compared to other industries. This problem caused by several factors, one of them relies on the resistance to change by service providers of using an independent-decisions approach that employing their clinical judgment to make treatment; rather than relying on protocols of big data [85].

Additionally, critical information inside a single hospital, payor, or pharmaceutical company, is often kept silent and isolated within a single group or department. That caused by the lack of procedures in organizations to integrate data and report results. Other factors are more structural. Recently, healthcare stakeholders like pharmaceutical industries and hospitals, now have access to promising new threads of knowledge and adopting with the facilities provided by big data analytics [86].

Although big data technology is in the initial stage in healthcare, these technologies help the industry make critical decisions, which in turn play an important role in enhancing the ability of healthcare organizations to make a good profit by providing services to patients. These reasons have made the healthcare industry a significant commercial system. Typically, healthcare stakeholders are using online frameworks to publish diagnostical reports, analysis reports, schedule appointments, monitor patient status, and preserve records; these systems represent a principal resource for big data.

The growing generating of tremendous varieties of data of healthcare systems like clinical trials, medications, exercise directions, allergies, insurance data, visiting schedule, treatment follow-ups, etc., represents a rich source of data that make pharmaceutical industry experts, payers, and providers, for turning to analyze such data to obtain insights.

The volume, complexity, and diversity of these types of data make traditional relational database systems are unable to manage and process those large datasets. However, the capability of big data techniques to manipulate whatever the type of data allows the manipulation of those datasets efficiently. So, recent technological advances in the industry have improved their ability to handle this data, as well as developing secure frameworks [87-89].

The introduced solutions aim not only for treatment identification but also for improving the process of rendering healthcare. Additionally, big data has a high impact on reducing consumption of money and time; it enables the development of new infrastructure and emergency medical services. The use of big data has many advantages that are difficult to list fully in this context, for example, the evaluation of symptoms and identification of many diseases at the early stages based on the availability of medical databases, which plays a significant role in disease prediction.

4.4 Smart Cities

According to the United Nations estimations, 1.3 million people are moving into cities each week, and by 2050, 68% of the world's population is expected to be living in cities, with close to 90% of this urban population growth set to occur in Africa and Asia [90]. Therefore, experts are spending numerous efforts attempting to improve the quality of life in our cities.

Generally, the term Smart City mainly supplies by the advent of IoT (Internet of Things) and Big Data [91]. A simple definition of a smart city is "A city equipped with the basic infrastructure to provide a high-quality lifestyle to its citizens". There are many areas of development in the city that must identify to reconfigure the current situation like water management, waste management, transportation, and safety. Smart cities of the future must comprise the necessities and higher technologies for effortless and elementary living.

The massive expansion in the use of IoT technologies has made it easy to communicate with devices without human intervention to collect data in real-time [92, 93]. Sensors installed all over the city producing data regarding critical infrastructures like rails, airports, seaports, roads, power, water, and communication. These devices, in addition to other types like cameras, GPRS, etc., generate enormous amounts of data, effective use enables achieving of many improvements. Collected data from these resources serve as a tremendous source for big data [94].

Big data plays a significant role in processing the gathered data and represents an emerging trend in the field of information systems; so that further analysis can be made to recognize the patterns and needs in the city. Analytical solutions based on big data aim to get insights and extract correlations relationships to improve services provided to residents by optimizing the usage of resources, and managing maintenance activities, as well as to support the decision-making processes. Many challenges are facing the development process of city-level smart information services like the integration of generated datasets from various city domains, and analysis process as well [95, 96].

The analysis of these resources offers many benefits of smart cities, particularly three main areas attest to why data-driven innovation is crucial to the future of urban life as follows [26, 97, 98]:

1. intelligent traffic management through the proper utilization of historical data;
2. promoting public safety by using predictive analytics that helps to examine historical and geographical data to recognize when and where crimes occur;
3. managing smart cities infrastructure, the evaluation of the current situation helps to enhance city planning, effective spending, and maintain sustainability.

4.5 Criminal Analysis & Fraud Detection

With the continuous development of criminal methods, which makes it more professional and sophisticated, it becomes necessary to resort to advanced techniques that rely on data-driven analytics to meet growing risks and challenges. Criminal operations differ in their forms and domains, and among the most complex are those carried out through using electronic devices and the disappearance from behind them to cover up the eyes. The determination of fraud against enterprises that involves any type of operations like claims or transaction processing is one of the most compelling examples of big data applications.

Recently, governments, security agencies, and enterprises have started using big data analytics in the fields of security and law enforcement. Various domains are affected by well-planned crimes such as drug trafficking, kidnaps, terror attacks, fraud, and robberies in an increasing manner. Big data offers a large variety of solutions to handle these types of crimes effectively, which already proved to be very beneficial in preventing criminal activities [99-101].

Historically, the discovery of fraud has proven to be an elusive purpose. Usually, the detection of fraud done long after it occurs. That means actual injury has already happened, and all that remains is to reduce the damage and set policies to decrease the opportunity of repeating occurrence. The advantages of using a well-designed big data framework play a significant role and could change the game of fraud detection. It can investigate claims and transactions in real-time, discover general patterns across multiple transactions, or recognize abnormal behavior from an individual user [102-104].

