



The Influence of Big Data Analytics in the Industry

Haya Smaya

Mechanical Engineering Faculty, Institute of Technology, MATE Hungarian University of Agriculture and Life Science, Gödöllő, Hungary

Email: hayasmaya1993@gmail.com

How to cite this paper: Smaya, H. (2022) The Influence of Big Data Analytics in the Industry. *Open Access Library Journal*, 9: e8383.

<https://doi.org/10.4236/oalib.1108383>

Received: January 20, 2022

Accepted: February 13, 2022

Published: February 16, 2022

Copyright © 2022 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Big data has appeared to be one of the most addressed topics recently, as every aspect of modern technological life continues to generate more and more data. This study is dedicated to defining big data, how to analyze it, the challenges, and how to distinguish between data and big data analyses. Therefore, a comprehensive literature review has been carried out to define and characterize Big-data and analyze processes. Several keywords, which are (big-data), (big-data analyzing), (data analyzing), were used in scientific research engines (Scopus), (Science direct), and (Web of Science) to acquire up-to-date data from the recent publications on that topic. This study shows the viability of Big-data analysis and how it functions in the fast-changeable world. In addition to that, it focuses on the aspects that describe and anticipate Big-data analysis behaviour. Besides that, it is important to mention that assessing the software used in analyzing would provide more reliable output than the theoretical overview provided by this essay.

Subject Areas

Information Management

Keywords

Big Data, Information Tools

1. Introduction

The research background is dedicated to defining big data, how to analyze it, the challenges, and how to distinguish between data and big data analyses. Therefore, a comprehensive literature review has been carried out to define and characterize Big-data and analyze processes. Several keywords, which are (big-data),

(big-data analyzing), (data analyzing), were used in scientific research engines (Scopus), (Science direct), and (Web of Science) to acquire up-to-date data from the recent publications on that topic.

The Problem this paper wants to solve is to show the viability of Big-data analysis and how it functions in the fast-changeable world. In addition to that, it focuses on the aspects that describe and anticipate Big-data analysis behaviour. Big Data is omnipresent, and there is an almost urgent need to collect and protect whatever data is generated. In recent years, big data has exploded in popularity, capturing the attention and investigations of researchers all over the world. Because data is such a valuable tool, making proper use of it may help people improve their projections, investigations, and decisions [1]. The growth of science has driven everyone to mine and consume large amounts of data for the company, consumer, bank account, medical, and other studies, which has resulted in privacy breaches or intrusions in many cases [2]. The promise of data-driven decision-making is now widely recognized, and there is growing enthusiasm for the concept of “Big Data,” as seen by the White House’s recent announcement of new financing programs across many agencies. While Big Data’s potential is real—Google, for example, is thought to have given 54 billion dollars. In 2009 to the US economy—there is no broad unanimity on this [3].

It is difficult to recall a topic that received so much hype as broadly and quickly as big data. While barely known a few years ago, big data is one of the most discussed topics in business today across industry sectors [4].

This study is dedicated to defining the Big-data concept, assessing its viability, and investigating the different methods of analyzing and studying it.

2. Status Quo Overview

This chapter will provide a holistic assessment of the Big-data concept based on several studies carried out in the last ten years, in addition to the behaviour of big-data, features, and methodologies to analyze it.

2.1. Big-Data Definition and Concept

Big data analytics is the often-complex process of analyzing large amounts of data to identify information such as hidden patterns, correlations, market trends, and customer preferences that can assist businesses in making better decisions [5]. Big Data is today’s biggest buzzword, and with the quantity of data generated every minute by consumers and organizations around the world, Big Data analytics holds a huge potential [6].

To illustrate the importance of Big-data in our world nowadays, (Spotify) could be a good example of how Big-data works. (Spotify) has nearly 96 million users that generate a vast amount of data every day. By analyzing this data and based on it, the cloud-based platform suggests songs automatically using a smart recommendation engine. This huge amount of data is the likes, shares, search

history, and every click on the application. Some researchers estimate that Facebook generates more than 500 terabytes of data every day, including photos, videos, and messages. Everything we do online in every industry uses mostly the same concept; therefore, big data get all this hype [7].

