UDK: 004.6-022.59

10.58898/ijmt.v4i1.47-57

BIG DATA TREND IN SOFTWARE DEVELOPMENT, IN MEDIA AND COMMUNICATION, IN THE DEVELOPMENT OF ECONOMY AND SOCIETY

Tatjana Davidov[1*], Maja Vojinovic[2]

[1] Akademija Dositej, Beograd, Serbia, tanja.davidov1@gmail.com

[2] Faculty of Management, Sremski Karlovci, Serbia, vojinovic@famns.edu.rs

Abstract: With the development of society, the economy, the development of innovative technology and technological equipment, the development of modern web software solutions, there is a constant tendency to increase the amount, variety and quality of data. The need to accommodate a huge amount of data implies the use of new technologies that are able to process new data dimensions. Virtually all participants in the social system, from large web companies to traditional entrepreneurs, researchers and scientists, are facing a disproportionate and lightning-fast growth in the amount of data (Jinchuan CHEN, 2013). The available data is of diverse content and format, and the method of their processing, generation and analysis is a challenge for modern Big Data technologies, which we will deal with in this paper. Big Data technologies create new opportunities for exploiting the value and potential of increasingly extensive and diverse data. This paper describes Big Data technologies, the reasons for their application and the challenges of big data from the point of view of data management. The technology of storing, hosting, mining and manipulating big data is described in several dimensions supported by Big Data. We compared the role and way of using traditional databases, versus Big Data technologies that take on a role in the economy and society, in web companies or classic commercial and non-commercial organizations. In this regard, it was pointed out the irreplaceable need for the implementation of Big Data technologies in all economic entities and the entire social system, as well as their role on the Internet in the development of social networks.

Keywords: Big Data, traditional databases, communication, software system, Hadoop technology.

Introduction

Today, large data systems have an important role and an even more important influence on the development of economic and non-economic activities, as well as all categories and structures of social development. A large system is a complex entity and involves a variety of skills: technique, methodology, technology, infrastructure, organization, knowledge, connectivity, accuracy, utility, and more. A great system means great responsibility.

*Corresponding author: tanja.davidov1@gmail.com

In front of such systems, the question is constantly raised: in what way will the modern system be able to improve its scalability, technical readiness and operability, so that such a system would support the improvement of business based on the development of the business environment? Technologies significantly help and participate in the development of businesses. Development implies knowledge, information, and data. The higher the quality of the information, the more substantial the knowledge. The system must necessarily support the use of data and work with data at all levels and in all data. The more perfect the technology for working with data, the more capable and cost-effective the system is. With the development of technologies, the content, amount and range of data increases. There is a growing need for a system that will support the processing of a huge amount of data, so that the data becomes useful information for the area in which the business activity of the company will develop and improve. Only companies that have to the ability to accommodate, develop and support a huge amount of scattered data have a high degree of certainty to survive and compete in the modern market game. Flexible work with big data must be the goal of every company or institution of modern society. This work is dedicated to the opportunities offered by modern technologies that work with big data. Technology and its influence on the future development of economic subjects, the health system and other physical subjects will be presented, which is multidimensionally useful to the social community.

BigData technologies are the key to business

Technologies are a reflection of the development of society, companies, and individuals. Depending on how developed technologies are, it is possible to develop the system, the environment, and the individual. Finally, technologies have a significant economic relationship to the system in which they exist. First, modern technologies cost more and more, because it means investing in equipment, knowledge, and personnel. However, it gives results in the long term, such business conditions provide support for accelerated and increasingly complex development. Developed technologies participate in maximizing profits and minimizing costs, with an emphasis on product or service quality, with one very simple goal: to make the end user satisfied. Today, current big data technologies are becoming an almost irreplaceable parameter in the development of a software system in any serious company or institution. Big data has become an innovative but also a necessary trend. Big data technologies are able to create, process, formalize and generate increasingly demanding and detailed data and information in all areas. Backend software data is increasingly complex and requires more and more information. The information that is accepted from the outside world is less and less formalized, that is, it has more and more scattered content. The data, and in the last instance, the generated information, are more and more comprehensive and voluminous.