Many areas are using big data technologies to detect fraud and electronic theft. The collected information for criminal analysis comes from various sources like bank transaction records, mobile call records, web, and social media (e. g. Social meeting sites like Facebook, Twitter, and LinkedIn), etc. [105, 106]. These areas include what is related to stealing money through fake electronic cards and using them in purchases and obtaining services illegally.

Credit card fraud considers one of the most common aspects of electronic fraud. Therefore, many approaches introduced to handle this problem and related fields, including illegal purchased for goods and services [107-109]. Another example compromises to healthcare services, that individuals and criminal networks whose commit fraud for nefarious reasons, personal gain, and obtaining medicines and medical supplies to gain private benefits illegally. That works on eliminating the ability of enterprises to effectively provide the healthcare needs of the elderly and other qualifying people [110-112].

4.6 Government

Government work provides an opportunity to help the public or provide services that add real benefit to the lives of citizens. They need to deal with various complex local, national, and global issues daily. One of the greatest strengths of big data is flexibility and the overall application of many different industries. Besides many other areas, big data in government can have a tremendous impact. Much of this work is being done in the public sector using big data and analysis, due to the use of analyzes can improve the results of general programs.

The implementation of a big data platform can leave an enormous impact on the governmental sector [113]. Governments can access vast amounts of relevant information about millions of people necessary to accomplish their daily functions, as well as help to make any decision regarding locals. Governments have to try to make sense and analyze the impact and opinions about vital decisions that affect millions of people, and to decide if any change is needed or not [114, 115].

The positive impact of using big data solutions is almost endless. It is very significant because it not only allows the government to identify areas that need attention but also gives it that information in real-time [116, 117]. Big data analytics has proven to be very useful in the government sector regarding the significant role it plays in election campaigns. Additionally, governments utilize numerous techniques to ascertain how the electorate is responding to government action, as well as ideas for policy augmentation [118].

Big data analytics provide tremendous benefits to the public sector by improving the outcomes that have a direct impact on citizens' lives [119]. Examples of these sectors that can be applied to achieve tangible results may include Emergency response, anti-money laundering, insider threats, workforce effectiveness, etc. Moreover, it can help in the success of government campaigns to eliminate problems that affect national security, such as the drug problem and poverty. Besides, it helps in predict any terrorist attack and take necessary action to prevent unwanted conditions. Finally, the analytical ideas obtained from owned big data stores could make a difference and make the government more effective [120, 121].

4.7 Marketing & E-Commerce

The tremendous expansion in the use of modern technology and various devices in commercial transactions has made marketing and electronic commerce, one of the most informative areas. This development has brought about drastic changes in the map of the global economy due to the successive changes in trade and marketing strategies to keep pace with this tremendous development. Marketing trends for companies have completely changed [122].

Digital marketing is the key to success in any company. Now, any business can manage marketing promotional activities and run successful advertising campaigns and promote their products and services regardless of their size. Big data has made digital marketing powerful and has become an essential part of any business. Various forms of marketing campaigns appear in different places, such as social media platforms (e. g. Facebook, LinkedIn, Twitter, etc.), ads placed on YouTube and TV, companies' websites, E-Markets, Text messages, E-mails, etc.

Big data analytics offers several solutions for helping enterprises in analyzing available tremendous datasets [123]. One example investigates at what kinds of advertisements compel viewers to continue watching and what turns viewers off. It uses facial-recognition software to learn how well their advertising succeeds or fails at stimulating interest in their products. That helps marketers to create widely accepted ads to increase sales.

Another example of using big data solutions for analyzing customer calls. Analyzing the content of customer contact records with call centers helps determine their sentiment that represents a powerful barometer and influencer of the market sentiment. Big data solutions can help recognize repeating issues or patterns of customer and employee behavior not only by understanding time/quality accuracy metrics but also by recording and analyzing the content of the call.

The advent of social media is one of the most substantial contributors to big data. Based on its significant importance, various solutions introduced to analyze user's activities [124-126]. Big data analytics can provide valuable insights in real-time about how the market is responding to products and campaigns [127]. According to obtained results, companies can adjust their prices, forecasting demand and its distribution, adjusting production, update distribution strategies, promotions, and campaign placements accordingly [128-131]. Therefore, to know the consumer mindset, it is necessary to apply smart decisions derived from big data [132].

The comprehensive development of e-commerce and the expansion of establishing electronic markets for online shopping like Amazon, eBay, Walmart, Best Buy, Wish, etc., have created a competitive environment between companies to attract the highest number of customers. Enterprises seek to measure several factors like customer satisfaction, loyalty, and the success of marketing strategies by analyzing customer reviews through electronic platforms such as websites and social media [133-136]. But the analysis process must involve excluding fake reviews and negative comments released by competitors [137].

5. Conclusions

Digital data generated every second throughout the world is producing in a structured, semi-structured, and unstructured format. Unfortunately, traditional data analytics techniques are not able to handle these volumes of data considering their complex structures. Therefore, big data analytics has emerged as a substantial research area, and intensively researched to handle these problems.

In this paper, we present theoretical conceptualizations of big data and its characteristics to reveal the challenges of tackling big data analytics in various fields of applications. Uncovering these challenges helps to determine, at a high level, essential functional and non-functional requirements that should put into consideration while the process of designing and developing big data analytics frameworks. Also, we gave a demonstration of the concepts and basics of big data analytics and the benefits of exploiting available assets of big data regarding Business Intelligence.

