Знание и бизнес, Knowledge and business

Икономика и компютърни науки
Economics and computer science

Научно списание

A scientific journal

бр. 1 / 2021

Volume: 2021, Issue: 1

Издателство „Знание и бизнес“ Варна
Publishing house „Knowledge and business“ Varna

ECONOMICS AND COMPUTER SCIENCE

Научно списание

A scientific journal

Брой 1 / 2021

Volume: 2021, Issue: 1

© НАУЧНО СПИСАНИЕ ИКОНОМИКА И КОМПЮТЪРНИ НАУКИ

©SCIENTIFIC JOURNAL ECONOMICS AND COMPUTER SCIENCE

брой 1, 2021 г.

Volume: 2021, Issue: 1

ISSN 2367-7791, Varna, Bulgaria

ЦЕНТЪР ЗА ПРОФЕСИОНАЛНО ОБУЧЕНИЕ „ЗНАНИЕ И БИЗНЕС“

Главен редактор: проф. д-р ик. н. Петко Илиев – тел: 0893 492 350

Зам. гл. редактор: доц. д-р Светлозар Стефанов – тел: 0878 362 245

Зам. гл. редактор: проф. д-р Юлиан Василев, vasilev@ue-varna.bg

Научен секретар: доц. д-р Анастасия Кондукторова – тел: 0882 164 846

Предпечат: Юлиан Василев

CENTER FOR PROFESSIONAL EDUCATION "KNOWLEDGE AND BUSINESS"

Editor-in-Chief: Prof. Dr.Sc. Petko Iliev - tel: 0893 492 350

Editor: Assoc. Prof. PhD Svetlozar Stefanov - phone: 0878 362 245

Editor: Prof. PhD Julian Vasilev, vasilev@ue-varna.bg

Scientific Secretary: Assoc. Prof. PhD Anastasia Konduktrova - tel: 0882 164 846

Prepress: Julian Vasilev

9000, Varna, 1, Dr. L. Zamenhof Str.

email: eknigibg@gmail.com

CONTENTS

Dear readers, we are pleased to inform you that on the initiative of the Editorial Board of "Knowledge and Business" Ltd., "Knowledge and Business" is a licensed publisher of the scientific journal "Economics and Computer Science".

Our thesis is that information technologies are used in all areas of human activity. That is why the scientific journal "Economics and Computer Science" is open to publications from various fields of science and practice.

We are confident that the publications in the journal contribute as enrichment of knowledge and to enhance the effectiveness of the use of information technology in business.

Sincerely: Prof. Dr. Sc. Petko Iliev

Big data processing in the logistics industry

Snezhana SULOVA[1]

[1] University of Economics, Varna, Bulgaria
ssulova@ue-varna.bg

Abstract. The aim of modern logistics is to achieve maximum connectivity in the supply chain. Companies are using increasingly innovative technological solutions, which creates the opportunity of generating a wide variety of data. This leads to several challenges and the need to change data storage and processing models. The aim of the study is to analyze the technological aspects of the digital transformation in logistics and to propose a conceptual framework for big data management and processing in the logistics industry. It is based on the discovery of existing prototype methodologies for big data processing which are used in all areas of business, as well as on the research of existing specific approaches to the processing of different types of big data in logistics. Basic principles for building a modern architecture for managing and processing big data in logistics are presented. The defined framework can be used by the companies to process structured, semi-structured and unstructured data in real time or for batch processing and to help optimize several business processes in the logistics industry. As a result, using it will help the analytical processes in these companies and it will be possible to make informed business decisions in dynamic conditions and in globalization. A software implementation of a conceptual framework with the Apache Handoop open-source software is proposed. The study is part of Project BG05M2OP001-1.002-0002-C02 "Digitalization of Economy in a Big Data Environment"

Key words: logistics, big data, digitization, conceptual framework.

1. Introduction

Developments in the modern economy are mainly due to the process of digitalization. According to current research, 87% of small companies and 6% of large companies in Western Europe have invested in the digitalization and automation of their production (pwc.de, 2017). In the field of logistics, old ways of distributing physical goods have been replaced with modern solutions, which are based upon the application of innovative technologies that have been integrated into the existing supply chain management (SCM) or have created completely new working concepts.

The use of innovative digital solutions, with the goal of optimizing and changing the functioning model of business, creates enormous prospects for businesses to develop, but at the same time, the new models of work also bring with it many challenges.

Modernization, digitization, and cloud computing have resulted in the appearance of ever greater volumes of data. The creation and use of data is growing rapidly throughout the world. In 2020, data volume was 59 zettabytes, and it is expected to be 149 zettabytes by 2024 (statista.com, 2021). On the one hand, this creates more opportunities to improve business intelligence analyses, which could be provided in real-time allowing for informed decisions to be made, but on the other hand, it will create a need to change management models and how organizations process data.

The goal of the current research is to analyze the technological aspects of the digital transformation of logistics, by focusing on the increasing volumes of data and to propose a conceptual framework for the management and processing of data in the logistics industry.

2. Technological aspects of the digital transformation of logistics

The initiative “Industry 4.0” began with the acceptance of a high-tech strategy for the development of industry in Germany (Lydon, 2014). Following that decision, many studies have analyzed the content, scope and technology that led to a fourth industrial revolution. Kondratiev (2019) points out that the revolution is due to: the use of integrated sensors through which devices are able to communicate with each other or better known as the Internet of Things (IoT) technology; working in real-time with the goal of optimizing expenses and the

quality of production through the application of Big Data Analytics (BDA); autonomous robots and 3-D printers. Another comprehensive study of the main technological trends of Industry 4.0 identifies: virtual and augmented reality (VR, AR); cloud computing; robotization; BDA; cyber-physical system (CPS); cybersecurity; Internet of People (IoP), Internet of Services (IoS); Industrial Internet of Things (IIoT); semantic technologies; simulation and modeling (Ghobakhloo, 2020).

The processes of digital transformation in logistics led to the emergence of the Logistics 4.0 concept. The term, Logistics 4.0, means the specific application of Industry 4.0 to the field of logistics (Raksoy, Koçhan, Ali, 2020, p. 21).

The range of technologies, which are the basis for the transformation of traditional organizations into smart businesses, are constantly growing (Lazarova, 2019). It is necessary to note that in order to create a digitally connected world, a complete change in business models is needed. Turchi (2018) has developed a framework for the digital transformation of a business, which is called “The Pyramid of Digital Transformation”. He defines 3 levels: technologies, new business models and strategies for development (Fig. 1).

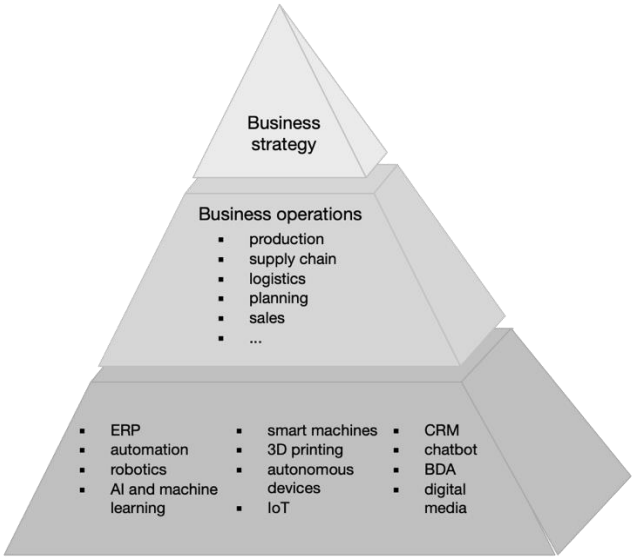


Figure 1. Pyramid of Digital Transformation
Source: (Turchi, 2018)

Each of the three levels of the pyramid affects each of the other elements of the framework and therefore, a successful digital transformation must use an integrated method and must create new strategies through Industry 4.0.

The focus of our research is the technological changes, which form the necessary base for the application of modern concepts to logistics activities. The key technological changes are**: vertical and horizontal system integration; Internet of Things; cloud computing; blockchain; autonomous robots; virtual and augmented reality** (Fig. 2).

The processes of digital transformation of logistics requires system connectivity and the presence of vertical integration as is related to the systems of the organization and to what extent they are integrated with each other as well as the horizontal integration with business partners. The creation of large, distributed, complex systems, which work flexibly as a self-organized structure using a new type of production facility are the foundation of **cyber-physical systems**. CPS ensures flexible mass production on demand and flexibility in the quantities produced (Rojko, 2017). Some authors point out that CPS has the following main characteristics: heterogeneous character; connectivity and networking abilities; as well as their ability to work based on the “Software as a Service” model; modularity; autonomy; decentralization; real-time operation; powerful computing abilities; dynamic possibilities for reconfiguration and adaptability (Napoleone, Macchi, Pozzetti, 2020).

