BIG DATA AND DECISION-MAKING SUPPORT

Peter Procházka

Abstract:
INTRODUCTION: Nowadays, Big Data is created in previously unimaginable quantities. Newly generated data from various Internet of Things (IoT) sensors and their use have never reached their current dimensions. Along with this trend, the availability of devices capable of collecting this data increases, the time for their evaluation is reduced and the volume of data collected at the same time increases. The most important task of research and development in this area is to bring solutions suitable for processing large amounts of data because our current storage and processing capabilities are limited and unable to compete with the storage, processing and publication of the resulting data.

OBJECTIVES: Point out the importance of implementing Big Data technology.

METHODS: To achieve the goal, the following methodological approach was chosen: study and processing of foreign and domestic literature, acquaintance with similar solutions for data processing, definition of Big Data and IoT, proposal for using Big Data solution to support decision-making, risk definition and evaluation.

RESULTS: With the growing amount of disparate and incoherent data and the further growth of the Internet of Things, it is now almost impossible to evaluate all available information correctly and in a timely manner. Without this knowledge, the company loses its competitive advantage and is unable to respond in a timely manner to client requests.

CONCLUSION: Implementing a solution for processing Big Data to support decision-making in the company is a complex process. As part of the implementation and use of the Big Data solution to support decision-making, the company must be prepared for the emergence of various problems. We can assume that Big Data technology will constantly be evolving in terms of streamlining analytical tools for obtaining information from large volumes of generated data. Therefore, it is appropriate to create space for the implementation of Big Data technology.

UDC Classification: 004.6, DOI: https://doi.org/10.12955/pns.v2.148

Keywords: Big, Data, IoT.

Introduction
Developments in the field of information and communication technologies are aimed at expanding technical products and services, which will be increasingly available to both businesses and ordinary users. This means that the amount of generated data will also increase, which may contain potentially significant information value for managers' decisions. Therefore, it can be assumed that solutions for processing disparate data will serve as a basic information system or an analytical platform for the needs of capturing, storing, organizing and processing data for real-time decision-making. In this context, new concepts such as Big Data and the Internet of Things (IoT) have emerged. In recent years, there has been significant development and generation of business opportunities in these areas.

Internet of things (IoT)
An important source of data today, thanks to technological advances, is the Internet of Things, where devices are connected that generate data that can be captured, stored and processed. Within the given technological solution, communication or interconnection is created on three levels:

- machine - machine,
- machine - human,
- man - man.

This means that data is generated from communication between connected devices, between users and devices, and between multiple users. As the IoT network envisages the global interconnection of devices via the Internet, it is possible to create a variety of data that can be distributed to the company via the IoT network, thus managing internal processes at a global level. Through the IoT network, there is a rapid exchange of information in real-time, which has a positive effect on the speed of response in the event of emergencies in the company's environment. At the same time, in this way, IT trends are adapting, which allows companies to have a constant overview or the opportunity to work with modern Information and Communications Technology (ICT) to support decision-making.

Ashton (1999) first used the term "Internet of Things" in an article published in the RFID Journal. The IoT allows various devices (smart home, car, life sensor, security system, material handling devices, etc.) to be reached via the Internet or another network, or in some cases, allows communication between

1 University of Economics in Bratislava, Faculty of Economic Informatics, Department of Applied Informatics, Bratislava, Slovakia, peter.prochazka@euba.sk
devices. In recent decades, most of the data available on the Internet has been recorded by humans, which has severely limited the types and amount of data available. In order for humanity to continue to streamline the systems it operates, it was necessary to collect data on the things and devices used and on the basis of the data thus obtained to enable intervention in the functioning of the system. For example, if we equip devices with sensors that detect a defect or damage before they fail, then we can intervene, make repairs and avoid production downtime.