The primary goal of big data applications is to help companies make more informative-business decisions by analyzing large volumes of data, given the tremendous competition where we are living overall the world. In this paper, we review the most significant fields that employ big data applications to demonstrate how it positively changes our life for the better and smoother as well as to improve our efficiency and productivity. Also, we state opportunities for researchers to develop complicated platforms, frameworks, and algorithms to struggle with the challenging of big data. In the future, the research will move towards the landscape of big data tools and specialized techniques offered for big data analytics, investigating its functionality and limitations.

References

[1]   Ghani N.A., Hamid S., Hashem I.A.T., Ahmed E. Social Media Big Data Analytics: A Survey. Computers in Human Behavior, vol. 101, 2019, pp. 417-428.

[2]   Emani C.K., Cullot N., Nicolle C. Understandable Big Data: A Survey. Computer Science Review, vol. 17, 2015, pp. 70-81.

[3]   Stieglitz S., Mirbabaie M., Ross B., Neuberger C. Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation. International Journal of Information Management, vol. 39, 2018, pp. 156-168.

[4]   Tsai C.W., Lai C.F., Chao H.C., Vasilakos A.V. Big Data Analytics: A Survey. Journal of Big Data, vol. 2, no. 21, 2015, pp. 1-32.

[5]   Yadav K., Rautaray S.S, Pandey M. A Prototype for Sentiment Analysis Using Big Data Tools. In Proc. of the First International Conference on Computational Intelligence, Communications, and Business Analytics, 2017, vol. 775, pp. 103–117.

[6]   Eckroth J. A Course on Big Data Analytics. Journal of Parallel and Distributed Computing, vol. 118, no. 1, 2018, pp. 166-176.

[7]   Smirnova E., Ivanescu A., Bai J., Crainiceanu C.M. A Practical Guide to Big Data. Statistics & Probability Letters, vol. 136, 2018, pp. 25-29.

[8]   Siddiqa A., Hashem I.A.T., Yaqoob I., Marjani M., Shamshirband S., Gani A., Nasaruddin F.A Survey of Big Data Management: Taxonomy and State-of-the-Art. Journal of Network and Computer Applications, vol. 71, 2016, pp. 151-166.

[9]   Soufi A.M., El-Aziz A.A.A., Hefny H.A. A Survey on Big Data and Knowledge Acquisition Techniques. IPASJ International Journal of Computer Science (IIJCS), vol. 06, no. 07, 2018, pp. 15-29.

[10] Halevi G., Moed H.F. The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature. Research Trends, no. 30, 2012, pp. 3-6.

[11] Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A. H. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011, 143 p.

[12] Gartner Glossary: Big Data. Available at: https://www.gartner.com/en/information-technology/glossary/big-data, accessed 14.10.2019.

[13] Chapter 1. Market and Business Drivers for Big Data Analytics. In Loshin D. Big Data Analytics, Morgan Kaufmann, 2013, pp. 1-9.

[14] Chapter 1. Introduction to Big Data. In Krishnan K. Data Warehousing in the Age of Big Data, Morgan Kaufmann, 2013, pp. 3-14.

[15] Davis K. Ethics of Big Data: Balancing Risk and Innovation. O'Reilly Media, 2012, 82 p.

[16] Zikopoulos P.C., Eaton C., deRoos D., Deutsch T., Lapis G. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media, 2012, 176 p.

[17] Owais S. S., Hussein N. S. Extract Five Categories CPIVW from the 9V's Characteristics of the Big Data. International Journal of Advanced Computer Science and Applications (IJACSA), vol. 7, no. 3, 2016, pp. 254-258.

[18] Laney D. 3D Data Management: Controlling Data Volume, Velocity and Variety. Application Delivery Strategies, META Group Research Note, 2001.

[19] Kemp S. Digital 2019: Internet Trends in Q3 2019. Available at: https://datareportal.com/reports/digital-2019-internet-trends-in-q3, accessed 06.11.2019.

[20] Kemp S. Digital Trends 2019: Every Single Stat You Need to know About the Internet. Available at: https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-internet/, accessed 06.11.2019.

[21] Reinsel D., Gantz J., Rydning J. The Digitization of the World: From Edge to Core, 2018, Available at: https://www.seagate.com/our-story/data-age-2025/, accessed 06.11.2019.

[22] Hale J.L. More Than 500 Hours of Content Are Now Being Uploaded to YouTube Every Minute. Available at: https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/, accessed 07.11.2019.

[23] Mention.com. 2018 Twitter Report. Available at: https://mention.com/en/reports/twitter/, accessed 07.11.2019.

[24] Wiener J., Bronson N. Facebook's Top Open Data Problems. Available at: https://research.fb.com/blog/2014/10/facebook-s-top-open-data-problems/, accessed 07.11.2014.

[25] Torrecilla J.L., Romob J. Data Learning from Big Data. Statistics and Probability Letters, vol. 136, 2018, pp. 15-19.

[26] Osman A.M.S. A Novel Big Data Analytics Framework for Smart Cities. Future Generation Computer Systems, vol. 91, 2019. pp. 620-633.