Generally, Big-data is a massive amount of data set that cannot be stored, processed, or analyzed using traditional tools [8]. This data could also exist in several forms, such as structured data and semi-structured data. The structured data might be an Excel sheet that has a definite format. At the same time, Semi-structured data could be resembled by an email, for example. Unstructured data are undetermined pictures and videos. Combining all these types of data creates what is so-called (Big-data) (Figure 1) [9] [10].

2.2. Characteristics of Big-Data

Firstly, it is essential to differentiate between Big-data and structured data (which is usually stored in relational database systems) based on five parameters (Figure 2) which are:

- 1-Volume
- 2-Variety
- 3-Velocity
- 4-Value
- 5-Veracity

And usually, it can be referred to these parameters as (5V's) that are the main challenges of Big-data management:

1-Volume:

Volume is the major challenge for Big-data and the paramount aspect that distinguishes it. Big-data volume is not measured by gigabytes but by terabytes

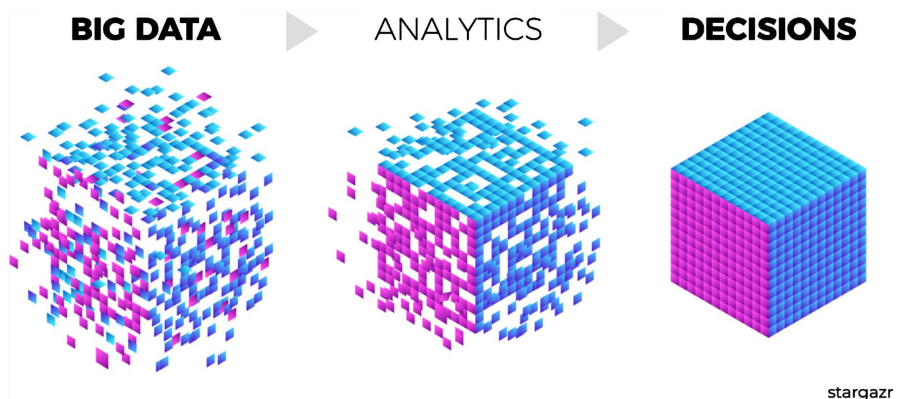


Figure 1. Scheme of big-data analyzing the output [9].

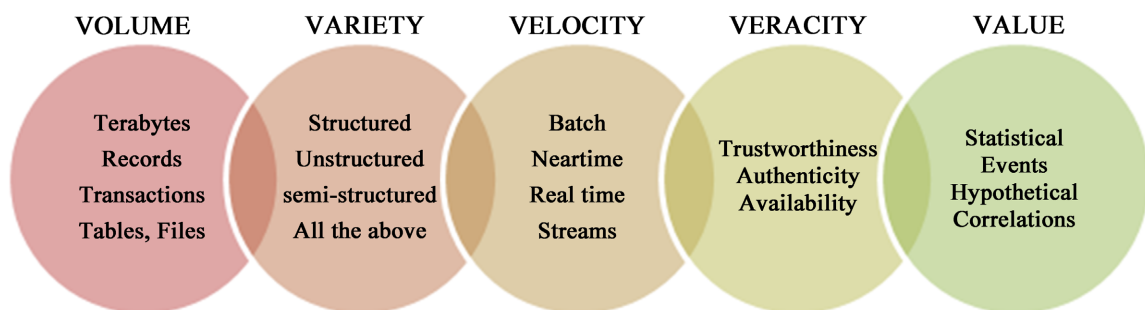


Figure 2. 5V concept. The source [12].

(1 tera = 1000 Giga) and petabytes (1 Peta = 1000 terra). The cost of storing this tremendous amount of data is a hurdle for the data scientist to overcome.

2-Variety

Variety refers to the different data types such as structured, unstructured, and semi-structured data in relational database storage systems. The data format could be in the forms as documents, emails, social media text messages, audio, video, graphics, images, graphs, and the output from all types of machine-generated data from various sensors, devices, machine logs, cell phone GPS signals and more [11].