The tremendous growth of data has improved the infrastructure in the information technology (IT) industry. Traditional relational databases have become insufficiently flexible and too formalized for informal, overly voluminous, unstructured data. Unusual data cannot be placed in a formally controlled model, such as the Relational Data Model, on which the development of relational databases in current IT practice is based. In the field of data storage, IT technologies are being developed according to a new and completely different approach, which is called: a new dimension of databases, i.e., Big Data. Big data

technologies have become an irreplaceable category for processing these overly voluminous amounts of data.

The simplest definition of Big Data was given by the well-known researcher Dumbill, who says that Big Data refers to information that cannot be processed or analyzed in a traditional way, using conventional processes and tools (e.g. relational databases) (Edd Dumbill, 2013). Big Data is large-scale data, of different content and form. The distribution of that data implies the use of new technologies of contemporary architecture, where the data is viewed statically, content-wise, functionally, temporally (Nathan Marz & James Warren, 2015). The concept of Big Data also means the visualization and organization of large amounts of data. The focus of the data is on the user - the consumer. Such data are of different forms, structures, and contents. Organizational alignment and visualization is a part of Big Data analytics, which refers to the graphical presentation of the results of processing a large amount of data, with the aim of easier and more accurate interpretation of the results obtained. (Brad Brown, Johnson Sikes, & Paul Willmott, 2013). Data analysis graphs have a wide range of applications both in academia and industry. They can be used to model parts of human knowledge and are called semantic networks. Graphs for Big Data analysis dominate all platforms of technology giants, such as Facebook, LinkedIn, Twitter, X, as well as many other companies, which present, model and process their data as graphs (Brij B Gupta & Mamta, 2024).

BigData develops technology in 5 dimensions

Big Data technology is a challenge of modern technological development. Databases are built on mathematical set theory. Thus, Big data represents a technology based on sets of data that synchronize several dimensions. It is estimated that these data sets exceed the sizes and dimensions that can be implemented in standard software solutions for data storage, management and processing in an acceptable amount of time. These complex data sets are implemented in Big Data dimensions that support five important Big Data concepts that go beyond the capabilities of current data processing technologies:

1. Volume,
2. Velocity i
3. Variety.
4. Veracity and
5. Value

The five main "dimensions" of big data technology are usually presented as a formula: 5V Big Data technology. This technology is used today in solving real problems of large amount of data in real time flow.

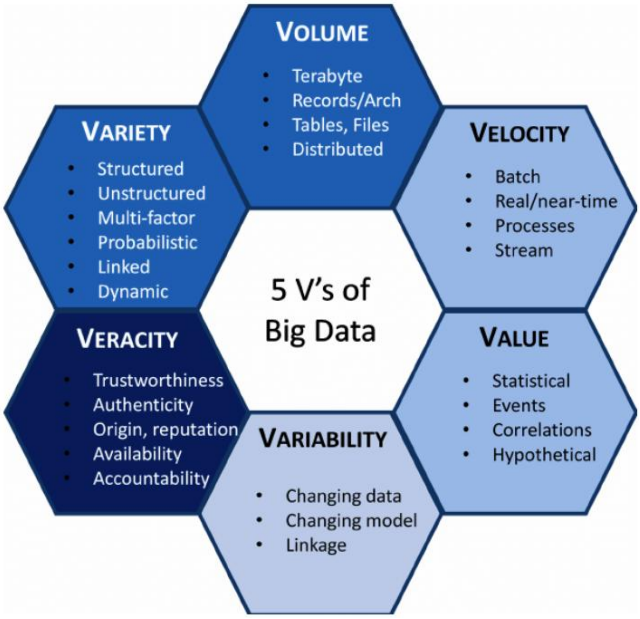

Figure 1. The five V's of Big Data (Jose Andre Moura & Carlos Serrão, 2015)

Volume of data in Big Data technology

A challenge such as the volume of data is actually the most significant in the development of BigData technology. In sciences such as biology, meteorology, astronomy, etc., there is a constant need to increase computing performance that meets the possibility of storing increased volumes of data. Today, in a technological sense, data is informalized, presented in graphs, images, and large texts, which generate an ever-increasing volume, and the management of such data is increasingly complex, difficult and demanding. Social networks manage a huge amount of data of complex structure (Jinchuan CHEN, i drugi, 2013). Communication between individuals requires the reception of large data in the shortest possible time so that communication can be realized in a timely and high-quality manner.