[27] Jin X., Wah B.W., Cheng X., Wang Y. Significance and Challenges of Big Data Research. Big Data Research, vol. 2, no. 2, 2015, pp. 59-64.

[28] Reeve A. Chapter 21. Big Data Integration. In Reeve A. Managing Data in Motion, Morgan Kaufmann, 2013, pp. 141-156.

[29] Dhupia B., Rani M.U. Research Challenges in Big Data Solutions in Different Applications. In Social Network Forensics, Cyber Security, and Machine Learning, Springer, 2019, pp. 105-116.

[30] Baig M.I., Shuib L., Yadegaridehkordi E. Big Data Adoption: State of the Art and Research Challenges. Information Processing & Management, vol. 56, no. 6, 2019, article 102095.

[31] Malik S.U.R., Khan S.U., Ewen S.J., Tziritas N., Kolodziej J., Zomaya A.Y., Madani S.A., Min-Allah N., Wang L., Xu C.-Z., Malluhi Q.M., Pecero J.E., Balaji P., Vishnu A., Ranjan R., Zeadally S., Li H. Performance Analysis of Data Intensive Cloud Systems Based on Data Management and Replication: A Survey. Distributed and Parallel Databases, vol. 34, no. 2, 2016, pp. 179-215.

[32] Bellatreche L., Furtado P., Mohania M.K. Guest Editorial: A Special Issue in Physical Design for Big Data Warehousing and Mining. Distributed and Parallel Databases, vol. 34, no. 3, 2016, pp. 289–292.

[33] Lakshman A., Malik P. Cassandra: A Decentralized Structured Storage System. ACM SIGOPS Operating Systems Review, vol. 44, no. 2, 2010, pp. 35-40.

[34] Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM, vol. 51, no. 1, 2008, pp. 107-113.

[35] Rotondi Azevedo D.N., Parente de Oliveira J.M. Application of Data Mining Techniques to Storage Management and Online Distribution of Satellite Images. Studies in Computational Intelligence, vol. 169, 2009, pp. 1-15.

[36] Agrawal D., Abbadi A.E., Antony S., Das S. Data Management Challenges in Cloud Computing Infrastructures. Lecture Notes in Computer Science, vol. 5999, 2010, pp. 1-10.

[37] Buza K., Nagy G.I., Nanopoulos A. Storage-Optimizing Clustering Algorithms for High-Dimensional Tick Data. Expert Systems with Applications, vol. 41, no. 9, 2014, pp. 4148-4157.

[38] Mateus R.C., Siqueira T.L.L., Times V.C., Ciferri R.R., de Aguiar Ciferri C.D. Spatial Data Warehouses and Spatial OLAP Come Towards the Cloud: Design and Performance. Distributed and Parallel Databases, vol. 34, no. 3, 2016, pp. 425–461.

[39] Merino J., Caballero I., Rivas B., Serrano M., Piattini M. A Data Quality in Use Model for Big Data. Future Generation Computer Systems, vol. 63, 2016, pp. 123-130.

[40] McGilvray D.. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. Morgan Kaufmann, 2008, 352 p.

[41] Russom P. Data Quality in the Age of Big Data. Available at: https://tdwi.org/Articles/2019/04/19/DIQ-ALL-Data-Quality-in-the-Age-of-Big-Data.aspx?Page=1, accessed 18.11.2019.

[42] SAS. Data Integration Déjà Vu: Big Data Reinvigorates DI - White Paper. Available at: https://www.sas.com/ru_ua/whitepapers/data-integration-deja-vu-107865.html, accessed 18.11.2019.

[43] FlyData I. The 6 Challenges of Big Data Integration. Available at: https://www.flydata.com/the-6-challenges-of-big-data-integration/, accessed 18.11.2019.

[44] Akusok A., Björk K.-M., Miche Y., Lendasse A. High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications. IEEE Access, vol. 3, 2015, pp. 1011–1025.

[45] Ji C., Li Y., Qiu W., Jin Y., Xu Y., Awada U., Li K.,Qu W. Big Data Processing: Big Challenges and Opportunities. Journal of Interconnection Networks, vol. 13, no. 03 & 04, 2012, article 1250009.

[46] White T. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale, 4 ed. O'Reilly Media, Inc., 2015, 756 p.

[47] Candela L., Castelli D., Pagano P. Managing Big Data through Hybrid Data Infrastructures. ERCIM News, no. 89, 2012, pp. 37-38.

[48] Tao H., Bhuiyan M.Z.A., Rahman M.A., Wang G., Wang T., Ahmed M.M., Li J. Economic Perspective Analysis of Protecting Big Data Security and Privacy. Future Generation Computer Systems, vol. 98, 2019, pp. 660-671.

[49] Tawalbeh L.A., Saldamli G. Reconsidering Big Data Security and Privacy in Cloud and Mobile Cloud Systems. Journal of King Saud University - Computer and Information Sciences, May 2019, 10 p. DOI: 10.1016/j.jksuci.2019.05.007

[50] Kantarcioglu M., Xi B. Adversarial Data Mining: Big Data Meets Cyber Security. In Proc. of the ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 1866-1867.

[51] Cavoukian A., Chibba M., Williamson G., Ferguson A. The Importance of ABAC to Big Data: Privacy and Context. The Privacy and Big Data Institute, Ryerson University, Toronto, Canada, 2015.

