

Milan Kubina - Gabriel Koman*

BIG DATA TECHNOLOGY AND ITS IMPORTANCE FOR DECISION-MAKING IN ENTERPRISES

Rapid development in the field of information-communication technologies, which has been recorded in recent years, caused the increase from 40 to 50% in the volume of data in companies last year. At present enterprises are able to use only a quarter of the data available. It is expected that by 2020, the volume of data will have reached 35 zettabytes [1]. By analyzing large amounts of data, it is possible to get the information important for an enterprise which will facilitate decision-making process of managers. Technology for large volume data called Big Data is nowadays largely employed for effective use of data.

Keywords: *Big Data, Technology, Management, Decision-making.*

1. Introduction

The occurrence of large amounts of data was recorded in the 1970s, when researchers from Berkeley identified that there should be more than 1.5 trillion bits of information generated by 1999. Subsequently, in 2003, the repetitive study showed that the amount of information would be doubled within three years. Thus it was established that the amount of data continues to grow. The basis for the definition of large data (Big) in 2001 was formulated by an analyst Doug Laney, who described the ever-increasing amount of data, i.e. diversity, volume and speed. The problem of large amounts of data was firstly introduced in 1997, when NASA scientists identified the problem with the graphics as data visualization, this amounted to such a size that the data files could not be placed in computer's memory or on the hard drive of the computer [2] and [3].

2. Big Data

In terms of information-communication technologies Big Data (BD) represents rapidly changing and expanding field. Big data is associated primarily with modern, trendy technologies that generate a wealth of data, where appropriate, the amount of data used, as for example, semantic technology, sound processing and voice, whether or not the internet of things. For this reason, it is very difficult to clearly define the concept of Big Data. Other reasons are also temporal changes in the meaning

of rapid technological development, i.e. data, which represented large data-intensive processing in the 1990s, can be handled by common computers or mobile devices just because of the influence of developed technology [4].

As it is not clearly defined there is a number of definitions of data, from which it is possible to mention the following:

Big data represents a loosely defined term that describes a large amount of complex data sets, and, at the same time, it describes advanced technology for the collection and storage of the data [5].

Big Data is currently a modern way of working with information, the importance of which depends on the number of enterprises. The very concept of Big Data can be misleading, as shows the survey conducted by IBM, in which 18% of Directors see Big Data just as a larger amount of data, and 8% as a new term, or a term to describe the amount of data [6].

In general, it is possible to talk about big data in three basic senses. These data consist of three different types - data too bulky to be processed in a reasonably short period of time with the appropriate performance (e.g. information about the gravitational potential of each star located in two galaxies, which are currently subjected to collision), flat data (e.g. searching for information within text, video or audio files), and real time data of the process, (such as traffic information from thousands of cameras and satellites) [7].

Big Data constitutes data of very large size, typically up to such an extent that their handling and management brings significant logistical challenges [7].

* Milan Kubina, Gabriel Koman

Department of Management Theories, Faculty of Management Science and Informatics, University of Zilina, Slovakia
E-mail: Milan.Kubina@fri.uniza.sk

Big Data shall relate to data files whose size is within the current database systems and it captures storage and processes capabilities over their analysis [8].

Large data relate to large amounts of data collected over time and are difficult to analyze and handle using ordinary tools to manage a single database. Analyzed data include the data in the field of marketing trends, as well as in the field of production, medicine and science. Types of input data are, e.g., commercial transactions, e-mail messages, photos, camera records, logs of activities, unstructured text, for instance from blogging and social media, as well as a large amount of data, which are generated by a variety of sensors [9].

Big Data is a term describing exponential growth and availability of structured as well as unstructured data. Big Data can have the same meaning for business and society as the Internet at present, but these data may lead to more accurate analyses [10].

The term Big Data refers to a large amount of information coming from a variety of sources, such as transaction records, boot files, social media, sensors, third parties, Web applications, etc. Big Data, however, is not only a large amount of data, but also extremely varied type of data, distributed at different speeds and frequencies [11].