As it is 2021 and the COVID 19 pandemic is spreading around the world, successes or failures in the evaluation of data related to this area, or seemingly unrelated but having an impact on solutions, strongly affect the health and lives of people and economies across countries. Groups of crisis managers receive daily reports from hospitals, local governments, mobile collection points about the number of positively or negatively tested people, the sick or the deceased, who are tied to some territorial units and take measures based on them. However, in a successful fight against a pandemic, it is also possible to use IoT data of devices that have become an integral part of our lives, whether they are the already well-known smart quarantines (eQarantines) or forecasting the behavior of the population based on weather monitored by weather station sensors; mobile devices or GPS devices of vehicles on the road, where traffic jams are formed for early warning of rescuers to choose other routes; or common temperature sensors connected to a tablet, which monitor the temperature of customers to better estimate how many potentially ill people are still moving in public. A large amount of data from such devices also reaches social networks in various forms, whether they are text messages, images or videos. These can point in the same serious way to the mistakes that are being made in the fight against a pandemic if we can process and evaluate them in a timely and correct manner. It is therefore clear from the outline that if we obtained the relevant data and were able to evaluate it correctly, we would have a very powerful tool in our hands. And the number of devices that record various data or events in our world is increasing day by day. This means that the amount of generated data will also increase, which may contain potentially significant information value for managers' decisions.

We can also consider, for example, the case of smart cities, where the requirement to evaluate large volumes of data in real-time, to achieve the required level of operation and security, in the conditions of management of the city or its parts, comes to the fore. Another important area for the use of the Big Data solution in the future, in terms of decision support, is the so-called Industry 4.0. In this case, Big Data solutions are important in terms of the rapid evaluation of the amount of diverse data generated within a comprehensive production management system in order to streamline production processes and improve the production according to the customer needs and requirements.

Based on the above, we can assume that the use of IoT will grow side-by-side with the ever-expanding portfolio of various devices connected to the Internet, while the main product of these devices will be data. The amount of disparate data will increase with the number of these devices. The information obtained from the data will bring various benefits to the whole society.

**Big Data**

The development of common devices that can communicate with each other via the Internet, ubiquitous sensors, affordable large data storages and many other circumstances have in recent years generated and stored huge amounts of disparate data, coming at a high pace. Traditional information systems are not able, in terms of time and money, to distribute relevant information from said data to managers in the decision-making process. This conditioned the emergence of new technologies aimed at the efficient and fast processing of this large volume of disparate and incoherent data, i.e., Big Data. The processing of such data is a financially and technologically demanding process, which if any company wants to implement, it is important to first understand the issue, identify the meaning of the solution and develop an implementation plan. In case of incorrectly defined requirements and expectations, an inappropriate solution may be implemented, which may prove to be unnecessarily financially and technologically demanding for decision support, and the company will not be able to use it.

The origin of the Big Data concept, its development and the term Big Data used today since its first discovery at the end of the last century have undergone a significant change in content. Since then, the term has become very widespread, but its full meaning is still little known. It, therefore, needs to be comprehensively examined and described.
Denning (1990) found that the volume of information exceeded the available amount of network capacity, storage capacity, or the ability of people to understand such a volume of data. He recommended the development of pattern recognition devices that can filter or predict important “bits”. Mashey (1998) gave a lecture entitled Big Data ... and the Next Wave of InfraStress. He highlighted the general needs of the ever-increasing amount and diversity of data and presented the expected needs of user-side data storage and the increasing pressure on networks.

The first definitions of Big Data naturally did not have an academic origin. They had much more technical interpretations, which were published in various lectures and blogs. This trend has changed since 2013 when the Big Data phenomenon erupted in the academic space, early technological interpretations have already expanded, and the concept has entered a period of practical implementation. Creating a definition of big data is not an easy task given the technological background, which is constantly changing. Likewise, the definition of Big Data based on data volume determination also does not hold because what was considered Big Data in terms of data volume 10 years ago may no longer represent Big Data today, just as today's data volumes will not represent another 10 years. For this reason, today, we know several different definitions of Big Data, of which we will list only a few to illustrate the matter.

Diebold (2000) - Big Data refers to the explosion in quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology.

Jacobs (2009) - Big Data should be defined at any point in time as “data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time.”

Wilder-James (2012) - Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn’t fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it.