Velocity of data in Big Data technology

The speed of processing large amounts of informal data is a major challenge for technologies. This challenge was generated due to the need to solve the automation of new data entry, where the speed of their processing is of great importance. Also, updating existing data is very important to process quickly and efficiently. The issue of input and updating is particularly related to data that is generated on the computer during the work process, such as that created by sensors or mobile devices. These are specific data formats that Big Data technologies recognize and process quickly.

Variety of data in Big Data technology

Application software that is used in a real environment, accepts input and generates output data. Input data generally does not come from a single source, nor are the forms of that data unambiguous and uniform. Big data implementations require handling data from different sources, which are often of different formats and models. Large amounts of data speak to the specificity and size of the data entering

the system. First of all, they are of very large dimensions, such as videos, images, recordings from various technical devices (magnetic resonance, monitoring system, sensor, various software tools, etc.). The different dimensions of data generate the problem of their acceptance: how to "capture" different types of data in a way that enables recognition and connection of their meaning for the purpose of generating information. Typically, diverse data can be classified into three general types of data:

1. structured,
2. semi-structured i
3. unstructured data.

The distribution of data and its processing in the big data system is shown in the figure. According to the steps shown in the picture, the data are integrated, their meaning is unified. Then there is a reduction of unnecessary data and the final presentation of the cleaned data on which queries are made and their indexing, which increases the speed of the search and obtaining quality information.

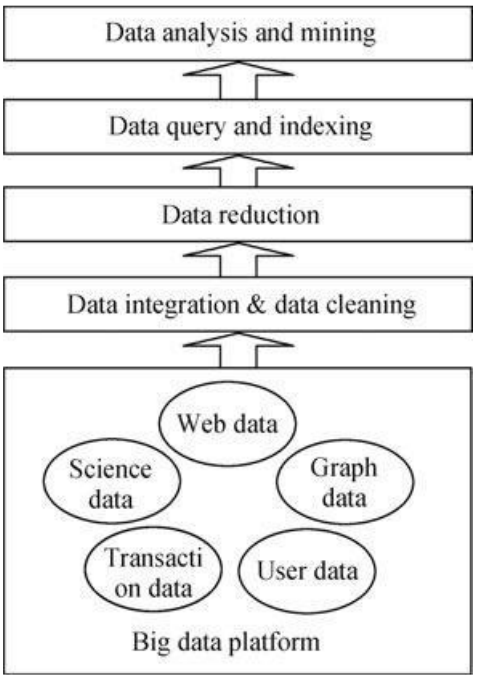

Figure 2 Big data processing steps (Jinchuan CHEN, i drugi, 2013)

Databases - relational or BigData

The data is formatted in the database within the framework of static formats of the relational database. Traditionally, data mainly consists of static paper documents in text formats or is formally structured so that it can easily be stored in a traditional relational database system. However, in recent times, the storage of data in uniform, traditionally structured formats have become insufficiently flexible and very limited. Data is increasingly bulky and is created in the form of different formats, given that it is formed by various new and modern technological devices. In Big Data databases, such data can be recognized and become usable for processing. The presentation of such data is the result of an

increasingly studious analysis of a phenomenon, information about that phenomenon is analyzed by the depth of insight, using Big Data technology. For example, if the manufacturer changes the attributes, this means adding product information by the depth of the product data analysis. The information is even more complete, because Big Data helps to chronologically store all new data, documents and information about the product. Thus, the information could be properly analyzed. As information accumulates, organizations can see trends, patterns, and the impact of change over time (Zeituni Baraka, 2014). Big Data technology takes over the role of storing and perceiving information, which until now was played by the relational database. Given that information is more and more complex, and data requires more and more dimensions, it exceeds the role and capabilities of traditional relational database technology (Tatjana Davidov, Saša Virijević Jovanović, Ilija Subotić, Maja Vojinović, & Jelena Zdravković, 2022).