[52] Talha M., Kalam A.A.E., Elmarzouqi N. Big Data: Trade-off between Data Quality and Data Security. Procedia Computer Science, vol. 151, 2019, pp. 916-922.

[53] Xu L., Jiang C., Wang J., Yuan J., Ren Y. Information Security in Big Data: Privacy and Data Mining. IEEE Access, vol. 2, 2014, pp. 1149–1176.

[54] Chardin B., Lacombe J.-M., Petit J.-M. Chronos A. NoSQL System on Flash Memory for Industrial Process Data. Distributed and Parallel Databases, vol. 34, no. 3, 2016, pp. 293-319.

[55] Sivarajah U., Kamal M.M., Irani Z., Weerakkody V. Critical Analysis of Big Data Challenges and Analytical Methods. Journal of Business Research, vol. 70, 2017, pp. 263-286.

[56] Ali S. M., Gupta N., Nayak G.K., Lenka R.K. Big Data Visualization: Tools and challenges. In Proc. of the 2nd International Conference on Contemporary Computing and Informatics, 2016, pp. 656-660.

[57] Yang A., Troup M., Ho J.W.K. Scalability and Validation of Big Data Bioinformatics Software. Computational and Structural Biotechnology Journal, vol. 15, 2017, pp. 379-386.

[58] Elgendy N., Elragal A. Big Data Analytics: A Literature Review Paper. Lecture Notes in Computer Science, vol. 8557, 2014, vol. 8557, pp. 214-227.

[59] Shim J.P., French A.M., Guo C., Jablonski J. Big Data and Analytics: Issues, Solutions, and ROI. Communications of the Association for Information Systems, vol. 37, 2015, pp. 797-810.

[60] Gandomi A., Haider M. Beyond the Hype: Big Data Concepts, Methods, and Analytics. International Journal of Information Management, vol. 35, no. 2, 2015, pp. 137-144.

[61] Russom P. Big Data Analytics. TDWI Best Practices Report, Fourth Quarter, 2011. Available at: https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx, accessed 18.11.2019.

[62] Jha A., Dave M., Madan S. A Review on the Study and Analysis of Big Data Using Data Mining Techniques. International Journal of Latest Trends in Engineering and Technology (IJLTET), vol. 6, no. 3, 2016, pp. 94-102.

[63] Berman J.J. Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information. Morgan Kaufmann, 2013, p. 288.

[64] van der Lans R. F. Analytics of Textual Big Data: Text Exploration of the Big Untapped Data Source. Independent Business Intelligence Analyst R20: Consultancy2013. Available at: http://www.data.net.ma/wp-content/uploads/2015/12/Analytics-of-Textual-Big-Data-Text-Exploration-of-the-Big-Untapped-Data-Source.pdf, accessed 18.11.2019.

[65] Malaka I., Brown I. Challenges to the Organisational Adoption of Big Data Analytics: A Case Study in the South African Telecommunications Industry. In Proc, of the Annual Research Conference on South African Institute of Computer Scientists and Information Technologists, 2015, Article No. 27.

[66] Lavalle S., Hopklns M.S., Lesser E., Shockley R., Kruschwitz N. Analytics: The New Path to Value. Research Report, Fall 2010, MIT Sloan Management Review and the IBM Institute for Business Value. Available at: https://sloanreview.mit.edu/projects/analytics-the-new-path-to-value/, accessed 18.11.2019.

[67] Fahmideh M., Beydoun G. Big Data Analytics Architecture Design - An Application in Manufacturing Systems. Computers & Industrial Engineering, vol. 128, 2019, pp. 948-963.

[68] Lopes C., Cabral B., Bernardino J. Personalization Using Big Data Analytics Platforms. In Proc. of the Ninth International C* Conference on Computer Science & Software Engineering, 2016, pp. 131-132.

[69] White C., Research B. Using Big Data for Smarter Decision Making. BI Research, IBM Big Data & Analytics Hub, 2011. Available at: https://www.ibmbigdatahub.com/whitepaper/using-big-data-smarter-decision-making, accessed 18.11.2019.

[70] Samosir R.S., Hendric H.L., Gaol F.L., Abdurachman E., Soewito B. Measurement Metric Proposed for Big Data Analytics System. In Proc. of the International Conference on Computer Science and Artificial Intelligence, 2017, pp. 265–269.

[71] Chapter 2. Business Problems Suited to Big Data Analytics. In Loshin D. Big Data Analytics, Morgan Kaufmann, 2013, pp. 11-19.

[72] Romary L. Data Management in the Humanities. ERCIM News, no. 89, 2012, p. 14.

[73] Lianzhi L. Evaluation Model of Education Service Quality Satisfaction in Colleges and Universities Dependent on Classification Attribute Big Data Feature Selection Algorithm. In Proc. of the International Conference on Intelligent Transportation, Big Data & Smart City, 2019, pp. 645-649.

[74] Li Y., Zhai X. Review and Prospect of Modern Education using Big Data. Procedia Computer Science, vol. 129, 2018, pp. 341-347.

[75] Xiong Z., Zhi L., Jiang J. Research on Art Education Digital Platform Based on Big Data. In Proc. of the IEEE 4th International Conference on Big Data Analytics, 2019, pp. 208-211.