With the development of CPS technologies over the past few years, physical devices, which use computers and networks to expand their functions, have led to modernized industrial products and technologies and have significantly improved their ability to compete in major industrial sectors (Zhang, D. et al., 2021)

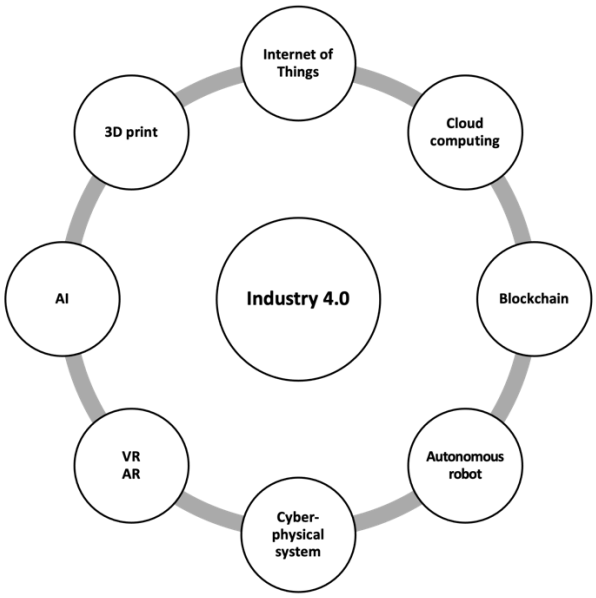


Figure 2. Key components for the digital transformation of logistics (Logistica 4.0)
Source: Own elaboration

There is a direct link between CPS and the **IoT** concept, whose main function is the collection of data from various physical entities using devices that have access to the internet. IoT “proposes the construction of communities in which the objects around us - often called ‘intelligent objects’ – can be connected through the Internet” (Stoyanov and Popchev, 2017). Thanks to these technologies, the concepts of smart production, smart buildings, automobiles, and logistics have been developed. Smart logistics pertains to the precision, reliability and effectiveness of logistics performance due to improved information and the use of data (Chaopaisarn & Woschank, 2020). In the field of logistics, smart logistics is related to real-time work, working conditions and the maintenance history of the production equipment. In the field of logistics, IoT can track the routes and alert to any deviations from the planned route, analyze the location of vehicles and route data and upon discovering difficulties in the supply chain, such as a traffic jam, it can generate alternative routes to optimize the supply chain.

Statistical data shows that at the end of 2018 22 billion IoT devices had been used throughout the world, and the data predicts that by 2030 there will be about 50 billion creating an extensive network of connected devices (Statista.com, 2021).

The development of the IoT concept has directly contributed to an increase in data volume coming from various sensors, GPS location tracking devices, radio frequency identification (RFID) tags, and others. This has led to an increase in demand for computer (computing) resources for storing and processing information. The cloud-based business model provides an opportunity for small and medium-sized businesses to meet the challenge of managing ICT infrastructure, platforms and services (mi.government.bg, 2017). Data shows that in 2020 more than 1/3 of EU businesses used cloud computing mainly for email services and file storage (Eurostat, 2021).

Technological factors that stimulate the development of cloud computing are: distributed processing; wide area networks and internet; high-performance hardware and virtualization (Emilova, 2016). Cloud computing can be found in every sector of the economy due to its advantages, such as easy access to shared resources, optimization of expenses, flexibility, and virtualization of resources.

Blockchain technology also plays a leading role in the process of digitizing logistics. The Gartner Research and Consulting company pointed out that blockchain was one of the ten most important strategic technologies for 2020 (Gartner, 2020). Blockchain is connected to the development of decentralized applications for the automation of business processes and data processing (Filipova, 2018, p. 83). It can be utilized in financial and insurance fields (Kirov, 2020), for “smart contracts”, for retail operations and the supply chain of products, goods and services (Dimitrova and Semova, 2018), property management (Stoyanova, Vasilev, Cristescu, 2021) as well as others.

Robots are used in many industries for completing complex tasks. The increased use of industrial robots by companies has accelerated the process of digital transformation. Intelligent technologies provide devices with new qualities and possibilities. Robotized processes can be found in the treatment of chemicals, the production of pharmaceuticals, food and drinks, and in logistics (Proctor and Wilkins, 2019).

Augmented virtual reality also contributes to the formation of an intelligent business. AR combines the physical and the real world with the computer world of generated data, with the goal of turning a virtual object into reality. **3D printing** (additive manufacturing) is important for some industries whose work necessitates the building of different complicated geometric constructions. In the past few years, larger investments have been made in 3D printing and studies show that by 2023 these investments will reach $32,78 billion (Proctor and Wilkins, 2019). Important benefits of 3D printing include decreasing the weight of components and waste reduction.

All of the technologies mentioned, combined with the possibilities, which intelligent technologies have for adapting their behavior and for autonomous work on the basis of the analysis of results from previous actions is fundamental for the formation of a technological base for the processes in digital transformation.

To summarize, it can be noted that the processes of greater interconnection between information systems, the presence of dynamic activities, the market conditions, the use of constantly enriching number of technological solutions in logistics has led to the generation, storage and processing of large volumes of data.

3. Data as a condition of digital transformation in logistics

The word “data” comes from the Latin word dare, which means “something, which is given” - observation or fact about a given object (Kotu and Deshpande, 2019, p. 39). Data is a vital and irreplaceable commodity for every organization. Today businesses work with operational data, as well as with large volumes of data, which are generated with the help of internet applications; transactions, which happen in real-time; publications in social media; pictures; mobile phone messages, etc. This necessitated the emergence of the concept “Big Data”.

Big data is a collection of chosen data, whose volume, speed or variety are so large that it is difficult to store, manage, process and analyze the data using traditional databases and instruments for data processing (Bahga and Madisetti, 2019, p. 25). The Gartner consulting company has defined three main characteristics of big data known as the so-called “3V” model:

- Volume. Data volume can be distinguished as the most essential characteristic of big data. As it has already been pointed out, all the digital transformation processes lead to the generation of large volumes of new data. Internet, social networks, use of mobile devices, measuring devices are all capable of generating large volumes of data.
- Velocity. Currently, data appears extremely fast. With the increased use of sensors and devices working with IoT, data is generated and transferred with great speed, which creates the need to process it in a timely manner, in real-time mode;
- Variety. Big data has a heterogeneous character. Data is in different formats - operational data from databases, text documents, emails, videos, audio files, sensor data, etc.

Some authors (Sharda, Delen, Turban, 2020, p. 125) divide data into three different types depending upon the data structure as it can be seen in Fig. 3.

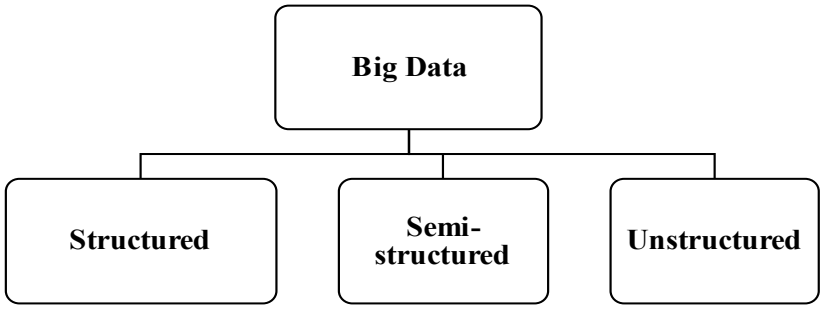


Figure 3. Types of Big Data
Source: (Sharda, Delen, Turban, 2020, p. 125)

Structured data has a form, which is suitable for computer processing. Often the data is stored and processed by Database Management System (DBMS). The data is generated by ERP, CRM or other business systems, sensor devices working with IoT technology, and surveys. **Semi-structured data** as a whole has some sort of organizational structure that eases its ability to be processed. It includes tags and elements (metadata), which are used to group and describe them. Typical examples are JSON, XML and CVS files.

Unstructured data does not have a specific structure or model of organization. It is in the form of text files, email messages, devices for video and audio recordings. According to prevailing assessments, unstructured data constitutes 80-90% of all data (Cio.com, 2019).

Other authors identify data according to its source and categorize it as: social media data; smart and IoT devices data, sensor data and transaction data (Storeya & Song 2017).

In order to better define the characteristics of big data including the challenges faced in processing big data, seven characteristics have been added to the three main ones. Firican (2017) uses the “10V” model to describe big data. In table 1, the seven additional distinguishing features of big data are given.

Table 1.

Characteristics of big data

Characteristics	Description
Value	It shows that with proper processing and analysis of big data, significant benefits can be obtained.
Variability	Data has many dimensions due to its varying types and sources.
Veracity	Data must have a reliable source and origin in order to be relevant for analysis.
Validity	Data must be accurate and true.
Vulnerability	Describes how long data that has been obtained from the original source has been available and how long it should be stored.
Volatility	Describes how long the obtained data is and how long it should be stored.
Visualization	Shows how the data processing results are presented.

Source: Own elaboration

With an aim of characterizing big data in greater detail and defining the challenges connected with its storage and processing, Shafer (2017) has presented an updated list with 42 V’s.