3-Velocity

The motion of the data sets is a significant aspect to categorize data types based on it. Data-at-rest and data-in-motion is the term that deals with velocity. The major concern is the consistency and completeness of fast-paced data streams and getting the desired result matching. Velocity also includes time and latency characteristics: the data being analyzed, processed, stored, managed, and updated at a first-rate or with a lag time between the events.

4-Value

Value deal with what value should be resulted from a set of data.

5-Veracity

Veracity describes the quality of data. Is the data noiseless or conflict-free? Accuracy and completeness are concerned.

3. Big-Data Analysis

3.1. Viability of Big-Data Analysis

Big data analytics assists businesses in harnessing their data and identifying new opportunities. As a result, smarter business decisions, more effective operations, higher profits, and happier consumers are the result. More than 50 firms were interviewed for the publication Big Data in Big Companies (Figure 3) [13] to

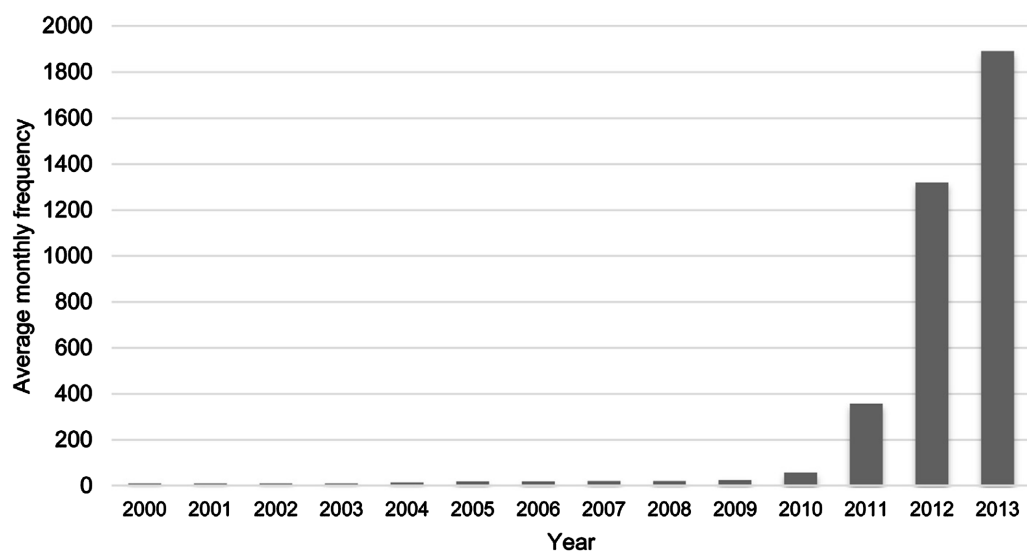


Figure 3. Frequency distribution of documents containing the term “big data” in ProQuest Research Library. The source [6].

learn how they used big data. According to the report, they gained value in the following ways:

Cost reduction. When it comes to storing massive volumes of data, big data technologies like Hadoop and cloud-based analytics provide significant cost savings—and they can also find more effective methods of doing business.

Faster, better decision-making. Businesses can evaluate information instantaneously—and make decisions based on what they’ve learned—thanks to Hadoop’s speed and in-memory analytics, as well as the ability to study new sources of data.

New products and services. With the capacity to use analytics to measure client requirements and satisfaction comes the potential to provide customers with exactly what they want. According to Davenport, more organizations are using big data analytics to create new goods to fulfill the needs of their customers.