[76] Kim Y.H., Ahn J.-H. A Study on the Application of Big Data to the Korean College Education System, Procedia Computer Science, vol. 91, 2016, pp. 855-861.

[77] Santoso L.W., Yulia. Data Warehouse with Big Data Technology for Higher Education. Procedia Computer Science, vol. 124, 2017, pp. 93-99.

[78] Ramos T.G., Machado J.C.F., Cordeiro B.P.V. Primary Education Evaluation in Brazil Using Big Data and Cluster Analysis. Procedia Computer Science, vol. 55, 2015, pp. 1031-1039.

[79] Huang Y., Chen Z., Yu T., Huang X., Gu X. Agricultural Remote Sensing Big Data: Management and Applications. Journal of Integrative Agriculture, vol. 17, no. 9, 2018, pp. 1915-1931.

[80] Sabarina K., Priya N. Lowering Data Dimensionality in Big Data for the Benefit of Precision Agriculture. Procedia Computer Science, vol. 48, 2015, pp. 548-554.

[81] Klerkx L., Jakku E., Labarthe P. A Review of Social Science on Digital Agriculture, Smart Farming and Agriculture 4.0: New Contributions and A Future Research Agenda. NJAS – Wageningen Journal of Life Sciences, vol. 90-91, 2019, article 100315.

[82] Gonzalez-Sanchez A., Frausto-Solis J., Ojeda-Bustamante W. Predictive Ability of Machine Learning Methods for Massive Crop Yield Prediction. Spanish Journal of Agricultural Research, vol. 12, no. 2, 2014, pp. 313-328.

[83] Senthilvadivu S., Kiran S.V., Devi S.P., Manivannan S. Big Data Analysis on Geographical Segmentations and Resource Constrained Scheduling of Production of Agricultural Commodities for Better Yield. Procedia Computer Science, vol. 87, 2016, pp. 80-85.

[84] Palanisamy V., Thirunavukarasu R. Implications of Big Data Analytics in Developing Healthcare Frameworks – A Review. Journal of King Saud University – Computer and Information Sciences, vol. 31, no. 4, 2019, pp. 415-425.

[85] Patel J.A., Sharma P. Big Data for Better Health Planning, In Proc. of the International Conference on Advances in Engineering & Technology Research, 2014, pp. 1-5.

[86] Pashazadeh A., Navimipour N.J. Big Data Handling Mechanisms in the Healthcare Applications: A Comprehensive and Systematic Literature Review. Journal of Biomedical Informatics, vol. 82, 2018, pp. 47-62.

[87] Abouelmehdi K., Beni-Hssane A., Khaloufi H., Saadi M. Big Data Security and Privacy in Healthcare: A Review. Procedia Computer Science, vol. 113, 2017, pp. 73-80.

[88] Kaur P., Sharma M., Mittal M. Big Data and Machine Learning Based Secure Healthcare Framework. Procedia Computer Science, vol. 132, 2018, pp. 1049-1059.

[89] Khaloufi H., Abouelmehdi K., Beni-hssane A., Saadi M. Security Model for Big Healthcare Data Lifecycle. Procedia Computer Science, vol. 141, 2018, pp. 294-301.

[90] United Nations, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2018 Revision. Available at: https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf, accessed 18.11.2019.

[91] DeRen L., JianJun C., Yuan Y. Big Data in Smart Cities. Science China Information Sciences, vol. 58, no. 10, 2015, pp. 1-12.

[92] Rathore M.M., Paul A., Ahmad A., Chilamkurthi N., Hong W.-H., Seo H. Real-Time Secure Communication for Smart City in High-Speed Big Data Environment. Future Generation Computer Systems, vol. 83, 2018, pp. 638-652.

[93] Rathore M.M., Paul A., Hong W.-H., Seo H., Awan I., Saeed S. Exploiting IoT and Big Data Analytics: Defining Smart Digital City Using Real-Time Urban Data. Sustainable Cities and Society, vol. 40, 2018, pp. 600-610.

[94] Hashem I.A.T., Chang V., Anuar N.B., Adewole K., Yaqoob I., Gani A., Ahmed E., Chiroma H. The Role of Big Data in Smart City. International Journal of Information Management, vol. 36, no. 5, 2016, pp. 748–758.

[95] Lima C., Kimb K.-J., Maglio P.P. Smart Cities with Big Data: Reference Models, Challenges, and Considerations. Cities, vol. 82, 2018, pp. 86-99.

[96] Pal D., Triyason T., Padungweang P. Big Data in Smart-Cities: Current Research and Challenges. Indonesian Journal of Electrical Engineering and Informatics, vol. 6, no. 4, 2018, pp. 351-360.

[97] Allama Z., Dhunny Z.A. On Big Data, Artificial Intelligence and Smart Cities. Cities, vol. 89, 2019, pp. 80-91.

[98] Doku R., Rawat DB. Chapter 8. Big Data in Cybersecurity for Smart City Applications. In Smart Cities Cybersecurity and Privacy, Rawat D.B., Ghafoor K.Z., eds. Elsevier, 2019, pp. 103-112.

[99] Hayes M.A., Capretz M.A. Contextual Anomaly Detection Framework for Big Sensor Data. Journal of Big Data, vol. 2, 2015, article no. 2.