Big data from logistics is generated by the following sources:

- **Traditional computer systems**, which maintain the logistics business processes. They are: Enterprise Resource Planning systems (ERP), Warehouse Management Systems (WMS), Transportation Management Systems (TMS), Customer Relationship Management systems (CRM), Supply Chain Management systems (SCM) and others. A detailed overview of information systems used by modern logistics as well as the volume of information in supply chains has been made by Vasilev (2017).
- **Data from Geographic Information Systems** (GIS), which has a rather diverse nature - numerical values, text, images (pictures), graphics, audio, animation and video (Gergova et al., 2017). Global Positioning System (GPS) is of great importance to transport and logistics companies, as it allows for the development and application of spatial and temporal models for the optimization of the routes.
- **Data from devices using IoT** concepts. IoT systems work with a variety of large volumes of data, which is collected from sensors, RFID and mobile devices. Internet of things is one of information technologies which can be used in 5PL logistics to collect and analyze large quantities of data (Zaychenko et. al, 2021). The data is most often found in the XML format (eXtensible Markup Language), JSON (Javascript Object Notation), PNG (Portable Network Graphics), CSV (Comma-separated values), XDR (eXternal Data Representation), RDF (Resource Description Framework);
- **Data from Augmented Reality** technology applications, which can be integrated into vehicles and provide a video stream of data and computer-generated graphics.
- **Company websites, web and mobile applications, social media profiles**. Web resources, sites, web applications and mobile applications are sources of various data. On the one hand, when one interacts with them usable data is generated, and on the other hand, these sources, themselves, have their own content and they also allow the creation of new content, which can also be a source of data. Currently, a tendency is being observed that traffic from wireless and mobile devices is quickly increasing and this generates new data for the analysis of users’ behavior.

- **Text documents and email messages**. Text documents with different structures and formats are used and exchanged in the logistics industry. Transforming these documents using natural language processing (NLP), they can become a useful source of knowledge.
- **Data, collected by governments through open databases**, such as those connected to intellectual property, civil infrastructure, scientific developments, and protection of the environment.
- **Data, connected with government regulation** – patents, regulatory activities, tax information.
- **Other data** from accompanying activities, such as data from advertising campaigns for example.

The importance of big data is since analyses can be performed on it which an organization will value. Many businesses are currently unable to process the large volume of data that they possess, or they are unable to cope with the speed with which new data is produced by sensors and social networks. Although traditional relational data management systems ensure speed in processing transactions, they do not have the necessary scalability to work with semi-structured and unstructured data. The digital environment, as well as the peculiarity and variety of work with large volumes of data requires the use of specific and improved new approaches for the processing and storage of the data. Several technologies exist for the storage, processing and analysis of big data. MapReduce, Hadoop and NoSQL have been highlighted by analysts as the most important technologies for digital transformation (Sharda, Delen, Turban, 2020, p. 523).

MapReduce is a computing technology, which can be implemented on many systems. “It can be viewed as a programmed data-processing module, whose purpose is to compact large quantities of information into useful summary results” (Petrov and Trifonova, 2019, p. 127). MapReduce performs two tasks: Map and Reduce. The first phase is to process inputted data. In order to complete this phase, one of the computers is assigned to be the main node and it receives the inputted data, divides it into parts and sends it to other computers (working nodes) for initial processing. In the Reduce phase, previously processed data is reduced. The main node organizes the results - the solution to the problem, which was initially formulated- based upon the answers received from the working nodes (Fig. 4).

High performance is reached when the process is reduced to small units of work, which can be completed in parallel across hundreds, potentially thousands, of cluster nodes. This process helps in processing and analyzing large volumes of data. MapReduce is available both commercially and for libraries with open code providing a wide spectrum of analytical possibilities. Apache Mahout, for example, is a library for open source machine learning, which includes cluster algorithms, classification and it works with MapReduce.

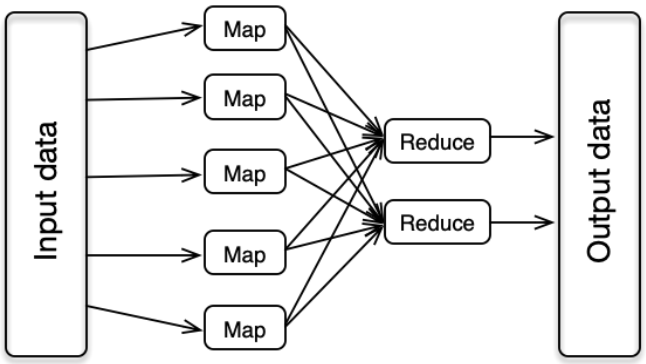


Figure 4. MapReduce Technology
Source: Own elaboration

Hadoop is a software framework with open code for processing, storing and analyzing large quantities of distributed unstructured data. It was created by Doug Cutting, Mike Cafarella and is now a software project of the Apache Foundation, who are continuously working to perfect it. Hadoop has two main components: Hadoop Distributed File System – HDFS and an environment for MapReduce inquiries. The client sends a Map query written in Java to one of the cluster nodes (Job Tracker), it initiates, coordinates the tasks and sends a query to the respective nodes. Processing happens on each node simultaneously or in parallel.

The Hadoop software instrument package includes NoSQL databases like Cassandra and HBase, which are also used for storing the results of MapReduce tasks in Hadoop. In addition, it has a specially created open-source language Pig, which can be used for some editions of MapReduce. Another instrument is Hive – database with open source, initially created for Facebook, which allows for analytical modeling in Hadoop. The most commonly used software tools that are part of Hadoop are summarized in Table 2.

Table 2.

Main software tools in Hadoop

Software Tool	Description
Apache Hive	System for data storage, based on Hadoop. Allows the user to write tasks in a language, similar to SQL, called HiveQL, which are then converted for MapReduce.
Apache Pig	Apache Pig is a platform for program creation, which works with Apache Hadoop. The language, which this platform uses, is called Pig Latin.
Apache HBase	DBMS for non-relational databases. It works with column databases - the data is organized as families of columns.
Apache Flume	Flume is software for populating Hadoop with data. The data can come from a log file, data generated by user behavior, from clicks on links, generated from various sensors, etc.
Apache Oozie	Workflow processing system, which allows one to define a series of tasks written in MapReduce, Pig and Hive and then to connect them.
Apache Ambari	Web based selection of instruments for installation, administration and observation of Apache Hadoop clusters.
Apache Avro	Data serialization system using JSON for describing data in compact binary format.
Apache Mahout	Knowledge extraction library based on machine learning algorithms.
Apache Sqoop	SQOOP is a software tool for transferring data between relational databases in Hadoop.
Apache HCatalog	Centralized service for management and sharing of metadata for Apache Hadoop.

Source: Own elaboration

The third main concept – **non-relational databases** (NoSQL – Not only Structure Query Language) have developed as a consequence of the emergence of the processes of semi-structured and unstructured data and the need for software that can process it quick enough. Some authors (Petrov and Trifonova, 2019) point out that an important aspect of NoSQL is its flexible schemes for data management, which can be easily changed, and which do not create obstacles for already stored data (Petrov and Trifonova, 2019). Examples of NoSQL databases are HBase, MongoDB, Cassandra, Accumulo, Riak, CouchDB, DynamoDB. In Fig. 5, a summary of existing NoSQL database models is given (Mazumdar, et al. 2019).

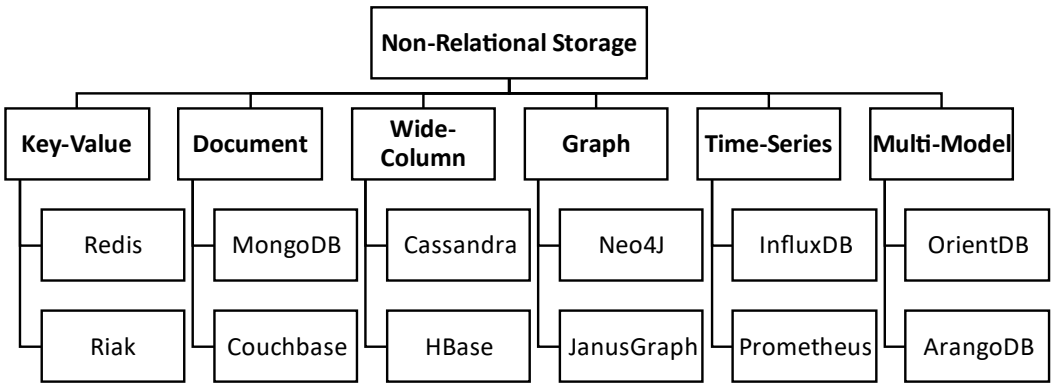


Figure 5. Classification of NoSQL
Source: (Mazumdar, et al. 2019)

NoSQL databases have characteristics that allow for scalability, work in a cloud environment and the use of parallel processing. Their other defining characteristic is that they can be used to work with real-time information flows.

Big data has also led to a change in the concepts used for data storage. Traditional repositories, based on Data Warehouse (DW) have been improved with the goal of being able to work with new types of data. The data lake (DL) concept has also appeared, which allows for the relatively inexpensive storage of different types of data, and subsequently, the application of differentiated methods of analysis.

Large volumes of fast incoming data have also led to a change in the processes of retrieval, conversion and loading of data into repositories. The traditional ETL (Extract, Transform, Load) process is when a conversion

happens. Before the data is loaded it is converted into ELT (Extract, Load, Transform) data first and then transferred to the repository where it is transformed.

In conclusion, we can point out that there are many various data sources for logistics. Among the data sources are not only complex information systems, but also real-time data from different measuring instruments, events, radio frequency identifiers, cell networks, video surveillance devices, social networks and other internet sources. The research gives us reason to conclude that in these areas there is a tendency toward an ever-increasing volume of stored data, as well as the use of data from various new sources. This leads to the need for companies to seek out and implement innovative models for the storage and processing of this data so that they can then apply appropriate modern forms of business analysis.