Big Data is generated everywhere around us at all times. It is created by each digital process and the exchange of information through social media. It also generates different systems, sensors and mobile devices. Big Data comes from multiple sources at very great speed, in large volume and wide diversity. In order to obtain meaningful values from the major data, it is essential to feature optimum computing power, analytical abilities and skills [12].

Big Data can be defined in situations where sets of data are so great that traditional technology, technique and tools for their extraction are no longer usable within a reasonable time frame, and they are also ineffective due to the cost [13].

Big Data represents an ocean of information, in which we swim every single day. These are the extensive data from our computers, mobile devices and zettabytes machine sensors. The right solution will allow organizations to delve into all of the data and to gain valuable insights, which were previously unthinkable [14].

The concept of big data generates data that are too big, too demanding or too quick for processing of existing instruments. Too big means that enterprises must constantly deal with the petabytes data, which come from reports, transaction systems, sensors, etc.. Too quick data for processing of existing systems designate detection of frauds at the point of sale or detection of ads which should be provided to users on a website. Too demanding data means that processing of data by using existing tools for particular analyses is complicated and difficult and it cannot be provided [15].

The term Big Data defines large volumes of data, which are so complex that traditional data processing systems have problems to

work with. Three main problems in large data include a number of data collections, speed at which it is necessary to analyze the data, and a variety of collected data formats [16].

Big Data is a collection of large amounts of disparate information, usable for an understanding of the environment, medicine and the human experience [17].

The concept of Big Data currently defines the broad use of data obtained through digital technology and analog sources. Big data is used for better understanding of business environment and markets, leading to better understanding of customers and at the same time increasing the performance of organizations [17].

Big Data constitutes data files whose size exceeds the possibilities of most current hardware and software technologies, so they can be paired and processed within a reasonable time. In addition, these data have a different texture, are heterogeneous, or they may be completely unstructured (such as multimedia or text documents) [18].

Big Data is a term used to describe the storage and analysis of large and complex data sets, by using a variety of technologies (MapReduce, NoSQL, etc.) [19].

With the advent of new technological options, such as social sites, mobile devices, cameras, and a variety of interfaces with sensors to data networks, many companies and institutions have become capable of generating huge volumes of data that potentially contain interesting information, but they are not directly evaluated by man or by the usual calculation means, only to a limited extent. These kinds of data are referred to as big data [20].

Big Data designates such amounts of data that are difficult to handle by using traditional, commonly available means or financially demanding [21].

Big Data represents the next generation of data warehousing and business analyses. The concept is designed to provide maximum support for the needs of the investigation costs and increasing the efficiency of businesses [22].

Big Data is data that require an inordinate amount of time/space for storage, transmission, processing, and use of available resources [23].

Based on the similarities of the aforementioned definitions, it is possible to establish Big Data as large amounts of data that are constantly generated from a variety of sources (social networks, sensors, Internet, text files, etc.), and whose processing by traditional technologies is financially and temporally unfeasible for enterprises. There are also new technologies focused on the collection, storage and analysis of data through a variety of advanced analytical tools. This work with the data (diverse data that is structured, flat and semi-structured) in the context of Big Data technology is characterized by high speed. The ability to process large amounts of disparate data has a major impact on the acquisition of relevant information, the detection of business opportunities and streamline activities in the framework of all the processes in the enterprise.

3. Characteristic of Big Data technology

Big Data is a new generation of technologies and structures for processing of large amounts of disparate data. This technology allows rapid collection, sorting, analyzing and extracting valuable data from large volumes of data.