Table 1

Basic differences between the characteristics of traditional databases and the characteristics of big data (Zeituni Baraka, 2014).

	Traditional DBMS characteristics	Big data database characteristics
Data characteristics	• Weak in handling non-structured data • Does not learn from user access behavior	• Real time live data • Environment supports all types of data and from all types of sources • Appropriate for handling petabytes and exabytes of data • Learns from user access behavior
Considerations for analytics	• Is appropriate for analysis of data containing information that will answer to information gaps that are known • Stable data structure • Passive user experience- the user simply retrieves information • Focus on attribute search • Data management simplified • Historical data management 2 dimensional • Limited inter organizational data access • Analytical capability expectations are minimal	• A lot of parallel processing often strains supporting systems • Can be truly ground-breaking for organization's as previously completely unknown gaps of information can be revealed randomly rather than just providing information about what is known- is not know • Emergent data structure • Active user experience- the system may initiate discourse that may need attention • Focuses on historical pattern, trend, multi-dimensional search • Rich multi-channel knowledge access • Real-time analytics technologies plays a vital role • Analytical capabilities expectations may be too high
Relevant technologies	• Implementation straight forward • SQL is most common language • Relational database function model • Not open source • Analytics is done on batch jobs containing aggregated data which is historical data rather than real-time data	• Implementation more complex due to lack of industry standards and direction • No SQL but near on SQL compliant • Hadoop framework • Open Source is appropriate • Stream processing capability is often relevant

Structured data has a clearly defined type, format and structure, and therefore belong to traditional data structures generated by statistical methodology. Structured data can be classified into the category of metric or numerical variables whose processing, analysis and interpretation is very precisely defined and relatively simple (Kelly Rainer & Efraim Turban, 2008). However, the problems that appear indicate that the data is more voluminous and its storage, preservation and recognition become more complex. It is not possible to structure such data so that it can be easily and simply added according to the format of traditional relational databases

In today's conditions of technological development, the growth of all types of data is constantly and certainly evident, with the difference that the generation of structured data follows a linear trend, unlike unstructured data whose growth is exponential. Traditional infrastructures and analytics platforms cannot keep up with this diversity, where mostly unstructured, quasi-structured and semi-structured data appear. This kind of data can basically be seen as nominal variables. They require different and more demanding methods of processing and analysis than structural variables. Therefore, techniques of advanced statistical analysis (cluster analysis), methods of artificial intelligence (machine learning), and Data Mining are most often applied in practice (Jure Leskovec, Anand Rajaraman, & Jeffrey David Ullman, 2014).

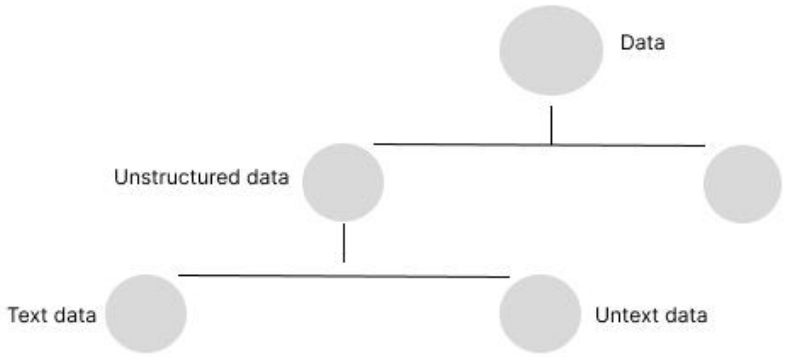

Figure 3. Structured and unstructured data (source: author)

A branch of **unstructured** data and textual data, or non-textual data, such as graphics, images, videos, sensor formats, etc. all together become complicated and complex to process, and that branch is directed towards the development of new technologies for storing and processing data, BigData.