3.2. Analyzing Process

3.2.1. Analyzing Steps

Data analysts, data scientists, predictive modellers, statisticians, and other analytics experts collect, process, clean, and analyze increasing volumes of structured transaction data, as well as other types of data not typically used by traditional BI and analytics tools. The four steps of the data preparation process are summarized below (Figure 4) [7]:

1) Data specialists gather information from a range of sources. It’s usually a mix of semi-structured and unstructured information. While each company will use different data streams, the following are some of the most frequent sources::

- clickstream data from the internet

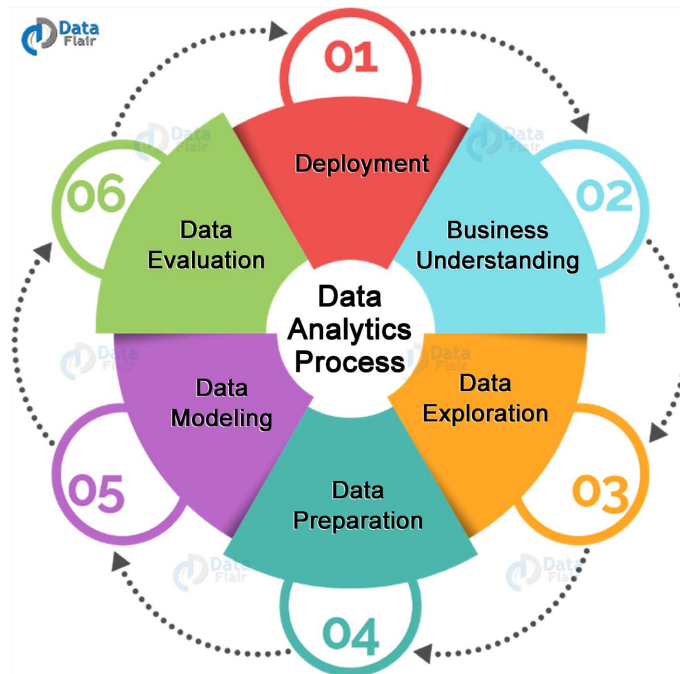


Figure 4. Circular process steps of data analysis [7].

- web server logs
- cloud apps
- mobile applications
- social media content
- text from consumer emails and survey replies
- mobile phone records
- machine data collected by the internet of things sensors (IoT).

2) The information is analyzed. Data professionals must organize, arrange, and segment data effectively for analytical queries after it has been acquired and stored in a data warehouse or data lake. Analytical queries perform better when data is processed thoroughly.

3) The data is filtered to ensure its quality. Data scrubbers use scripting tools or corporate software to clean up the data. They organize and tidy up the data, looking for any faults or inconsistencies such as duplications or formatting errors.

4) Analytics software is used to analyze the data that has been collected, processed, and cleaned. This contains items such as:

- Data mining, which sifts through large data sets looking for patterns and connections.
- Predictive analytics, which involves developing models to predict customer behaviour and other future events.
- Machine learning, which makes use of algorithms to evaluate enormous amounts of data.
- Deep learning, a branch of machine learning that is more advanced.
- Program for text mining and statistical analysis.
- Artificial intelligence (AI).
- Business intelligence software that is widely used.

3.2.2. Analyzing Tools

To support big data analytics procedures, a variety of tools and technologies are used [14]. The following are some of the most common technologies and techniques used to facilitate big data analytics processes:

- **Hadoop** is a free and open-source framework for storing and analyzing large amounts of data. Hadoop is capable of storing and processing enormous amounts of structured and unstructured data.
- **Predictive analytics** large volumes of complicated data are processed by hardware and software, which uses machine learning and statistical algorithms to forecast future event outcomes. Predictive analytics technologies are used by businesses for fraud detection, marketing, risk assessment, and operations.
- **Stream analytics** are used to filter, combine, and analyze large amounts of data in a variety of formats and platforms.
- **Distributed storage data** is usually replicated on a non-relational database. This can be a safeguard against independent node failures, the loss or corruption of large amounts of data, or the provision of low-latency access.
- **NoSQL databases** non-relational data management methods come in handy

when dealing with vast amounts of scattered data. They are appropriate for raw and unstructured data because they do not require a fixed schema.