4. Conceptual framework for the management and processing of logistical data

The researched literature sources indicate that there are some methods, which have been developed for data processing in the logistics industry. Solutions for collecting logistics data, the completion of analysis in real-time and forecasting have been offered by a group of scientists (AlShaer, et al., 2019). The framework is called IBRIDIA and it represents a further development of their approach called ProLoD. IBRIDIA is a hybrid solution, which can be used for processing logistics data in real-time and in batch style. It allows one to collect logistics data in real-time from multiple heterogeneous sensors, social media and business processes, as well as effective data processing in real-time or in batch style and modelling and analysis of data for forecasting delays. The suggested method of IBRIDIA includes: preparation of data – collection, filtration, cleaning, integration; processing of data packets; real-time data processing; data storage.

Other authors have proposed an architecture for the platform, which is called SWeTI and they have determined that it is suitable for application in the organizations of the Industry 4.0 concept (Patel, et al., 2018). Their solution has 5 layers:

- device layer, which includes all the machines working according to an IoT standard;
- edge layer, which is based upon a network protocol for data volume and techniques for data filtration and cleaning for data refinement;
- cyber layer, which represents a center for data distribution. Stores data from different distributed sources and prepares it for processing;
- data analytic layer, distributed data lake, which facilitates analytical activities by using AI algorithms;
- application layer, presents the knowledge so that experts can make the right decisions.

In essence, the presented model is good, but we think that there is a bit of discrepancy between the name of layers and their actual function.

Another suggested solution for processing logistics data is the model of Haße and others (2019), which is a “digital twin” type. Digital twin is defined as a continuously changing digital profile, which contains historical data and new data for a given physical object or process. The main goal of this model is to optimize the effectiveness of a business, based on a large quantity of accumulated data obtained from measurements of a number of objects in the real world. The analysis of the accumulated data allows one to obtain accurate information about the operation of the system, which also provides the basis for conclusions on the necessity to make changes to the product, which is being manufactured, as well as to the manufacturing process (Parrot & Warshaw, 2020).

The model of Haße et al. applies the Lambda architecture in the logistics industry, which ensures a scalable and powerful infrastructure for collecting, processing and visualizing data from IoT. It has four layers: data acquisition; data processing with Lambda architecture; data visualization; semantic layer for digital twins in logistics. The proposed solution is flexibly scalable. The architecture operation can be adapted depending on the type and number of sensors, complexity of analysis models and the presence of computing resources in the logistics company.

The researchers of Oracle point out that with the implementation of management systems of big data in logistics, businesses will be able to achieve greater benefits and become more flexible (Oracle, 2015). They have concluded that the kind of data used, the way in which it was extracted and structured, as well as the frequency of updates and quality of the data, are crucial in deciding which is the right technology. It must be determined whether it will be used for processing in real time or in batch mode. The frequency of processing, which needs to be done based upon the availability of data, is also important. The proposed conceptual framework permits one to operate with structured, semi-structured, and unstructured data in logistics.

In another study (Lv, et al., 2020) that focused on the design of logistics parks for steel, they suggested a data-based approach aimed at improving the operational efficiency of these types of logistics structures. They

identified the frequency and correlation between the different products and used analyses to facilitate the product distribution of steel logistics parks. The data analysis module includes three steps: data pre-processing; model building and setting of model parameters.

The aim of the module is to use the obtained results to assist the operations management in order to clarify the characteristics of incoming orders, monitor the work in real-time and achieve an optimal distribution and deployment of steel products in the logistics parks.

Another solution proposed by a team of scientists (Jiang et al., 2020), which can be utilized for data processing and monitoring, includes an architecture consisting of: perception layer, network layer, big data layer and application layer. In the first layer, one finds hardware such as mobile devices, RFID, sensors, all of which collect appropriate data for the logistics process. The network layer is based on urban networks, networks for mobile communications. The data layer uses basic technologies for virtualization and distributed data storage. The application layer is divided into modules for: client relations management, warehouse management, orders and distribution management.

The data management model for logistics systems in e-commerce are proposed by another group of scientists (Zhao et al., 2020). They researched how to improve efficiency of logistics distribution in e-commerce through big data analysis. The idea of constructing a logistics model for distribution management in e-commerce includes these stages: data acquisition; data extraction and analysis; derivation of forecasts, optimization models, etc. In relation to logistics processes in e-commerce, Zampou et al. (2018) proposed a solution. In their model, they focus on security and include special connectors, which allow a higher level of security through combining various mechanisms for secure data exchange that is only between certified and connected partners. They use platforms for storage and computing of big data, which consist of a collection of network servers with resources for storage and computing working in a cluster.

This study of the existing logistics-specific solutions for processing big data gives us grounds to define the following basic principles for the construction of a modern management architecture and big data processing in this field:

1. **Adaptability** and **flexibility**. Ability to adjust to quickly changing conditions. Allowing the quick redirection of data flow in accordance with the business' goals. Every object that uses data to be utilized several times and, in this way, to ensure a constant stream of high quality, relevant data for business. The ability to maintain various types of business users, operations and the processing of different types of data.
2. **Scalability**. Variable workloads in organizations require an elastic architecture, which adapts to the changing requirements for data processing upon demand. This type of architecture allows for work to be done even when there is a change in the capacity, i.e, the use of additional applications and analytical platforms.
3. **Automation**. The processes of absorption, preparation, processing of data sets to be implemented by mechanisms that operate with minimal human intervention. So that anomalies can be discovered in real time and then to display warnings to the instituted operational control panels.
4. **Intelligent**. Necessity of the data architecture to use intelligent resources for training, settings, and data administration. Based upon algorithms for machine training, which prepares the data, identifies errors in quantity and suggests actions.
5. **Managed**. Developed by different cooperating specialists. The process of extraction, storage, and data processing is the responsibility of IT specialists as well as business analysts, who know the business context of the operations. To clearly define the different types of users and their rights to access the data.
6. **Personalized**. Compliance with the organization and its business needs. If it is a small business for example, it could work with intelligent tools in an integrated management environment. If it is a big company, it should use a system for parallel processing.
7. **Secure**. Protected from unauthorized access and compliant with the general regulation for data protection adopted by the European Union (Europa.eu, 2016).
8. **Stable**. Any data architecture should be reliable and allow recovery upon need.

On the basis of all the presented and researched methods for data processing and analysis in logistics, as well as, the principles above, we would like to offer a conceptual framework for the management and processing of data in the logistics industry (Fig. 6).

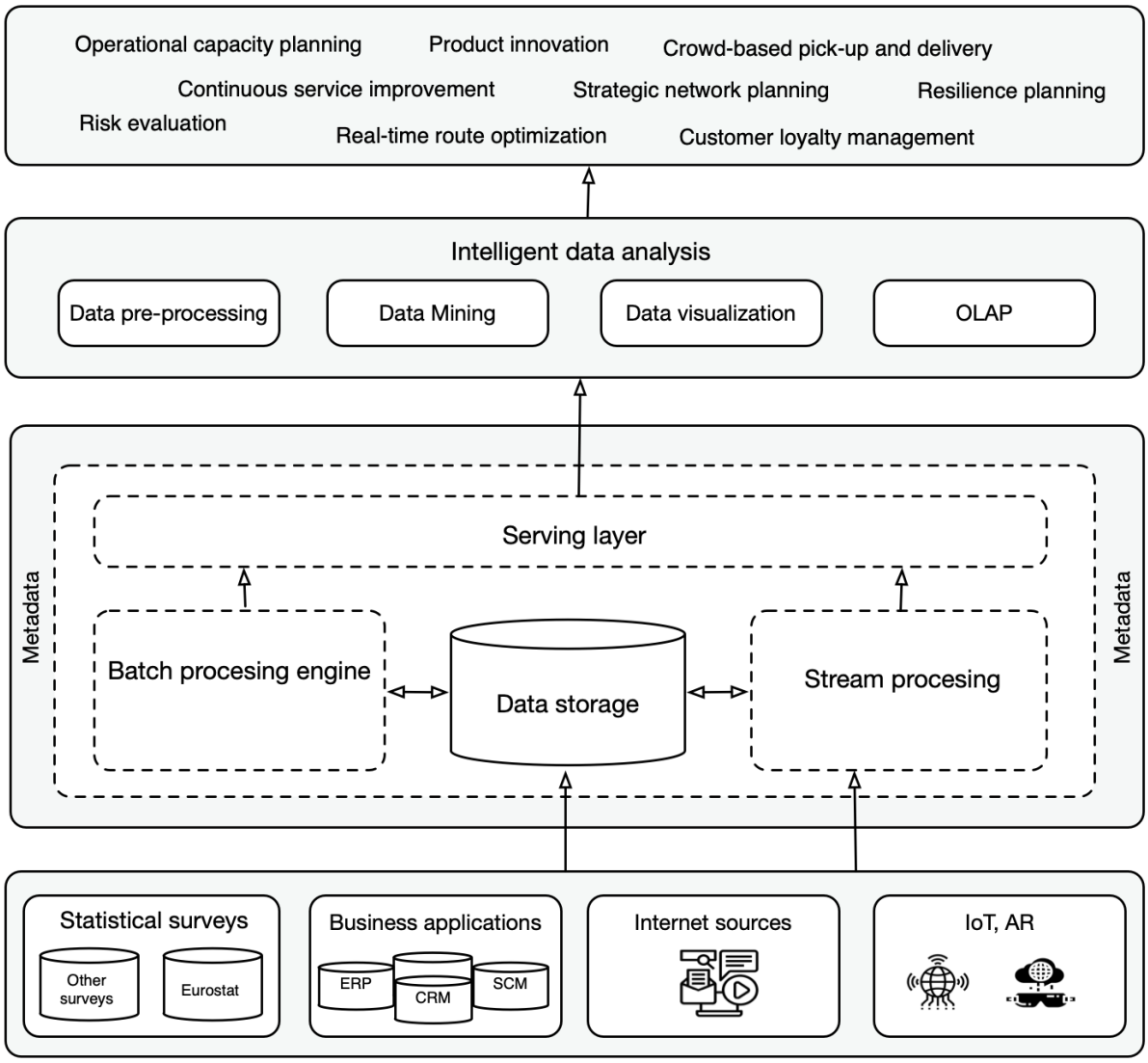


Figure 4. Conceptual framework for the management and processing of data in logistics
Source: Own elaboration