Gartner (2014) - Big data is high -volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

Grady and Chang (2017) - The Big Data characteristics of volume, velocity, variety, and variability demand a versatile management platform for storing, processing, and managing complex data.

In several sources, we meet with opinions on the social benefits of Big Data innovations. We see opinions and reports on their impact on economic, legal, social, ethical, and political issues. The societal usefulness of innovations thanks to Big Data technology is defined by identifying target areas and examining them in detail. These target areas are, in general, public health, public and national security, governance, the environment and disaster management, education, economic development, etc. We consider the original use of Big Data technology to be the area of economic analysis (predictive analyzes of the banking sector, economic habits based on the analysis of customer habits, business planning, etc.). It is possible to observe that there are few areas of activity where Big Data does not bring changes.

At present, we define Big Data by three, or in some publications by four characteristics. In 2001, the META Group analyst (now Gartner) Doug Laney introduced the well-known 3V characteristics - volume (amount of data), velocity (speed of data in / out), and variety (range of data types, sources) in his publication on three-dimensional data processing. It is worth noting that the name Big Data does not appear here, however, we consider it one of the first sources published in the field of Big Data. Several other sources tried to expand this set of characteristics, but in the end, only veracity remained, creating the 4V designation attributed to IBM.

*Volume*

Volume represents the amount of raw data generated. Exabytes are typical. The extent of the data is not only a storage problem (e.g., what data should be kept and for how long, the periodic deletion of rarely used data, etc.) but also causes significant problems with their processing. The boundary between large and “normal” volumes of data is unlikely to be accurately defined, expressed, or quantified, as technological performance is constantly increasing due to technological progress and capacity. In this case, we can rather talk about an imaginary interval, where both of its borders are continuously shifting to higher values in the long run, while currently, various sources estimate the total volume of data at several tens of zettabytes.
Velocity
The speed at which data is generated and changed, and the speed at which data needs to be received, interpreted, and processed, should be close to real-time work, or at least within a reasonable time. By reasonable time we mean such a maximum response time, during which the processing of freshly captured data still makes sense and fulfills the desired purpose. The "time value" of the data is therefore also important. Real-time processing of up-to-date data and its timely analysis enables better decision-making, such as product recommendation, safety risk analysis, etc.

Variety
The diversity of Big Data lies in the different types and forms of data captured and processed, as well as in their origin. Large volumes of data are obtained from text, images, audio and video recordings, logs in text files or other unstructured sources. With the help of data fusion, it is also possible to supplement the missing parts of the whole. Sources of large volumes of data are mainly the activities of people in the online space, sensors on various devices, data from financial transactions, etc.

Veracity
By the credibility of Big Data, we mean the evaluation of data quality and the value it carries. The quality and accuracy of the data captured can vary significantly, which can have a significant impact on the accuracy of the analyses. Therefore, it is not enough just to collect data, it is also necessary to maintain its quality and accuracy at a sufficiently high level and verify it regularly. Sources of problems that need to be addressed are, for example, conflicting or ambiguous data, duplicates, errors caused by delayed processing, false data, spam, etc.

Figure 1: Big Data characteristics by IBM

Source: Edit according to Zikopoulos (2011)

Decision support
Every company is not only people, machines, devices but also various processes that are performed in interaction with external and internal environments. To ensure the achievement of the necessary results or outputs from these processes, it is important to manage them effectively. The most important part of the decision-making process is the identification of the problem and the setting of goals because, in case of inappropriate specification of the problem, the final decision or solution will be inefficient. This activity is performed by managers. Management was defined by Hitmár (2006) as a process in which managers, with the help of scientific knowledge, but especially practical recommendations, know how to perform basic managerial activities in evaluating available resources to determine and achieve goals.

Among the various data generated in the company, we can include:

- structured data, i.e., data generated by the company’s systems, in a structured form,
- multimedia content (video, pictures, spoken word),
- documents, e-mails, and text format records,
• sensor data,
• machine-generated data,
• data from various applications and utility software, which show an unstructured character,
• data from the IoT network, etc.