According to IBM, Big Data brings platform with options how to handle large amounts of data in relation to business opportunities of an enterprise. Big Data platform includes technologies for processing of structured and unstructured data with traditional links to new technologies focused on speed, flexibility, and task force exploring, discovering, analyzing the data. According to IBM Big Data is defined by the following features [24] - Fig. 1:

- **volume** represents a very large number of collected data for analytical processing (e.g. Airbus aircraft generates up to 40 TB of data every half an hour, Twitter generates 12 TB of data a day, Facebook 25 TB, etc.) [25 and 26], which presents an opportunity for businesses to process voluminous data in a single database summary structure.
- **diversity** means that data are well structured in a structured form, i.e. the data entered into the form of messages, images, and other types of data generated by the GPS signals via the Internet and telecommunications equipment.
- **speed** indicates that data must be gathered and processed very quickly, i.e. in real or near real time, which enables businesses to respond flexibly to changes in the market, or gain a competitive advantage.
- **truthfulness** represents the possibility of obtaining distorted output as processing a large number of diverse data is accompanied by certain amount of data that contain noise or distortion (e.g. data from social networks).

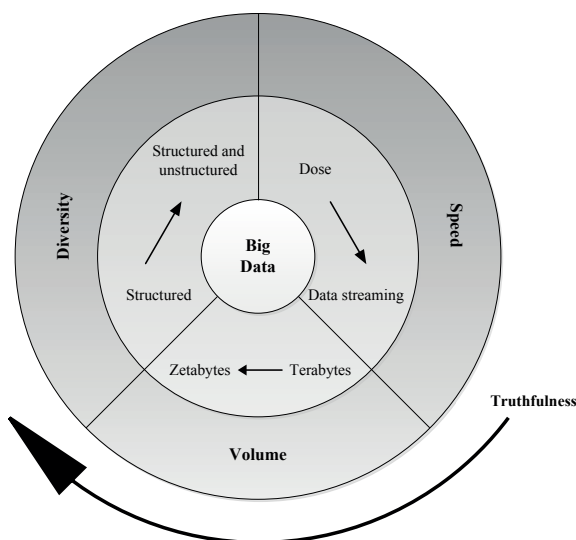


Fig. 1 Characteristics of Big Data, according to the IBM
Source: Edited according to ZIKOPOULOS, 2011, p. 5

The original characteristic of Big Data technology is defined by the three above-mentioned factors: volume, speed of data processing and diversity. The fourth factor, truthfulness, represents the extension of the three preceding characteristics of Big Data, as defined by IBM [26].

4. Benefits of Big Data technology

Big Data solutions are able to analyze almost all data sources that contain structured or unstructured data. Precisely for this reason, it is very rewarding, especially for large technologies and medium-sized enterprises that generate bulks of disparate data. The technology of Big Data can be utilized by small businesses to create positive influence on decision making and processes in enterprises, to ensure increase in sales and market growth. An example of this may be small businesses that operate call centers. Big Data platform is convenient exactly for this kind of enterprises as it generates a number of unstructured data. Assuming that a company knew how to analyze the quality of large volumes of unstructured data, in particular, it would receive more detailed information on the basis of which would be possible for example [6]:

- **Streamlining marketing enterprise**, for example, in the context of the call centers which carry out Direct Marketing. Success rate in mass marketing in this case is currently between 0.2% and 3.1% [3], which is inefficient from a marketing perspective. By analysis of unstructured data, using Big Data technology in the form of e.g. Event Driven Marketing, it is possible to increase the rate to a range from 18% to 34% [6], which represents a significant difference compared to the present.
- **Reducing the risk of loss**, by analyzing the behavior of clients using the Big Data technology. Analysis of the characteristics of the management of individual clients, the system can warn you about potentially risky procedures and thus, ultimately, to save money and time to prepare corrective measures against possible threats.
- **Adjust bids according to the needs of customers**, using information obtained in the analysis of unstructured data. On the basis of this information it is possible to predict purchase behavior of consumers and subsequently adapt the products according to their needs, requirements and expectations or situations. The information provided can be used not only by Internet sales and marketing campaigns, but also in the marketing of a brick and mortar store.
- **Streamlining work of the call center**, based on the processing of recorded interviews. Information that can be obtained from this data, they can improve the work of the operators, who then can offer the customer service that you need, or to solve his problem in a short time.