Big data processing in logistics refers to the overall process of: selection and extraction of data; storage and management of data; and the application of resources for intelligent business analyses in order to identify models, predictions, forecasts, discovery of tendencies, etc.

We have identified the following as main data sources:

- surveys;
- user-inputted data from software systems, as well as data generated from work on ERP, CRM, SCM and other systems;
- IoT devices with internet access and sensors and Augmented Reality (AR) technology;
- internet sources – server log files, webpages, social media, etc.

The data storage process should be implemented though data repositories with flexible architecture, which allows for the organization and processing of structured, semi-structured and unstructured data.

The implementation of the suggested concept can be made on the basis of various software tools for data warehouse management and business analysis. We would suggest that the implementation can be accomplished with the collection of open-source software Apache Handoop. Open-source software solutions are constantly evolving, and they allow for an easier adaptation for a specific business and the ability to make improvements. The Hadoop platform is designed for the organization of distributed processing of large volumes of data using

the map/reduce concept and for that reason, we believe that it is appropriate for the implementation of the proposed conceptual framework.

It is advisable to use the Apache Sqoop tool to transfer the data coming from the relational databases of the logistics business application in the Hadoop Distributed File System. One of its advantages is that it is easy to use and it can work with different relational database systems. Also, it was designed based on the module principle and that allows for specialized additions to be included and for optimized transfers for certain DBMSs.

Batch data processing is based on HDFS, Map-Reduce, Oozie, Pig, Hiv and Impala. HDFS distributed file system and MapReduce are the main components of Hadoop, providing parallel processing of clusters. Oozie is suitable for task planning and workflow. Pig is a platform for creating and executing tasks, it can be extended with the help of user-defined functions written in Java, Python, JavaScript and Ruby. Apache Hive is a database management system that supports SQL-based query language. Apache Impala is used for queries of data stored in the HDFS and Apache HBase.

Different sensor measurements, website and application clicks, and completed financial transactions thankfully provide an almost constant flow of data in logistics. The Apache Kafka software platform is used to receive and process real-time streaming data. It was initially designed to work with LinkedIn data, but later it transformed into a highly effective system for processing many types of streaming data. It can be used with data from websites, social networks and monitoring systems. Apache Flume is most often used for the collection and summary of streaming data, such as log files, but it must be noted that Apache Flume is not limited to working with log files only. It is used for transporting large quantities of data for events, network traffic, social media data, email messages, etc. Apache Kafka and Flume are both reliable and provide guarantees against the loss of data.

Apache Spark open-source platform can be used to achieve real-time processing and the necessary performance for distributed processing of semi-structured and unstructured streaming data. Its main advantage is speed, it can complete tasks much faster than MapReduce. Apache Spark Streaming can work with real-time streams by diverting input data streams into packets, which are processed by the Spark Engine and then it generates a final stream. It integrates well with Spark SQL and Spark MLib. It is suitable for processing data from IoT devices.

Streaming data can also be accomplished using Apache Flink – distribution mechanism for streaming data, which provides libraries for batch and streaming processing, machine learning and graphic processing. During the processing of the stream, each stream has input from one or more sources and sends the data to output streams (file or database). The data is transformed while in the stream so that the data will be suitable for real-time analyses.

In order to service the data warehouse requests, it is best to use these technologies: Cassandra, Impala, Pig Table Export, SqoopExport, Hive Tables, Impala Tables.

Intelligent analyses of big data in logistics can be performed using a large variety of data mining technologies. For this purpose, the R and Python scripts were specially designed or one can use other data mining software tools that are suitable for working with big data in the logistics industry. Some of the appropriate software tools, which the research has shown to be widely used (Landset et al., 2015) are:

- MLlib – Spark library for machine learning. It uses algorithms for data pre-processing and transformation, classification, regression, and clustering (spark.apache.org, 2021);
- Apache Mahout – framework for creating algorithms focused on linear algebra for machine learning (mahout.apache.org, 2021);
- SAMOA (Scalable Advanced Massive Online Analysis) – open-source platform for data mining from large data streams. Offering a collection of distributed streaming algorithms for the most often encountered data mining and machine learning tasks such as classification, clusterization and regression;
- H2O – open-source platform for machine learning with memory with linear scalability. It works with Handoop, which provides the ability for data importation from various sources and it has a quick, scalable and distributed computing mechanism written in Java. It supports the most commonly used statistics and algorithms for machine learning. The platforms includes interfaces for R, Python, Scala, Java, JSON and CoffeeScript/JavaScript and built-in Flow web interface;
- RapidMiner Radoop – machine training for Hadoop and Spark. It provides an easy-to-use graphic interface for analysis of data in the Hadoop cluster with the Hive working server.

The proposed conceptual framework and its implementation fully comply with the above defined principles for constructing a modern management architecture and the processing of big data in logistics.

5. Conclusion

The collection and processing of data from diverse and heterogeneous sources in the logistics industry creates problems for conventional logistics information systems. Existing software solutions are not able to process large volumes of data from sensor devices or social media in real-time. Our research proposes a conceptual framework for the processing of specific structured, semi-structured and unstructured data in logistics. The framework would provide logistics companies with many advantages, such as forecasting events, route optimization in real-time, prevention of unexpected delivery delays, implementation of innovative solutions and an overall adaptation to digitalization processes.

Literature

Alessia Napoleone, A., Macchi, M., Pozzetti, A. (2020). A review on the characteristics of cyber-physical systems for the future smart factories. Journal of Manufacturing Systems, 54, pp. 305–335.

AlShaer, M., Taher, Y., Haque, R. & Hacid, M., 2019. IBRIDIA: A hybrid solution for processing big logistics data. Future Generation Computer Systems, Том 97, pp. 792-804.

Bahga, A. and Madisetti, V. (2019). Big Data Analytics: A Hands-On Approach. Published by Arshdeep Bahga & Vijay Madisetti.

Chaopaisarn, P. and Woschank, M. (2020). Maturity Model Assessment of SMART Logistics for SMEs. Chiang Mai University Journal of Natural Sciences. 20(2), pp. 1-8.

Cio.com, 2019. AI Unleashes the Power of Unstructured Data. [Online] Available from: https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html. [Accessed 06/02/2021].

Dimitrova, V., Semova, M., 2018. Blockchain and Smart Contracts in Retailing. Commerce 4.0 – Science, Practice and Education: International Scientific Conference, pp. 126-136.

Domo.com, 2021, Data Never Sleeps 8.0, [Online] Available from: https://www.domo.com/learn/data-never-sleeps-8 [Accessed 07/02/2021].

Emilova, P., 2016. Economic aspects of the use of cloud services. (in Bulgarian). Anniversary scientific conference "Challenges to information technology in the context of Horizon, Svishtov: Tsenov Academic Publishing House, pp. 126-132.

Europa.eu, 2016. Data protection in the EU. [Online] Available from: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_bg [Accessed 07/02/2021].

Eurostat, 2021. Cloud computing – statistics on the use by enterprises. [Online] Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php/Cloud_computing_statistics_on_the_use_by_enterprises. [Accessed 04/02/2021].

Filipova, N., 2018. Blockchain – an opportunity for developing new business models. Business management, SA „D. A Tsenov“ Svishtov, 2, pp. 76-92.

Firican, G., 2017. The 10 vs of big data, upside where data means business. [Online] Available from: https://upside.tdwi.org/Articles/2017/ 02/08/10-Vs-of-Big-Data.aspx?Page=1. [Accessed 06/02/2021].

fortunebusinessinsights.com 2020. Market Research report. [Online] Available from: https://www.fortunebusinessinsights.com/industry-reports/big-data-technology-market-100144 [Accessed 07/02/2021].

