



Journal
of Danubian
Studies
and Research

Entrepreneurial Perspectives and their Impact on the Danube

Big Data: Actuality and Challenges

Adina Barila¹, Mirela Danubianu², Corneliu-Octavian Turcu³

Abstract The volume of data is constantly growing due to the explosion of machine-generated data and human involvement in social networks, especially in the last period in which the pandemic forced most activities to take place online. Big Data refers to storage, manipulation and analysis of this huge data sets that come from variety of sources and are too large and too heterogeneous to be traditionally processed. This paper gives an overview of Big Data sources, Big Data analytics, its applications, advantages and limitations, and challenges that Big Data has to face nowadays.

Keywords: big data; data analytics; data storage; data quality; privacy

1. Introduction

Data is all around us. A while ago they were generated by employees. Nowadays almost every action, every word, every click creates data. There are more and more sensors that collect data. More and more devices are generating and transmitting more and more data these volume of data have to be gathered, stored, and explored. Big Data offers solutions. Initially described by the 3Vs, standing for volume, velocity and variety, now other Vs come to describe Big Data. Literature indicates 7 or even 10 Vs. The huge amounts of data must be explored and analyzed to provide meaningful information. This is the goal of Big Data Analytics. Constantly generating of data in a high rate and by different sources, the need of moving forward from

¹ Ștefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Corresponding author: adina.barila@usm.ro

² Ștefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, E-mail: mirela.danubianu@usm.ro

³ Ștefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, E-mail: cturcu@usm.ro.

traditional storage and analytics system to new storage and analytics system, have put the Big Data in situation of facing a number of challenges.

This paper aims to present an overview of Big Data. Section 2 defines the characteristics of Big Data. Section 3 introduces Big Data Analytics and presents the type of analytics. Section 4 presents some of the challenges Big Data has to face nowadays.

2. Big Data Characteristics

Big Data is defined as “datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze” (Manyika, et al., 2011). These datasets consist of structured data, unstructured data or semi-structured data. They may contain enterprise data, social data, sensor collected data or machine-generated data.

Although the definitions of Big Data in the literature consider the size aspect, this is not the only feature of Big Data. Now it is known about Big Data that “the data is too big, moves too fast, or doesn’t fit the structures of your database architectures” (Emami, Culot & Nicolle, 2015). Big Data has some characteristics known as *the three V’s*: volume, velocity and variety. Some authors added veracity and value and, recently, new V’s have come to characterize Big Data: variability and visualisation. So we can speak about *the seven V’s* of Big Data.

The first V – *volume* – indicates the size attribute of data. Big Data refers to large amounts of data. Nowadays huge quantities of data are stored: transaction-based data, reports, text data and videos constantly streaming in from social networks, clinical data, administrative data surrounding payments and payers, increasing amounts of sensor data. Volume of data is doubling every 12-18 months (Maheshwari, 2017). On the other hand, the cost of storing data and the cost of communication of data are coming down every year. There are organizations storing gigabytes, terabytes, petabytes or exabytes of data. The term volume is a relative one, because there is not a limit beyond which the data become big data.

The second V – *velocity* – refers to the speed at which new data is being created and accessed or delivered. It’s about not only the speed of creating new data but the speed of analyzing and giving feedback which lead to make a decision. The velocity of data is determined by the continuous increase of Internet speeds, and variety of devices. Gathering, processing and presenting data as close as possible in real-time

can provide companies with insights that will lead to better business results **Error! Reference source not found.**

The third V – *variety* – is related to the high diversity of data types. This is generated by combining data from different sources having different formats and different functions. Can be traditional relational databases and also can be text documents, emails, posts on social media, images, videos, financial transactions, data collected from sensors. Volume of unstructured data is larger than that of structured data. It's a typical use of data processing to extract ordered meaning from unstructured data for immediate use or as a structured input to an application (Syed, Gillela, & Venugopal, 2013).



Figure 1. The 7 Vs. of Big Data

The fourth V – *veracity* – is related to the quality of data. It's about accuracy, certainty, precision. In traditional databases and data warehouses there was always the assumption that the data is certain, clean, and precise, but Big Data has to deal with uncertain or imprecise data Emani, Cullot, & Nicolle, 2015.

The fifth V – *value* – refers to the usefulness of data, to the information which can be obtained by analysis of data. Regardless of its volume (or because of its volume) data isn't very useful. Analysis of data can offer useful information for a better decision-making process.

The sixth V – *variability* – refers to continuously changing of data. This means that the data offer different meaning at different time. Also the reason of this V is the fact that big data velocity is not consistent and has periodic peaks and troughs (Nimankar & Dagare, 2018).

The seventh V – *visualisation* – refers to the process of displaying data in graphical formats, such as charts, graphs, maps. This makes understanding and interpreting data faster and easier.