- *Streamlining maintenance processes*, for example, in the industrial sphere, where businesses use numerous sensors that generate large volumes of data. By analyzing these data, it is possible to prevent accidents, or more accurately plan their service and maintenance, allowing the company to reduce costs and downtime caused by malfunctioning equipment.
- *Understanding the spoken word*, in connection with the recognition of the voice, its analysis and sorting. This will allow businesses to save time and increase efficiency.
- *Improving the quality of provided service*, for example, in healthcare, where doctors decide on the treatment of their patients based on their experience. However, in the future the analysis of large scale of patients' data might enable doctors to make decisions on the basis of information and knowledge from a range of other doctors. Even in the field of business management, managers make decisions in various situations, not only on the basis of the information and outputs of the system, but also on the basis of their expertise.

Big Data technology can be applied in almost all areas of business where it is necessary to carry out decisions on the basis of readily available and relevant information obtained by analyzing large volumes of data.

5. Conclusion

The main problem in management and decision-making of enterprises is constantly growing amount of data generated within undertaking and its surroundings. These data reach the volumes and structures, which is not possible to manage through current management information systems due to time and cost.

The fastest increasing volumes of data are unstructured data, which may contain data with significant information value, for the purposes of decision-making in enterprises. Following the principle of processing data in the existing MIS, i.e. processing of structured data, companies do not have efficient tools to capture, analyze and transform the data.

Big Data brings solution to the problem of processing and integrating data of different types. This technology allows handling different kinds of data, from a variety of data sources, in a very short amount of time (in milliseconds).

Big Data technology provides business managers with the opportunity to carry out decision-making process on the basis of a larger quantity of available information. In general, the more information the better decision may be taken. Based on the above facts, it is thus possible to answer the question: *how much data do we need to achieve the best decision?* The answer to this question is: *it is never enough.*

References

- [1] *What we Need Big Data* (in Slovak) [on-line] [cit. 2016-02-25]. Available at: https://www.datalan.sk/tmp/asset_cache/link/0000014602/DATALAN%20NEWS-01-2014_web.pdf
- [2] BEZWEEK, S., EGBU, C.: *Impact of Information Technology in Facilitating Communication and Collaboration in Libyan Public Sector Organisations*, [on-line], [cit. 2016-02-02], available at: <http://usir.salford.ac.uk/12835/1/530.pdf>, 2010.
- [3] *What Is 'Big Data,' Anyway?*, [on-line], [cit. 2016-02-12], available at: <http://www.strategy-business.com/blog/What-Is-Big-Data-Anyway?gko=28596>
- [4] CERNY, M.: *Eight Technology Trends that Transform Libraries in the Information Society* (in Czech), [on-line] [cit. 2016-02-12]. Available at: http://itlib.cvtsir.sk/buxus/docs/30_osm%20technol%20trendu.pdf, 2013.
- [5] KIM, J.: *Big Data Sharing Among Academics*. In: HU, W., KAABOUCH, N. 2014. *Big Data Management, Technologies, and Applications*. USA: Information Science Reference, 2014, pp. 177-194, ISBN 9781466647008.
- [6] *Reasons Why Big Data Easier Life for Companies* (in Slovak), [on-line], [cit. 2016-02-22], available at: <http://www.zive.sk/clanok/66175/7-dovodov-preco-big-data-ulahcia-firmam-zivot>
- [7] *Oxford English Dictionary - Big Data*, [on-line], [cit. 2015-03-01], available at: <http://www.oed.com/view/Entry/18833?redirectedFrom=Big+Data#eid301162177>.
- [8] *Big Data: The next Frontier for Innovation, Competition, and Productivity*, [on-line], [cit. 2016-02-15], available at: <http://www.mckinsey.com/business-functions/business-technology/our-insig-hts/big-data-the-next-frontier-for-innovation>.
- [9] *Definiton of: Big Data*, [on-line], [cit. 2016-02-20], available at: <http://www.pcmag.com/encyclopedia/term/62849/big-data>.
- [10] *Big Data. What is it & why it matters*, [on-line], [cit. 2016-02-10], available at: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html.
- [11] STANIMIROVIC, Z., MISKOVIC, S.: *Efficient Metaheuristic Approaches for Exploration of Online Social Networks*. In: HU, W., KAABOUCH, N.: *Big Data Management, Technologies, and Applications*. USA: Information Science Reference, 2014, pp. 222-269, ISBN 9781466647008.
- [12] *What is Big Data*, [on-line], [cit. 2016-02-25], available at: <http://www.ibm.com/big-data/us/en/>.