Gergova, I. et al. 2017. Possibilities of GIS for optimization of localization processes in tourism. Management and sustainable development, 4 (65), [Online] Available from: http://oldweb.ltu.bg/jmsd/files/articles/65/65-05_I_Gergova_Minov_Momchev_Kovacheva_Ivanov.pdf

Ghobakhloo, M., 2020. Industry 4.0, digitization, and opportunities for sustainability. Journal of Cleaner Production 252, 119869, [Accessed 07/02/2021].

Haße, H. et al., 2019. Digital twin for real-time data processing in logistics. Chapters from the Proceedings of the Hamburg International Conference of Logistics (HICL), Institut für Logistik und Unternehmensführung, Technische Universität Hamburg, Том 27, pp. 4-28.

Kirov, S., 2020. Blockchain applications in the insurance industry. (in Bulgarian). Jubilee International Scientific Conference “Economic science, education and the real economy: development and interactions in the digital age” T. 1, Varna: Publishing house "Science and Economy", pp. 186-200.

Kondratiev, V., 2019. The fourth industrial revolution and globalization. (in Bulgarian) [Online] Available at: https://geopolitica.eu/2018/174-broy-6-2018/2941-chetvartata-industrialna-revolyutsiya-i-globalizatsiyata. [Accessed 3.02.2021].

Kotu, V. and Deshpande, B. (2019). Data Science. Concepts and Practice Second Edition. Morgan Kaufmann Publishers an imprint of Elsevier.

Lazarova, V., 2019. Digitalization and Digital Transformation in Accounting. Ikonomiceski i Sotsialni

Alternativi, Sofia: University of National and World Economy, issue 2, pp. 97-106.

Lv, Y., Xiang, S., Zhu, T. & Zhang, S., 2020. Data-Driven Design and Optimization for Smart Logistics Parks: Towards the Sustainable Development of the Steel Industry. MDPI, Open Access Journal, 12(17), pp. 1-13.

Lydon, B., 2014. Industry 4.0 - Only One-Tenth of Germany's High-Tech Strategy. [Online] Available at: https://www.automation.com/en-us/articles/2014-1/industry-40-only-one-tenth-of-germanys-high-tech-s. [Accessed 3.02.2021].

Mazumdar, S., et al., 2019. A survey on data storage and placement methodologies for Cloud-Big Data ecosystem. Journal of Big Data, Vol. 6, pp. 1-37.

mi.government.bg, 2017. Concept for digital transformation of the Bulgarian industry (Industry 4.0). (in Bulgarian) [Online] Available at: https://mi.government.bg/bg/themes/koncepciya-za-cifrova-transformaciya-na-balgarskata-industriya-industriya-4-0-1862-468.html [Accessed 3.02.2021].

Oracle, 2015. Improving Logistics & Transportation Performance with Big Data. [Online] Available at: https://wheels.report/Resources/ Whitepapers/9b500313-465f-48cb-90d1-719404791108_C.pdf [Accessed 04/03/2021].

Paksoy, T. Koçhan, Ç. Ali, S., 2021. Logistics 4.0. Digital Transformation of Supply Chain Management. CRC Press Taylor & Francis Group.

Parrott, A. & Warshaw, L., 2020. Industry 4.0 and the digital twin technology. [Online] Available at: https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html [Accessed 04/03/2021].

Patel, P., Sheth, A. & Ali, M., 2018. From Raw Data to Smart Manufacturing: AI and Semantic Web of Things for Industry 4.0. Intelligent Systems, IEEE, 33(4), pp. 79-86.

Petrov, Ph. and Trifonova, Ts., 2019. Non-relational databases – a practical guide. Sofia: University Publishing House "St. Kliment Ohridski ”.

Proctor, M. and Wilkins, J., 2019. 4.0 Sight—digital industry around the world. Technical report, EU Automation. [Online] Available at: http://www.4sightbook. com/. Accessed Feb 2020 [Accessed 04/02/2021].

pwc.de, 2017. Digital Factories 2020: Shaping the future of manufacturing. [Online] Available from: https://www.pwc.de/de/digitale-transformation/digital-factories-2020-shaping-the-future-of-manufacturing.pdf [Accessed 03/02/2021].

Rojko, A., 2017. Industry 4.0 concept: background and overview. International Journal of Interactive Mobile Technologies. (iJIM), 11(5): pp. 77-90.

Satatista.com, 2021. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024. [Online] Available at: https://www.statista.com/statistics/871513/worldwide-data-created/, [Accessed 3.02.2021].

Shafer. T., 2017. The 42 V’s of Big Data and Data Science. [Online] Available from: https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html [Accessed 04/02/2021].

Sharda, R. Delen, D. Turban, E., 2020. Analytics, Data Science, & Artificial Intelligence Systems for Decision Support. Pearson Education.

Sharda, R., Delen, D., Turban, E., 2020. Analytics, data science, & artificial intelligence systems for decision support. Pearson.

Statista.com, 2021. Number of Internet of Things (IoT) connected devices worldwide in 2018, 2015 and 2030. [Online] Available from: https://www.statista.com/statistics/802690/worldwide-connected-devices-by-access-technology/ [Accessed 04/02/2021].

Storeya, V. and Song, I., 2017. Big data technologies and Management: What conceptual modeling can do. Data & Knowledge Engineering, 108, pp. 50-67.

Stoyanov, S. and Popchev, I., 2017. Internet of Things. (in Bulgarian) [Online] Available at: https://www.researchgate.net/profile/Ivan_Popchev/publication/322953157_Internet_na_nesata/links/5a79909445851541ce5ce863/Internet-na-nesata.pdf [Accessed 3.02.2021].

Stoyanova, M., Vasilev, J., Cristescu, M. 2021. Big Data in Property Management. Applications of Mathematics in Engineering and Economics: Proceedings of the 46th Conference on Applications of Mathematics in Engineering and Economics (AMEE '20), AIP Conference Proceedings 2333, 070001.

Turchi, P., 2018. The Digital Transformation Pyramid: A Business-driven Approach for Corporate Initiatives. [Online] Available from: https://www.linkedin.com/pulse/digital-transformation-pyramid-business-driven-approach-turchi [Accessed 03/02/2021].

Vasilev, J., 2017. E-logistics in the context of globalization. (in Bulgarian) Varna: Publishing house "Science and Economy".

Zampou, E. et al., 2018. Big data analytics in e-commerce logistics: Findings from a systematic review and a

case study. Proceedings of 7th Transport Research Arena TRA 2018, Vienna, Austria, [Online] Available at: https://zenodo.org/record/1491581#.X6fU5y8RoXo. [Accessed 8.11.2020].

Zhang, D. et al., 2021. A survey on attack detection, estimation and control of industrial cyber-physical systems, ISA Transactions. [Online] Available from: https://www.sciencedirect.com/science/article/pii/S001905782100046X [Accessed 04/02/2021].

Zhao, Y. et al., 2020. Innovation Mode and Optimization Strategy of B2C E-Commerce Logistics Distribution under Big Data. Sustainability, 12, 3381, pp. 1-13.

Prototype model for big data predictive analysis in logistics area with Apache Kudu

Liliya MILEVA[1], Pavel PETROV[2], Plamen YANKOV[3], Julian VASILEV[4], Stefka PETROVA[5]

[1] University of Economics, Varna, Bulgaria
l.mileva@ue-varna.bg

[2] University of Economics, Varna, Bulgaria
petrov@ue-varna.bg

[3] University of Economics, Varna, Bulgaria
yankov.plamen@ue-varna.bg

[4] University of Economics, Varna, Bulgaria
vasilev@ue-varna.bg

[5] University of Economics, Varna, Bulgaria
s.petrova@ue-varna.bg

Abstract. Logistics is area, which is evolving rapidly, generating a lot of data lately. There are several problems that, even when partially digitized, require additional work. At the same time there is a need of big data analyses. These analyses are represented by machine learning and statistical analyses. One important problem is these with missing data for delivery in the whole process of transportation, other is about empty freight transportation. The purpose of this paper is to present an ICT prototype model with analytical tool Apache Kudu in logistics area, which will contribute to resolve such problems in area. Content analysis and systematic approach are used. Statistical methods and statistical data are used. Growth rates of empty road transport are calculated. The presented data are for Bulgaria and its neighbour’s county, excluding Romania, because of the missing data. Croatia is also included in the analysis. Growth rate analysis indicates problems with empty fright transport in some countries. The tendency for Bulgaria is to reduce empty road freight transport, even though it is better for logistics organizations to work for optimizing the process of delivery.

Key words: logistics, machine learning, big data, big data analyses, predictive analysis, Apache Kudu, Apache Hadoop, logistics software, growth rates;

1. Introduction and literature review

Acknowledgments. This work is supported by the "Digitization of the Economy in an Environment of Big Data" Project - BG05M2OP001-1.002-0002.

Organizations worldwide, use a variety of methods, models and techniques for data analysis (Calzon, 2021). They use them to create and restructure their business models as well as to avoid certain problem areas in their storage, incompatibility due to diversity, visualization problems, lack of tracking and analysis in real time, which in turn lead to large financial losses and other issues.

At present, more and more big data and their problems are entering science (Kuyumdzhiev, 2020). Previous years the scope of the term was the focus of many articles. Methods for storing and organizing big data are actually developing at present. Sometimes there is a line between traditional datasets and modern big data. A comparatively big dataset may become a similarly small “big data” (Todoranova and Penchev, 2020). The border between big datasets, data warehouse and big data are obvious (Todoranova *et al.*, 2020) (Nacheva *et al.*, 2019).