3. Big Data Analytics

The growing data sets are only useful if they can be analyzed. The basic challenge of Big Data is to explore large volumes of data in order to extract useful information and competitive knowledge which serves, ultimately to decision making (Danubianu & Barila, 2014). Figure 2 presents the process of transforming raw data into decision.

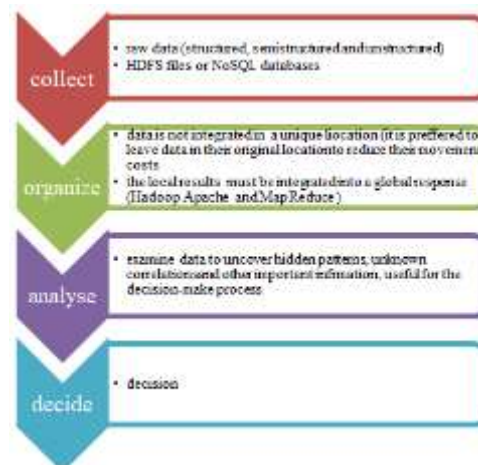


Figure 2. Process of Transforming Raw Data into Decision (Danubianu & Barila, 2014)

Big Data Analytics offers tools and methods to accumulate, manage, analyze, combine and assimilate large volumes of disparate, structured and unstructured data.

3.1. Types of Big Data Analytics

There are four types of Big Data analytics which use different technologies and architectures.

Descriptive Analytics combines past data from multiple sources into a readable form. This type of analytics offers insights into what has happened in the past without establishing the cause of a certain event or phenomenon. Can be especially helpful in tracking trends to help plan for the future. The results are shown in a form that can be easily interpreted by people. A common example of descriptive analytics are reports about revenues, sales and profits of a company (Shabana & Sharma, 2021).

Diagnostic Analytics analyzes the past data to understand what the cause a problem or what was the conditions in which a certain event has happened. A common example is the analyse of sales report of a company. If the sales decrease although customers are adding products in their shopping carts, the reason can be found by analytics. It can be the shipping fee, the low number of payment methods or unproperly load of form.

Predictive Analytics looks into the past and present data and, like then name shows, make prediction regarding the future. They estimate the evolution of a certain event or phenomenon. Typical use of this type of analytics is predicting market trends or customer trends. It can be used to predict fraudulent activities based on analyzing customer behavior. It is notable that all predictive analytics are probabilistic. They only forecast what might happen in the future, they do not tell what will happen in the future (Pitu & Gulia, 2019).

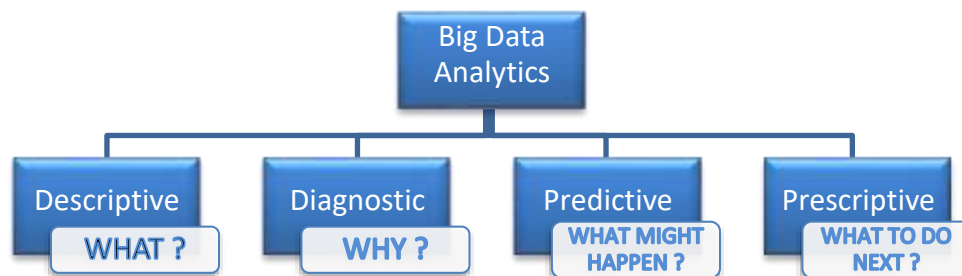


Figure 3 The Four Types of Big Data Analytics

Prescriptive Analytics can suggest solution to a specific problem, depending on the results of descriptive and predictive analytics, prescriptive analytics. The results of this type of analytics are rules and recommended actions. Prescriptive Analytics can be used to maximize the profit of a company by building algorithms that will automatically adjust the offers according to the clients' needs **Error! Reference source not found.**(Shabana & Sharma, 2021).

4. Big Data Challenges

Big data has some challenges, some of them are mentioned below:

- **Data storage** – In spite continuously growing of storage capacities and continuously decreasing of storage costs, Big Data push the limit on storage capacity. Storage systems of organizations and enterprises are facing major challenges from huge quantities of data, and the ever increasing of generated data (Agrawal & Nyamful, 2016). This systems have to ensure not only the storage but also quick acces to data. In addition, data backup has an important role in IT environment so huge amounts of data need backup and archive. Outsourcing data to the cloud seems to be a solution for this big challenge (Banu & Yakub, 2020).

- **Data management** - Nowadays the digital data are generated by companies, by individuals and by machines. These are in differents forms, such as text documents, spreadsheets, image files, audio files, video files. Due to volume, variety and velocity of Big Data, the data management systems have to adapt to ensure a properly administer of all data after storage and a properly and quick retrieval of them (Doshi, Agrawal, Kanani, & Padole, 2020).