- **A data lake** is a big storage repository that stores raw data in native format until it's needed. A flat architecture is used in data lakes.
- **A data warehouse** is a data repository that holds vast amounts of data gathered from many sources. Predefined schemas are used to store data in data warehouses.
- **Knowledge discovery/big data mining tools** businesses will be able to mine vast amounts of structured and unstructured big data.
- **In-memory data fabric** large volumes of data are distributed across system memory resources. This contributes to minimal data access and processing delay.
- **Data virtualization** allows data to be accessed without any technical limitations.
- **Data integration software** enables big data to be streamlined across different platforms, including Apache, Hadoop, MongoDB, and Amazon EMR.
- **Data quality software**, cleans and enriches massive amounts of data
- **Data preprocessing software**, which prepares data to be analyzed further. Unstructured data is cleared and data is prepared.
- **Spark**: which is a free and open-source cluster computing platform for batch and real-time data processing.

3.2.3. Different Types of Big Data Analytics

Here are the four types of Big Data analytics:

1) Descriptive Analytics

This summarizes previous data in an easy-to-understand format. This aids in the creation of reports such as a company's income, profit, and sales, among other things. It also aids in the tally of social media metrics.

2) Diagnostic Analytics

This is done to figure out what created the issue in the first place. Drill-down, data mining, and data recovery are all instances of techniques. Diagnostic analytics is used by businesses to gain a deeper understanding of a problem.

3) Predictive Analytics

This sort of analytics examines past and current data to create predictions. Predictive analytics analyzes current data and makes forecasts using data mining, artificial intelligence, and machine learning. It predicts customer and market trends, among other things.

4) Prescriptive Analytics

This type of analytics recommends a remedy to a specific issue. Both descriptive and predictive analytics are used in perspective analytics. Most of the time, AI and machine learning are used.

3.3. Big Data Analytics Benefits

Big data analytics has several advantages, including the ability to swiftly evaluate

massive amounts of data from numerous sources in a variety of forms and types (Figure 5).

Making better-informed decisions more quickly for more successful strategizing, can benefit and improve the supply chain, operations, and other strategic decision-making sectors.

Savings that can be realized because of increased business process efficiencies and optimizations.

Greater marketing insights and information for product creation can come from a better understanding of client demands, behaviour, and sentiment.

Risk management tactics that are improved and more informed as a result of huge data sample sizes [15].

3.4. Big Data Analytics Challenges

Despite the numerous advantages of utilizing big data analytics, it is not without its drawbacks [16]:

- Data accessibility. Data storage and processing grow increasingly difficult as the number of data increases. To ensure that less experienced data scientists and analysts can use big data, it must be appropriately stored and managed.
- Ensuring data quality. Data quality management for big data necessitates a significant amount of time, effort, and resources due to the large volumes of data coming in from multiple sources and in varied forms. it.
- Data protection. Large data systems pose unique security challenges due to their complexity. It might be difficult to properly handle security risks within such a sophisticated big data ecosystem.
- Choosing the appropriate tools. Organizations must know how to choose the appropriate tool that corresponds with users' needs and infrastructure



Figure 5. Benefits of big-data analytics [15].

from the huge diversity of big data analytics tools and platforms available on the market.

- Some firms are having difficulty filling the gaps due to a probable lack of internal analytics expertise and the high cost of acquiring professional data scientists and engineers.

3.5. The difference between Data Analytics and Big Data

1) Nature: Let's look at an example of the key distinction between Big Data and Data Analytics. Data Analytics is similar to a book where you can discover solutions to your problems; on the other hand, Big Data can be compared to a large library with all the answers to all the questions, but it's tough to locate them.

2) Structure of Data: Data analytics reveals that the data is already structured, making it simple to discover an answer to a question. However, Big Data is a generally unstructured set of data that must be sorted through to discover an answer to any question and processing such massive amounts of data is difficult. To gain useful insight from Big Data, a variety of filters must be used.

3) Tools used in Big Data vs. Data Analytics: Because the data to be analyzed is already structured and not difficult, simple statistical and predictive modelling tools will be used. Because processing the vast volume of Big Data is difficult, advanced technological solutions such as automation tools or parallel computing tools will be required to manage it. More information about Big Data Tools can be found here.