Unstructured data - Big Data

Unstructured data is basically information that either does not have a predefined data model or does not fit well into a traditional database: This data includes data presented as:

- text,
- pdf document,
- video,
- pictures,
- audio,

- geospatial data,
- internet data,
- slick streams,
- log files
- x-rays
- data of biometric and monitoring devices
- data generated by sensor devices, etc.

The role of Big Data in society, economic and non-economic branches

The term Big Data refers not only to data, but also to technologies that ensure their purposeful processing. Hadoop, currently the most famous Big Data Technology, is presented in the form of a distributed system for storing and processing large amounts of structured and unstructured data. Apache Hadoop is an open-source distributed storage and processing system. Apache Hadoop is an open-source system for distributed storage and processing of diverse data, created at a time when the amount of data exceeds the memory capacity of computer systems. A significant role in Big Data is played by Hadoop, NoSQL databases, Apache spark, etc. (Vernes Vinčević & Mahir Zajmović, 2021). With the development of technological achievements and technical devices that generate unstructured data formats, the need for data processing through Big Data technologies is created in companies. These complex data structures are not supported by traditional software products for processing data. They cannot download them, manage them or process them in a reasonable amount of time. How much data can one of the Big Data technologies accept or process? Most of the time it is not very clear, but usually it is about several terabytes, petabytes and even exabytes.

Every company has a need to process data as quickly and consistently as possible. Because if Big Data is the support for a product to be advertised, clearly, in-depth and detailed, graphically, then surely Big Data must studiously accept, accommodate, process and analyze a very scattered amount of data about the product. With this analysis, any information about the product is above all fast, accurate, clear, precise, unambiguous and presented in the most consistent way. Such numerous and varied information about a product can so clearly and graphically present and bring the product closer to the future customer that the sale of such a product could be much faster and more efficient, even up to 100% more

Big Data is represented on social networks, and this is actually the first and initial role of big data in communication with users. However, everything that surrounds us is turning into a global planet of producers and consumers of products and services. The more detailed and purposeful the description of both the raw material and the finished product, the more clearly some information is presented, the more details there are about an item, service, product, project, the more attractive the given object will be to the person for whom it is intended. Data is important. Data is all around us and Big Data technologies are the future of all technologies.

If we look at services, we can highlight services that are important to people's lives, such as the health system. In this regard, it is necessary to include a significant development of technologies and many professional experts in the development of an expert system for the needs of society. Big Data has an irreplaceable role in which new healthcare systems must be built. There is more and more data in the current

healthcare software, so the old technologies are not able to support it. Only realistic and accurate information in health care can have an impact on saving human life, if life is in danger. The speed, variety, size, accuracy, depth of analysis, precision and comprehensiveness of the information are of key importance here, which provides a sense of security in case of making important decisions for the patient's life or health.

The speed of storage, recognition, classification, structuring, processing, generation of information and reporting from the set of obtained results is unthinkable without Big Data technologies. Today, in society, every social phenomenon or natural analysis has become dependent on these technologies, such as weather forecasting. At the state level, state institutions can use Big Data technology to get the safest and most reliable information possible. If we look at various branches of social development, for example agriculture: how much only Big Data technologies would help in decision-making and the introduction of various modern systems into processes and workflows. The development of a software solution based on Big Data technology would help many activities: cultivation, sowing, storage, accommodation, purchase, sale, advertising or any segment of the production process of that economic branch.

Hadoop - BigData technology

Hadoop is a software framework derived from two prominent Big Data technologies: MapReduce and BigTable. Hadoop allows MapReduce-based applications to manage large clusters on user hardware. This project represents the foundation of the computer architecture of Yahoo! business processes. Hadoop is designed to speed up information processing and computation by processing across computing nodes, as well as to hide latency.