- **Data quality** – Due especially to big variety of data sources, Big Data to quality of raw data. The poor quality of data can be caused by human error, technical error or malicious intent. For example, The data from official governement or companies websites are trustworthy. But we could not say the same about the data from social networks. Some data provided can be intentionally incorrect. Also the sensors for colleting data or machines that transmite data may malfunction and may record or communicate wrong data (Maheshwari, 2017).

- **Privacy** – This is a big challenge of Big Data. Due to the fact that Big Data combine data from a variety of sources, confidential data it might be compromised. In healthcare domain, for example, sensitive data is accumulating and the management of them must ensure that they remaine confidential. The social media data is another example for privacy issues. Data about anyone could become open to the world. Confidentiality of personal data, especially the user location, is a big issue in Big Data (Jony, Rony, Rahman, & Rahat, 2016). In addition, the analytics methods appllied to personal information stored in Big Data might lead to the found out of new information of that person, so to taking insights in people's lives without their consent.

- **Security:** The sensitive information about peoples and companies collected in Big data can be a target for hackers. There are financial information, trade secrets, intellectual property, personal health information. The security mechanisms must to deal with huge volume of data which are being exposed to much more digital attacks. Security breach of organization can directly effect security and privacy of data in big data Analytics (Jamil, Abdullah, Javed, & Hassan, 2018).
- **Lack of skilled people** – The big data is a relative new domain and uses new technologies. Their advantages have determined companies to rush to use Big Data analytics. But there is a lack of skilled people for it. According to a QuantHub survey there was a shortage of 250,000 data science professionals in 2020. In 2020 the demand for data analysis skills has so far exceed suply (DuBois, 2020).
- **Value of data** - The goal of gathering, storing, streaming the huge amounts of data is to explore or analize and turn them into value .

5. Conclusions

Today companies are turning to Big Data tools and technologies for data analytics and decision making. The paper defined what is meant by Big Data and presented its seven important characteristics: volume, velocity, variety, veracity, value, variability, visualisation. This work also presented the four types of Big Data Analytics that help organizations improve their activities and increase profits. Finally we focused on Big Data challenges in the Big Data era.

6. Acknowledgement

“This work is supported by the project *ANTREPRENORDOC*, in the framework of Human Resources Development Operational Programme 2014-2020, financed from the European Social Fund under the contract number 36355/23.05.2019 HRD OP /380/6/13 – SMIS Code: 123847.”

References

- Agrawal, R.; Nyamful, C. (2016). Challenges of big data storage and management. *Global Journal of Information Technology*. 6(1), pp. 1-10.
- Banu, A. & Yakub, M. (2020). Evolution of Big Data and Tools for Big Data Analytics. *Journal of Interdisciplinary Cycle Research*, Volume XII, Issue X, 309-316.
- Danubianu, M. & Barila, A. (2014). Big Data vs. Data Mining for Social Media Analytics. *International Conference on Social Media in Academia - Research and Teaching –SMART2014*.
- Doshi, Z.; Agrawal, R.; Kanani, P. & Padole, M. (2020). *Big Data, Big Challenges*.
- DuBois, J. (2020, April). Retrieved July 2021. *QuantHub*. <https://quanthub.com/data-scientist-shortage-2020/>.
- Emani, C.; Cullot, N. & Nicolle, C. (2015). Understandable Big Data: A survey. *Computer Science Review*, Volume 17, pp. 70-81.
- Jamil, A.; Abdullah, M.; Javed, M. & Hassan, M. (2018). Comprehensive Review of Challenges & Technologies for Big Data Analytics. *2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET)*, pp. 229-233.
- Jony, R.; Rony, R.; Rahman, M. & Rahat, A. (2016). Big Data Characteristics, Value Chain and Challenges. *1st International Conference on Advanced Information and Communication Technology 2016*.
- Maheshwari, A. (2017). *Big Data*. McGraw Hill Education (India) Private Limited.
- Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C., et al. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*.
- Nimankar, S. & Dagare, S. (2018). 7 Dimensions of Big Data Analytics. *Global Journal of Engineering Science and Researches*, pp. 14-19.
- Ritu, R., Gulia, P. (2019). Big Data Tools and Techniques: A Roadmap for Predictive Analytics. *International Journal of Engineering and Advanced Technology (IJEAT)* Volume-9 Issue-2, 4986-4992. ISSN: 2249 – 8958. DOI: 10.35940/ijeat.B2360.129219
- Shabana, M. & Sharma, V. (2021). A Study on Big Data Advancement and Big Data Analytics. *Journal of Applied Science and Computations*, pp. 4099-4108.
- Syed, A.; Gillela, K. & Venugopal, C. (2013). The Future Revolution on Big Data. *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, Issue 6, 2446-2451. ISSN (Online), pp. 1021-2278.