4) Type of Industry using Big Data and Data Analytics:

Industries such as IT, travel, and healthcare are among the most common users of data analytics. Using historical data and studying prior trends and patterns, Data Analytics assists these businesses in developing new advances. Simultaneously, Big Data is utilized by a variety of businesses, including banking, retail, and others. In a variety of ways, big data assists these industries in making strategic business decisions.

Application of Data Analytics and Big Data

Data is the foundation for all of today's decisions. Today, no choices or actions can be taken without the data. To achieve success, every company now employs a data-driven strategy. Data Scientists, Data Experts, and other data-related careers abound these days.

Job Responsibilities of Data Analysts

1) Analyzing Trends and Patterns: Data analysts must foresee and predict what will happen in the future, which can be very useful in company strategic decision-making. In this situation, a data analyst must recognize long-term trends [17]. He must also give precise recommendations based on the patterns he has discovered.

2) Creating and Designing Data Report: A data scientist's reports are a necessary component of a company's decision-making process. Data scientists will

need to construct the data report and design it in such a way that the decision-maker can understand it quickly. Data can be displayed in a variety of ways, including pie charts, graphs, charts, diagrams, and more. Depending on the nature of the data to be displayed, data reporting can also be done in the form of a table.

3) Deriving Valuable Insights from the Data: To benefit the organizations, Data Analysts will need to extract relevant and meaningful insights from the data package. The company will be able to use those valuable and unique insights to make the greatest decision for its long-term growth.

4) Collection, Processing, and Summarizing of Data: A Data Analyst must first collect data, then process it using the appropriate tools, and finally summarize the information such that it is easily comprehended. The summarized data can reveal a lot about the trends and patterns that are used to forecast and predict things.

Job Responsibilities of Big Data Professionals

1) Analyzing Real-time Situations: Big Data professionals are in high demand for analyzing and monitoring scenarios that occur in real-time. It will assist many businesses in taking immediate and timely action to address any issue or problem, as well as capitalize on the opportunity [18]. Many businesses may cut losses, boost earnings, and become more successful this way.

2) Building a System to Process Large Scale Data: Processing large amounts of data promptly is difficult. Unstructured data that cannot be processed by a simple tool is sometimes referred to as Big Data. A Big Data Professional must create a complex technological tool or system to handle and analyze Big Data to make better decisions [19].

3) Detecting Fraud Transactions: Fraud is on the rise, and it is critical to combat the problem. Big Data experts should be able to spot any potentially fraudulent transactions. Many sectors, particularly banking, have important duties in this area. Every day, many fraudulent transactions occur in financial sectors, and banks must act quickly to address this problem. People will lose trust in the financial system if they continue to save their hard-earned money in banks.

4. Conclusions

Gradually, the business sector is relying more on its development on data science. A tremendous amount of data is used to describe the behaviour of complex systems, anticipate the output of processes, and evaluate this output. Based on what we discussed in this essay, it can be stated that Big-data analytics is the cutting-edge methodology in data science alongside every other technological aspect, and studying comprehensively this major, would be essential for further development.

Several methods and software are commercially available for analyzing big-data sets. Each of them can relate to technology, business, or social media. Further studies using analyzing software could enhance the depth of the know-