In business, processes generate amounts of data, which are normally stored in appropriate databases (Nacheva *et al.*, 2019), but even though sometimes databases could grow larger and larger within duration of

several minutes (Nacheva and Sulova, 2020, 2021). These are specific situations, in which companies are migrating from databases to “big data”. Measuring efficiency and hardware storage (Polkowski *et al.*, 2020) uses new opportunities to continuously increase the amount of data. The increased amount of data opens new perspectives for hardware storage (Polkowski *et al.*, 2020) and measuring the disk storage performance (Cristescu, 2019) (Stoyanova, 2020). Digitizing more and more business processes means that data are increasing (Miryanov and Petkov, 2017) (Nikolaev, Milkova and Miryanov, 2018). More energy is needed for their storage and cooling. Sometimes data centres are used. The question concerning the choice “own data centre” or “using a 3PL” is open.

Storing large amount of data needs the application of mathematical methods for creating forecasts with big data (Miryanov and Petkov, 2017) Retrieving data has many approaches (Medvedev and Sergeev, 2020). Creating high reliability forecasts (Miryanov and Petkov, 2017) means accepting, testing and validating all assumptions for the chosen mathematical models (Ana-Maria Ramona, Marian Pompiliu and Stoyanova, 2020) (Polkowski *et al.*, 2020). Sometimes for analyzing big data are used heuristic approaches (Ileanu et al., 2019) (Abdel-Badeeh M. Salem, 2018). Providing web access to big data is another challenge with multiple technological solutions.

At the same time, logistics, which is evolving rapidly and generating a lot of data, needs further digitalization of many business processes. However, there are a number of problems that, even when partially digitized, require additional work. Many of them stem from a lack of control (Kirova, 2009) and other issues. Many of the existing problems are typical for the combined transport or in the transport of goods from point A to point B, when the logistics is performed between different countries and by more than one logistics organization. Some of the problems start from poor or lack of communication between companies, as well as reluctance or inability to cooperate. Some of them could be solved with the help of Apache Kudu.

2. Creating methods for analysis of different types of data with the analytical tool Apache Kudu

Big data are evolving as a product that is growing rapidly, but at the same time make a connection between many businesses and other economic units. In this way, new forms of data extraction are emerging. Traditional data sets are also interpreted through the most widely used big data analysis. For the purposes of the study, they are considered articles that describe the methods of big data analysis. In the scientific literature, statistical analysis are known as well as analysis of social media (Baesens, 2019) stating that in the future many benefits are expected from their analysis. By social networks, it is meant not only social media sites such as Facebook, Twitter and LinkedIn, but also any network between customers or companies connected in a certain way (Baesens, 2017). It is made an overview of the considered methods of analysis based on the division into quantitative and qualitative (Fig. 1). Such a separation is present in each of the areas of the economy that involves analysis. The main characteristics of the methods shown in the figure 1 are quantitative analysis with a focus on a large part of the object of study and qualitative analysis - when the data are qualitative, non-numerical specialists turn to qualitative analysis by assigning (coding) numbers. The focus of this analysis is on a smaller part of the studied object.

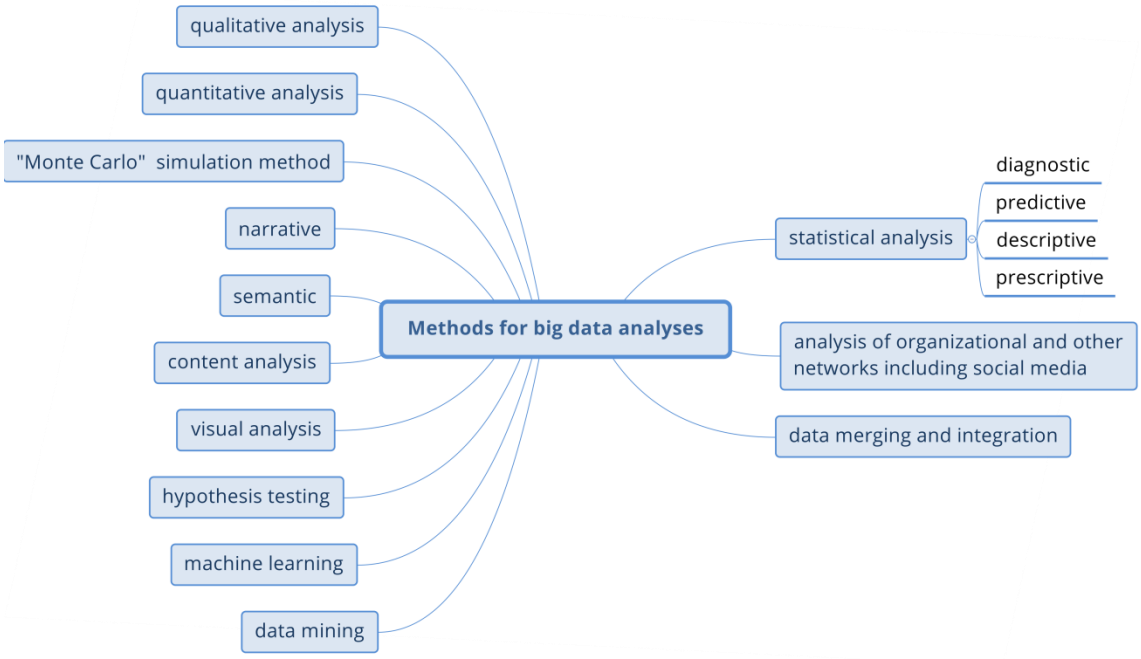


Figure 1 Methods for big data analysis

Source: The figure is based on a Dr. L. Mileva research on the existing methods in the scientific literature for big data analysis.

The paper presents the possible statistical analyses, which are of several types. 1) Descriptive which analyse the initial data. 2) Diagnostic analysis- establishes the reasons that have caused the changes in the studied phenomenon over a period. It mainly models the development trend and calculates indices of seasonal fluctuations. 3) Predictive analysis- by extrapolation of previously established development trends, different forecasts are developed. The analysis works by developing hypotheses about the expected future behavior of the development of an organization or part of it (T-testing). 4) Prescriptive analysis - a combination of descriptive and predictive, often preferred method, but it also has serious drawbacks, namely: although it is associated with, finding the best course of action for a situation usually requires a lot of labour and large financial costs. 5) Analysis of the connections between the phenomena - regression and correlation analysis, in which it is examined whether there is a connection between certain phenomena, what is the strength of the connection, etc. For a more comprehensive study, specialists focus on performing a combination of statistical or other analysis, because it will get more objectivity and higher quality of the desired result.

Statistical analyses are followed by other types that are widely used in theory and practice: 1) Machine learning, which works with computer algorithms creating assumptions about the studied data in the database. It makes predictions that cannot be made by humans because the amounts of information are too large. This makes machine learning one of the favourites for analysing large data sets. 2) Semantic analysis- collects the semantic information needed for the study from the source code. It is specific and used by information technology specialists. 3) Visual analysis - decision making based on imaged data. It is usually done, so that data users can get a better idea of the prospects for future work, such as rising and falling costs, sales revenue and more. 4) Narrative analysis, which implementation carries out through a story is exploring ideas, attitudes, opinions and stories, reveal the main preferences of the employees of an organization. 5) Content analysis, analysis of texts and performed by extracting data from the content of various formats such as video, text or other extracted from the press, social networks and others. 6) Data mining is similar to content analysis, but with the difference, that it is a search and discovery of useful and valuable information contained in large databases or data warehouses. It combines methods from statistics and machine learning, within the database management. 7) Monte Carlo - a series of computational algorithms that rely on the repetition of random phenomena to achieve numerical results, used for analysis mostly by organizations that want to minimize various risks. 8) Data merging and integration - by combining a set of techniques for integrating and analysing data from multiple sources, better and more accurate predictions are made. The presented list of studied analysis is not exhaustive, as researchers are constantly looking for new ways of analysis and working on new models or combinations of them. The decision usage specific method of analysis depends on the field of work, specific characteristics and purpose of the results

that specialists want to achieve. A combination of methods should achieve greater objectivity of the survey, higher quality of the results sought and higher added value for users of the results of surveys conducted using methods of big data analysis.

3. Big data analysis using machine learning in the Hadoop ecosystem

Data analysis is a process that relies on methods and techniques to extract information from raw data that are important for achieving business goals. They are represented by statistical methods and machine learning methods. Although statistical methods are derived as a separate branch, they are also contained as part of machine learning. In practice, both types of methods are used in almost all areas of economy and are therefore of great importance and can be used for different purposes, such as comparing different data, end results of different activities, comparing values, studying the presence or absence of mutual connections between individual phenomena. The need for in-depth consideration of machine learning models is due to its growing importance and use in the analysis of large data sets. Machine learning is becoming more popular due to the growing volume of data, which is so diverse that it requires different methods of processing and storage. In addition, we are looking for technology with lower costs and greater processing and storage capacity, something that machine learning provides as opportunities. There are also opportunities for making models that are characterized by the speed of their creation and automation options, while they have the ability to analyse larger, more complex data and accordingly to provide faster, more accurate results- even on a very large scale. Through the usage of machine learning, organizations have the opportunity to build models that help to make optimal decisions and avoid risks. One branch of data analysis is machine learning (ML), which is divided into three parts. These are: classification, clustering and evolutionary algorithms (Raeesi Vanani and Majidian, 2020). After analysing the most popular definitions of Machine learning in science, they are extracted the main features of machine learning (Fig. 2).