- [13] BAKSHI, K.: *Technologies for Big Data*, In: HU, W., KAABOUCHE, N. 2014. *Big Data Management, Technologies, and Applications*. USA : Information Science Reference, 2014, pp. 1-22, ISBN 9781466647008.
- [14] *Explore the World of Big Data*, [on-line], [cit. 2016-02-01], available at: <http://go.sap.com/solution/big-data.html>.
- [15] MADDEN, S.: *From Databases to Big Data*, [on-line], [cit. 2016-02-22], available at: <https://www.computer.org/csdl/mags/ic/2012/03/mic2012030004.pdf>, 2012.
- [16] MATTMANN, CH. A. et al.: *Scalable Data Mining, Archiving, and Big Data Management for the Next Generation Astronomical Telescopes*. In: HU, W., KAABOUCHE, N.: *Big Data Management, Technologies, and Applications*, Information Science Reference, 2014, pp. 196-221, ISBN 9781466647008.
- [17] *Best Practice Guideline: Big Data*, [on-line], [cit. 2016-02-05], available at: <http://www.adma.com.au/assets/Uploads/Downloads/Big-Data-Best-Practice-Guidelines2.pdf>.
- [18] SEVVAS, K. I. et al.: *Applying the K-Means Algorithm in Big Raw Data Sets with Hadoop and MapReduce Data*. In: HU, W., KAABOUCHE, N.: *Big Data Management, Technologies, and Applications*. Information Science Reference, pp. 23-46, 2014, ISBN 9781466647008.
- [19] *The Big Data Conundrum: How to Define It?* [on-line], [cit. 2016-02-05], available at: <http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/>.
- [20] STRASIK, S.: FlexPod Select Hardware Platform for Hadoop (in Slovak), *INFOWARE X.*, vol. 1-2, 2014, pp. 20-21.
- [21] SYNEK, T.: How to Save IT Costs with Big Data Technology (in Slovak), *INFOWARE X.*, vol. 1-2, 2014, pp. 23-24.
- [22] MINELLI, M., CHAMBERS, M., DHIRAJ, A.: *Big Data, Big Analytics. Emerging Business Intelligence and Analytic Trends for Today's Businesses*. John Wiley & Sons, 2013, 187 p., ISBN 9781118147603.
- [23] YILDIRIM, A., A. et al.: *Parallel Data Reduction Techniques for Big Datasets*. In: HU, W., KAABOUCHE, N.: *Big Data Management, Technologies, and Applications*, Information Science Reference, 2014, pp. 72-93, ISBN 9781466647008.
- [24] *Big Data for Beginners and Advanced* (in Czech), [on-line], [cit. 2016-02-10], available at: <http://data.computerworld.cz/file/BigData/BigData-2012-web.pdf>.
- [25] *How Can we Use Big Data* (in Slovak), [on-line], [cit. 2016-02-01], available at: <http://www.itnews.sk/2012-05-02/c148464-ako-mozno-vyuzit-big-data>.
- [26] *Data and Big Data - Prepare to Handle the New Paradigm?* (in Slovak), [online], [cit. 2016-02-20], available at: http://issuu.com/jozefsupsak/docs/zbornik_marec2013_final/3?e=2664486/8836631.