[100] Goswami K., Park Y., Song C. Impact of Reviewer Social Interaction on Online Consumer Review Fraud Detection. Journal of Big Data, vol. 4, 2017, article no. 15,

[101] Shalaginov A., Johnsen J.W., Franke K. Cyber Crime Investigations in the Era of Big Data. In Proc. of the IEEE International Conference on Big Data, 2017, pp. 3672-3676.

[102] Pramanik M.I., Zhang W., Lau R.Y.K., Li C. A Framework for Criminal Network Analysis Using Big Data. In Proc. of the IEEE 13th International Conference on e-Business Engineering, 2016, pp. 17-23.

[103] Hu J. Big Data Analysis of Criminal Investigations. In Proc. of the 5th International Conference on Systems and Informatics, 2018, pp. 649-654.

[104] Vaughan G. Efficient Big Data Model Selection with Applications to Fraud Detection. International Journal of Forecasting, June 2018, https://doi.org/10.1016/j.ijforecast.2018.03.002.

[105] Khan E.S., Azmi H., Ansari F., Dhalvelkar S. Simple Implementation of Criminal Investigation Using Call Data Records (CDRs) Through Big Data Technology. In Proc. of the International Conference on Smart City and Emerging Technology, 2018, pp. 1-5.

[106] Zhao Q., Chen K., Li T., Yang Y., Wang X. Detecting Telecommunication Fraud by Understanding the Contents of A Call. Cybersecurity, vol. 1, no. 8, 2018, p. 12.

[107] Chen Y.-J., Wu C.-H. On Big Data-Based Fraud Detection Method for Financial Statements of Business Groups. In Proc. of the 6th IIAI International Congress on Advanced Applied Informatics, 2017, pp. 986-987.

[108] Makki S., Assaghir Z., Taher Y., Haque R., Hacid M.-S., Zeineddine H. An Experimental Study with Imbalanced Classification Approaches for Credit Card Fraud Detection. IEEE Access, vol. 7, 2019, pp. 93010-93022.

[109] Zhou H., Sun G., Fu S., Jiang W., Xue J. A Scalable Approach for Fraud Detection in Online E-Commerce Transactions with Big Data Analytics. CMC: Computers, Materials & Continua, vol. 60, no. 1, 2019, pp. 179-192.

[110] Herland M., Khoshgoftaar T.M., Bauder R.A. Big Data Fraud Detection Using Multiple Medicare Data Sources. Journal of Big Data, vol. 5, 2018, article no. 29.

[111] Castaneda G., Morris P., Khoshgoftaar T.M. Maxout Neural Network for Big Data Medical Fraud Detection. In Proc. of the IEEE Fifth International Conference on Big Data Computing Service and Applications, 2019, pp. 357-362.

[112] Castaneda G., Morris P., Khoshgoftaar T. M. Evaluation of Maxout Activations in Deep Learning Across Several Big Data Domains. Journal of Big Data, vol. 6, 2019, article no. 72.

[113] Lnenicka M., Komarkova J. Developing A Government Enterprise Architecture Framework to Support the Requirements of Big and Open Linked Data with the Use of Cloud Computing. International Journal of Information Management, vol. 46, 2019, pp. 124-141.

[114] Yang P., Xia H., Liu W., Li Z. Research on Government Integrity Evaluation Based on Big Data. In Proc. of the 2nd International Conference on Artificial Intelligence and Big Data, 2019, pp. 28-35.

[115] LaBrie R.C., Steinke G.H., Li X., Cazier J.A. Big Data Analytics Sentiment: US-China Reaction to Data Collection by Business and Government. Technological Forecasting and Social Change, vol. 130, 2018, pp. 45-55.

[116] Laude H. Chapter 6. France's Governmental Big Data Analytics: From Predictive to Prescriptive Using R. In Federal Data Science: Transforming Government and Agricultural Policy Using Artificial Intelligence, Batarseh F.A., Yang R., eds. Academic Press, 2018, pp. 81-94.

[117] Yan Z. Big Data and Government Governance. In Proc. of the International Conference on Information Management and Processing, 2018, pp. 111-114.

[118] Aron J.L., Niemann B. Sharing Best Practices for the Implementation of Big Data Applications in Government and Science Communities. In Proc. of the IEEE International Conference on Big Data, 2014, pp. 8-10.

[119] Hardy K., Maurushat A. Opening up Government Data for Big Data Analysis and Public Benefit. Computer Law & Security Review, vol. 33, no. 1, pp. 30-37.

[120] Archenaa J., Anita E.A.M. A Survey of Big Data Analytics in Healthcare and Government. Procedia Computer Science, vol. 50, 2015, pp. 408-413.

[121] Lee Y., Park S. Design of A Government Collaboration Service Map by Big Data Analytics. Procedia Computer Science, vol. 91, 2016, pp. 751-760.

[122] Amado A., Cortez P., Rita P., Moro S. Research Trends on Big Data in Marketing: A Text Mining and Topic Modeling Based Literature Analysis. European Research on Management and Business Economics, vol. 24, no. 1, 2018, pp. 1-7.

[123] [Saidali J., Rahich H., Tabaa Y., Medouri A. The Combination Between Big Data and Marketing Strategies to Gain Valuable Business Insights for Better Production Success. Procedia Manufacturing, vol. 32, 2019, pp. 1017-1023.