When implementing the Big Data decision support platform in the company, we must precisely specify the needs and requirements for the system in connection with its use in the decision-making process in order to prevent unnecessary damage by incorrectly defining requirements and in order to obtain the most appropriate solution. The basic processes implemented within the application of the solution include:

• setting goals and resources to achieve them,
• setting requirements for the integrated Big Data platform,
• analysis of suppliers,
• sending requests for assessment by suppliers,
• assessment of information from suppliers,
• implementation of the audit,
• assessment of audit results,
• setting new requirements for the integrated Big Data platform.

In connection with the reasons for the implementation and use of the Big Data solution in companies, we can identify the benefits of using the Big Data solution in the decision-making process as follows:

• the possibility of implementing more precise decisions of company managers,
• integration with corporate information systems,
• uniform database structure of diverse data,
• the ability to process and visualize a variety of data,
• data analysis and evaluation in real-time,
• implementation of predictive analyses on all data,
• more accurate identification of customer needs and requirements,
• expansion of the product portfolio,
• improving product quality,
• increase in sales,
• cost reduction,
• elimination of waste,
• support of processes within the whole company,
• acceleration of communication,
• access to data via mobile devices.

Conclusion

Data enters the decision-making process as a resource that carries a certain amount of information that may relate to the decision-making process. The decision-making of managers in the company is therefore based primarily on the maximum use of information that is available in a correct and clear structure, in the current time and with immediate access using various devices.

Implementing a Big Data decision support solution to assist decision-making in a company is a complex process. This complexity is also increased by the intricacy of the Big Data platform itself in terms of hardware, software, personnel security, processes, and information flows. As part of the implementation and use of the Big Data solution for decision support, the company must be prepared for the emergence of possible problems, both in the design of a decision model with Big Data support, as well as in the implementation of the proposed solution in the company. Some problems can occur in duplicate in both models, which means that the company can design measures to eliminate them or minimize the impact when designing a decision model with Big Data support.

We can assume that Big Data technology will constantly be evolving in terms of streamlining analytical tools for obtaining information from large volumes of generated data. Therefore, it is appropriate to create space for the implementation of Big Data technology within the decision-making process of managers and executives of a company, who need relevant information to be available at all times when making decisions.
References

Ashton, K. (2009, June 22). That 'Internet of Things' Thing, in the real world things matter more than ideas. RFID Journal. [Web blog post]. Retrieved from https://www.rfidjournal.com/that-internet-of-things-thing

Denning, J.P. (1990). Saving All the Bits, The Science of Computing, Research Institute for Advanced Computer Science at the NASA Ames Research Center, 1-4.

Diebold, X.F. (2000). ‘Big Data’ Dynamic Factor Models for Macroeconomic Measurement and Forecasting, in M. Dewatripont, L.P. Hansen and S.Turnovsky (Eds.), Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society. Cambridge, 115–122.

Eaton, C., Deroos, D., Deutsch, T., Lapis, G., Zikopoulos, P. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw Hill Professional, ISBN: 978-0-07-179053-6

Gartner IT Glossary (2014). What is Big Data?

Grady, N., Chang, W. (2017). NIST Big Data Interoperability Framework, NIST Big Data Public Working Group (NBD-PWG). Vol. 1-9 Version 2., Gaithersburg, MD 20899 USA: U. S. Department of Commerce, 1-20.

Hittmár, Š. (2006). Manažment. Žilina : EDIS, ISBN 80-8070-558-5

Jacobs, A. (2009). The Pathologies of Big Data, Commun. ACM, 36–44.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Application Delivery Strategies, Stamford, Connecticut, U.S.A.: META Group Inc.

Mashey, J.R. (1998). Big Data ... and the Next Wave of InfraStress. USENIX meeting. Retrieved from https://static.usenix.org/event/usenix99/invited_talks/mashey.pdf

Wilder, J.E. (2012, January 11). What is big data? [Web blog post]. Retrieved from https://www.oreilly.com/ideas/what-is-big-data