ledge reported and validate the results.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Siegfried, P. (2017) Strategische Unternehmensplanung in jungen KMU—Probleme and Lösungsansätze. de Gruyter/Oldenbourg Verlag, Berlin.
- [2] Siegfried, P. (2014) Knowledge Transfer in Service Research—Service Engineering in Startup Companies. EUL-Verlag, Siegburg.
- [3] Divesh, S. (2017) Proceedings of the VLDB Endowment. *Proceedings of the VLDB Endowment*, **10**, 2032-2033.
- [4] Su, X. (2012) Introduction to Big Data. In: *Opphavsrett: Forfatter og Stiftelsen TISIP*, Institutt for informatikk og e-læring ved NTNU, Zürich, Vol. 10, Issue 12, 2269-2274.
- [5] Siegfried, P. (2015) Die Unternehmenserfolgsfaktoren und deren kausale Zusammenhänge. In: *Zeitschrift Ideen- und Innovationsmanagement*, Deutsches Institut für Betriebs-wirtschaft GmbH/Erich Schmidt Verlag, Berlin, 131-137.
<https://doi.org/10.37307/j.2198-3151.2015.04.04>
- [6] Gandomi, A. and Haider, M. (2015) Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, **35**, 137-144.
<https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- [7] Lembo, D. (2015) An Introduction to Big Data. In: *Application of Big Data for National Security*, Elsevier, Amsterdam, 3-13.
<https://doi.org/10.1016/B978-0-12-801967-2.00001-X>
- [8] Siegfried, P. (2014) Analysis of the Service Research Studies in the German Research Field, Performance Measurement and Management. Publishing House of Wroclaw University of Economics, Wroclaw, Band 345, 94-104.
- [9] Cheng, O. and Lau, R. (2015) Big Data Stream Analytics for Near Real-Time Sentiment Analysis. *Journal of Computer and Communications*, **3**, 189-195.
<https://doi.org/10.4236/jcc.2015.35024>
- [10] Abu-salih, B. and Wongthongtham, P. (2014) Chapter 2. Introduction to Big Data Technology. 1-46.
- [11] Sharma, S. and Mangat, V. (2015) Technology and Trends to Handle Big Data: Survey. *International Conference on Advanced Computing and Communication Technologies*, Haryana, 21-22 February 2015, 266-271.
<https://doi.org/10.1109/ACCT.2015.121>
- [12] Davenport, T.H. and Dyché, J. (2013) Big Data in Big Companies. *Baylor Business Review*, **32**, 20-21.
http://search.proquest.com/docview/1467720121?accountid=10067%5Cnhttp://sfx.lib.nccu.edu.tw/sfxlcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ:abiglobal&atitle=VIEW/REVIEW:+BIG+DATA+IN+BIG+COMPANIES&title=Bay
- [13] Riahi, Y. and Riahi, S. (2018) Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering*, **5**, 524-528.
<https://doi.org/10.21276/ijre.2018.5.9.5>
- [14] Verma, J.P. and Agrawal, S. (2016) Big Data Analytics: Challenges and Applications

-
- for Text, Audio, Video, and Social Media Data. *International Journal on Soft Computing, Artificial Intelligence and Applications*, **5**, 41-51. <https://doi.org/10.5121/ijscai.2016.5105>
- [15] Begoli, E. and Horey, J. (2012) Design Principles for Effective Knowledge Discovery from Big Data. *Proceedings of the 2012 Joint Working Conference on Software Architecture and 6th European Conference on Software Architecture, WICSA/ECSA*, Helsinki, 20-24 August 2012, 215-218. <https://doi.org/10.1109/WICSA-ECSA.212.32>
- [16] Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R. and Muharemagic, E. (2015) Deep Learning Applications and Challenges in Big Data Analytics. *Journal of Big Data*, **2**, 1-21. <https://doi.org/10.1186/s40537-014-0007-7>
- [17] Bätz, K. and Siegfried, P. (2021) Complexity of Culture and Entrepreneurial Practice. *International Entrepreneurship Review*, **7**, 61-70. <https://doi.org/10.15678/IER.2021.0703.05>
- [18] Bockhaus-Odenthal, E. and Siegfried, P. (2021) Agilität über Unternehmensgrenzen hinaus—Agility across Boundaries, Bulletin of Taras Shevchenko National University of Kyiv. *Economics*, **3**, 14-24. <https://doi.org/10.17721/1728-2667.2021/216-3/2>
- [19] Kaisler, S.H., Armour, F.J. and Espinosa, A.J. (2017) Introduction to Big Data and Analytics: Concepts, Techniques, Methods, and Applications Mini Track. *Proceedings of the Annual Hawaii International Conference on System Sciences*, Hawaii, 4-7 January 2017, 990-992. <https://doi.org/10.24251/HICSS.2017.117>