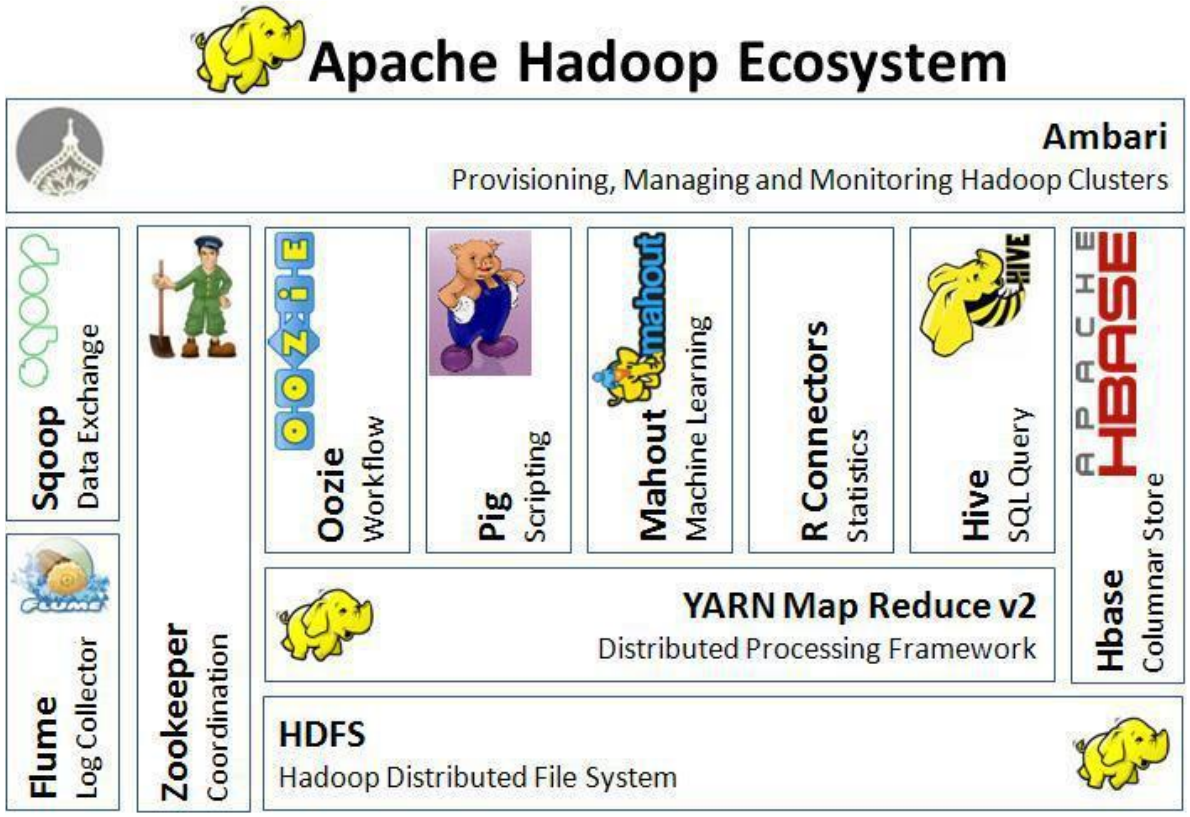

Figure 4: Apache Hadoop ecosystem (Stipe Batinica, 2018).

As already mentioned, Big Data technology is developing to process data of undefined sizes and to do the same at a disproportionate speed. Once technology master the way to analyze and "mine" large amounts of data, it can bring great value for the funds invested in Big Data. Hadoop technology processes data, but at the same time it faces great challenges. Given that unstructured data has dynamic, heterogeneous, interconnected, disconnected and sometimes insufficiently reliable properties, analyzing and "mining" big data is very challenging.