[124] Akter S., Wamba S.F. Big Data Analytics in E-Commerce: A Systematic Review and Agenda for Future Research, Electronic Markets, vol. 26, no. 2, 2016, pp. 173-194.

[125] Chong A.Y.L., Li B., Ngai E.W.T., Ch'ng E., Lee F. Predicting Online Product Sales Via Online Reviews, Sentiments, and Promotion Strategies: A Big Data Architecture and Neural Network Approach. International Journal of Operations & Production Management, vol. 36, no. 4, 2016, pp. 358-383.

[126] Erevelles S., Fukawa N., Swayne L. Big Data Consumer Analytics and the Transformation of Marketing. Journal of Business Research, vol. 69, no. 2, 2016, pp. 897-904.

[127] Jabbar A., Akhtar P., Dani S. Real-time Big Data Processing for Instantaneous Marketing Decisions: A Problematization Approach. Industrial Marketing Management, Sept. 2019, https://doi.org/10.1016/j.indmarman.2019.09.001.

[128] Li T. Using Big Data Analytics to Build Prosperity Index of Transportation Market. In Proc. of the 4th ACM SIGSPATIAL International Workshop on Safety and Resilience, 2018, no. 17, p. 6.

[129] See-To E.W.K., Ngai E.W.T. Customer Reviews for Demand Distribution and Sales Nowcasting: A Big Data Approach. Annals of Operations Research, vol. 270, no. 1-2, 2018, pp. 415–431.

[130] Kumar A., Shankar R., Aljohani N.R. A Big Data Driven Framework for Demand-driven Forecasting with Effects of Marketing-mix Variables. Industrial Marketing Management, June 2019, https://doi.org/10.1016/j.indmarman.2019.05.003.

[131] Zheng K., Zhang Z., Song B. E-Commerce Logistics Distribution Mode in Big-Data Context: A Case Analysis of JD.COM. Industrial Marketing Management, Oct. 2019. DOI: https://doi.org/10.1016/j.indmarman.2019.10.009.

[132] Salehan M., Kim D.J. Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach to Big Data Analytics. Decision Support Systems, vol. 81, 2016, pp. 30–40.

[133] Malhotra D., Rishi O.P. An Intelligent Approach to Design of E-Commerce Metasearch and Ranking System Using Next-Generation Big Data Analytics. Journal of King Saud University - Computer and Information Sciences, Mar. 2018, https://doi.org/10.1016/j.jksuci.2018.02.015.

[134] Wu P.-J., Lin K.-C. Unstructured Big Data Analytics for Retrieving E-Commerce Logistics Knowledge. Telematics and Informatics, vol. 35, no. 1, 2018, pp. 237-244.

[135] Zhaoa Y., Xu X., Wang M. Predicting Overall Customer Satisfaction: Big Data Evidence From Hotel Online Textual Reviews. International Journal of Hospitality Management, vol. 76, 2019, pp. 111-121.

[136] Liu X., Shin H., Burns A.C. Examining the Impact of Luxury Brand's Social Media Marketing on Customer Engagements: Using Big Data Analytics and Natural Language Processing. Journal of Business Research, May 2019, https://doi.org/10.1016/j.jbusres.2019.04.042.

[137] Kauffmann E., Peral J., Gil D., Ferrández A., Sellers R., Mora H. A Framework for Big Data Analytics in Commercial Social Networks: A Case Study on Sentiment Analysis and Fake Review Detection for Marketing Decision-making. Industrial Marketing Management, Aug. 2019, https://doi.org/10.1016/j.indmarman.2019.08.003.

Информация об авторах / Information about authors

Ноаман Мухаммед АЛИ в 2016 году получил степень магистра на факультете компьютерных наук Каирского университета. С 2016 года Ноаман является ассистентом на кафедре информационных технологий и систем университета Порт-Саида, Египет. Ноаман является также аспирантом кафедры информатики Санкт-Петербургского государственного университета. Его научные интересы включают анализ больших данных, распознавание образов, системы рекомендаций, обработку естественного языка.

No'aman Muhammad ALI received his M.Sc. degree from the department of computer science, Cairo University, in 2016. Currently, No'aman is an assistant lecturer at the Information Technology & Systems Department, Port Said University, Egypt, since 2016. No'aman is a Ph.D. student at the Department of Computer Science, Saint Petersburg State University. His research interests involve Big data analytics, pattern recognition, recommender systems, natural language processing.

Борис Асенович НОВИКОВ – доктор физико-математических наук, профессор, кафедра информатики в НИУ ВШЭ, Санкт-Петербург, Россия. Сфера научных интересов - широкая область управления данными и их анализа, включая системы и приложения для управления базами данных, структуры данных и методы доступа, обработку запросов, контроль параллелизма, обработку и анализ дискретных потоков, а также приложения машинного обучения.

Boris Asenovitch NOVIKOV, Dr. Sci. in mathematics and physics, professor, Department of Informatics at National Research University Higher School of Economics, Saint Petersburg, Russia. Research interests are in a wide area of data management and analytics, including database management systems and applications, data structures and access methods, query processing, concurrency control, discrete stream processing and analytics, and applications of machine learning.