Conclusions

Big Data is becoming an important branch of IT industry development. Soon these databases will assume the role of primary databases. Under these conditions, in the near future, relational databases will only sometimes be used on certain systems where it is necessary due to the accuracy of information that, for example, refers to financial or similar data, which are related to legal frameworks. With this data, their form and structure are of paramount importance. The use of big data to guide business activities, production processes, social and government decisions, where people have a primary role, requires that "deep data analysis" be carried out, instead of generating it with simple forms and formulas for reports. This type of complex analysis must depend on complex models and could hardly be expressed in a simple SQL language, which is applied to structured data. In order to make some active preparations, people must actively participate in the storage, analysis, generation and completion of big data that are processed by Big Data technologies. People must be an active parameter and factor in the creation of information, both through their work and mutual relationships and communication. People who participate must know what is happening now but also predict what will happen in the future. Users must analyze the data at their disposal with the help of modern technologies and modern software systems, or expert systems based on Big Data technologies.

References

Brad Brown, Johnson Sikes, & Paul Willmott. (2013). Bullish on digital: McKinsey Global Survey. Retrieved from https://www.mckinsey.de/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Bullish%20on%20digital%20McKinsey%20Global%20Survey%20results/Bullish%20on%20digital.pdf

Brij B Gupta, & Mamta. (2024). Big Data Management and Analytics (Vol. 1). World Scientific Publishing Co. Pte. Ltd. doi:10.1142/12869

Curry, E. (2016). New Horizons for a Data-Driven Economy. National University of Ireland Galway. doi:10.1007/978-3-319-21569-3

Edd Dumbill. (2013). Big Data A Revolution That Will Transform How We Live, Work, and Think (Vol. 1). doi:10.1089/big.2013.0016

Girdharee Saran . (2025). What is a BigTable? The Architecture of BigTable. Whizlabs Software Pvt. Ltd. All rights reserved. Retrieved from https://www.whizlabs.com/blog/what-is-a-bigtable/

Jinchuan CHEN, ueguo CHEN, Xiaoyong DU, Cuiping LI, Jiaheng LU, Suyun ZHAO, & Xuan ZHOU. (2013). Big data challenge: a data management perspective. Frontiers of Computer Science, 7(2), 157–164. doi:10.1007/s11704-013-3903-7

Jose Andre Moura, & Carlos Serrão. (2015). Security and Privacy Issues of Big Data. doi:10.4018/978-1-4666-8505-5.ch002

Jure Leskovec, Anand Rajaraman, & Jeffrey David Ullman. (2014). Mining of Massive Datasets. Cambridge University.

Kelly Rainer, & Efraim Turban. (2008). Introduction to Information Systems: Enabling and Transforming Business. Introduction to Information Systems.

Map-Reduce Jobs. (2015). Conceptual Overview of Map-Reduce and Hadoop. glennklockwood. Retrieved from https://www.glennklockwood.com/data-intensive/hadoop/overview.html

Nathan Marz, & James Warren. (2015). Big Data: Principles and best practices of scalable. Manning. Retrieved from https://www.amazon.com/Big-Data-Principles-practices-scalable/dp/1617290343#

Stipe Batinica. (2018). Primjena big data tehnologije u zdravstvu. VUS REPOSITORY - Repozitorij završnih radova. Retrieved from https://urn.nsk.hr/urn:nbn:hr:143:871676

Tatjana Davidov, Saša Virijević Jovanović, Ilija Subotić, Maja Vojinović, & Jelena Zdravković. (2022). Information and communication technologies and e-marketing on the Big Data database in the function of sustainable development strategy. 29, 306-314. doi:10.18485/ecologica.2022.29.107.2

Vernes Vinčević, & Mahir Zajmović. (2021). Big Data koncept i primena u savremenom kriznom poslovanju. Sveučilište/Univerzitet “VITEZ” Fakultet informacionih tehnologija. Retrieved from https://hrcak.srce.hr/file/389274

Zeituni Baraka. (2014). Opportunities to manage big data efficiently and effectively. Retrieved from https://esource.dbs.ie/server/api/core/bitstreams/6af4c720-195e-4cdb-b8d4-fba01e5ebe79/content