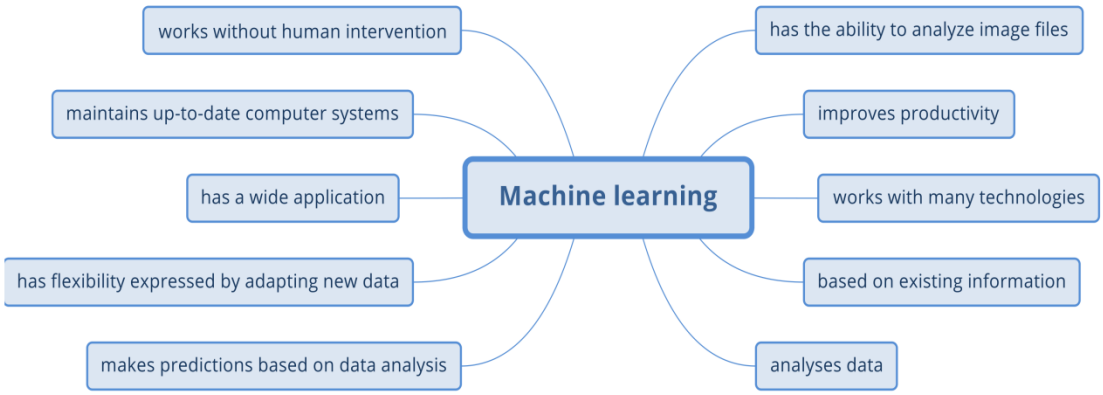


Figure 2 Characteristics of machine learning

Source: The figure is based on a Dr. L. Mileva research on the existing methods in the scientific literature for big data analysis.

3.1 General application of machine learning

Machine learning occupies a central place in the analysis of big data and is widely used in many areas of the economics and its most common usage is in the collection and analysis of data from video surveillance and information collection from social networks, but this is far from use up its application in practice. Machine learning is used to:

- image recognition (most often) video surveillance, customs clearance and airport security check; - in social networks at:
- friends suggestion;
- common acquaintances with other people;
- image recognition and tagging (image tagging on photos);
- markings of location, site, etc.;

- vehicles with a certain autonomy, such as TESLA cars and Google cars;
- personal assistants- Siri, Alexa, etc.- virtual assistants who help mobile phone users to search for objects on the Internet, save data, make schedules, mark and remind important events and more. Real-time translations have recently become especially popular, when the user does not know a certain foreign language and uses the help of a personal virtual assistant;
- analyses of consumer opinion, through surveys- for example, made after buying a product or using a service, after visiting a site, etc.
- whether they are satisfied with the purchase, the service, whether they would use it again, at what price, whether it costs them is that it is too expensive, whether they think there is a need for certain changes and what;
- forecasts based on consumer behavior, for example, a preliminary survey of consumers' opinions on whether they would invest in a new service, previously unknown or product, thus the organization will gain information on whether it is worth investing in a particular product or service or not, should it be soon consider introducing a product and/or service of another kind in order to overcome unforeseen losses and risks;
- analyses of consumer behavior, by collecting input data (usually include information on the number of purchases per month, amount spent, preference for a particular model of shoes, for example - sporty, elegant, etc.);
- real-time traffic forecasts - where passengers will avoid traffic jams and save travel time;
- fraud detection with bank cards and transactions;
- filtering spam and malware messages in e-mail;
- recommendations for visiting tourist sites such as accommodation and meals;
- recommendations for visiting historical and cultural landmarks;
- data collection for project implementation- collection of data for tracking the employees’ behavior within the working day- tracks the time of arrival and departure from the workplace, compliance or non-compliance with the time for breaks, even tracking the activity within the working day;
- collection of data for tracking the behavior of employees within a project- for example, movement of employees, performance of tasks, reporting results, etc.
- video surveillance of residential and construction sites- for security and protection of the site;
- tracking of logistic processes- for tracking of personal and company shipments, logistic processes in sites or from point A to point B during their movement, which helps for undamaged and timely deliveries in real time;

For purposes of the present article, they are derived the applications of Machine learning in logistics area.

Application of machine learning in logistics:

- tracking of logistics processes - for tracking of personal and company shipments, logistics processes in sites or from point A to point B, during their movement, which helps for undamaged and timely deliveries in real time;
- working through a supply chain;
- collecting data to track employee behavior within a project- for example, employee movement, task performance, reporting results, etc. The Hadoop ecosystem handles very large data sets, which requires various tools to retrieve, process, store and distribute them. It also performs multiple analyses using machine learning. In its development, the known basic species complement, modify and compose new subspecies. Each type of machine learning could apply to a different area, goal or task. It is characteristic of Machine learning that depending on the purpose of the work there are opportunities to mix two or more types of methods and techniques. In FIG. 3 and 4 are presented the types of ML, the variables with which they work in the analysis of big data, methods of work, as well as the tools for working in Hadoop.

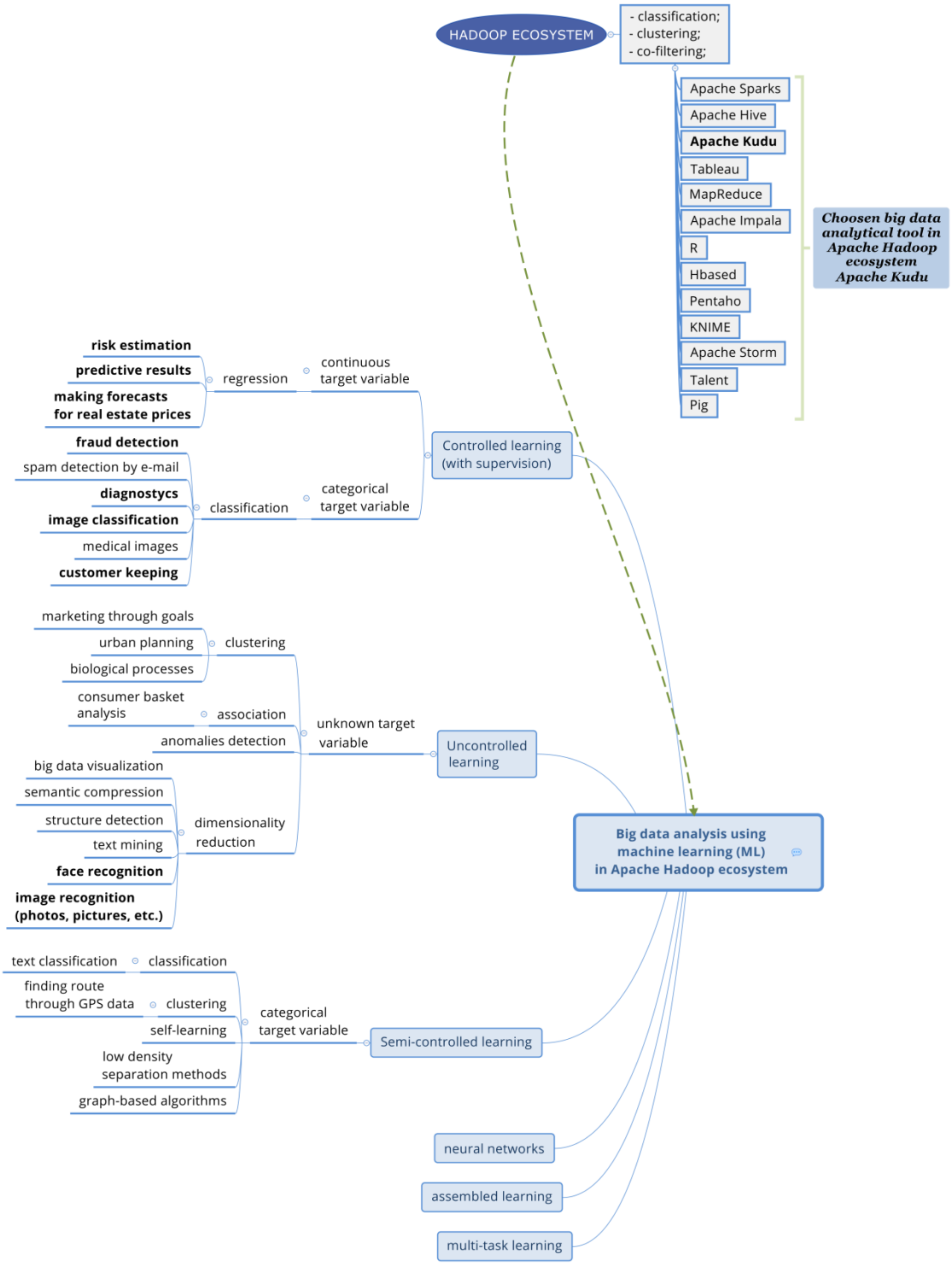


Figure 3 Big Data Tools

Source: The figure is based on a Dr. L. Mileva's research on existing in the scientific literature methods for big data analysis.

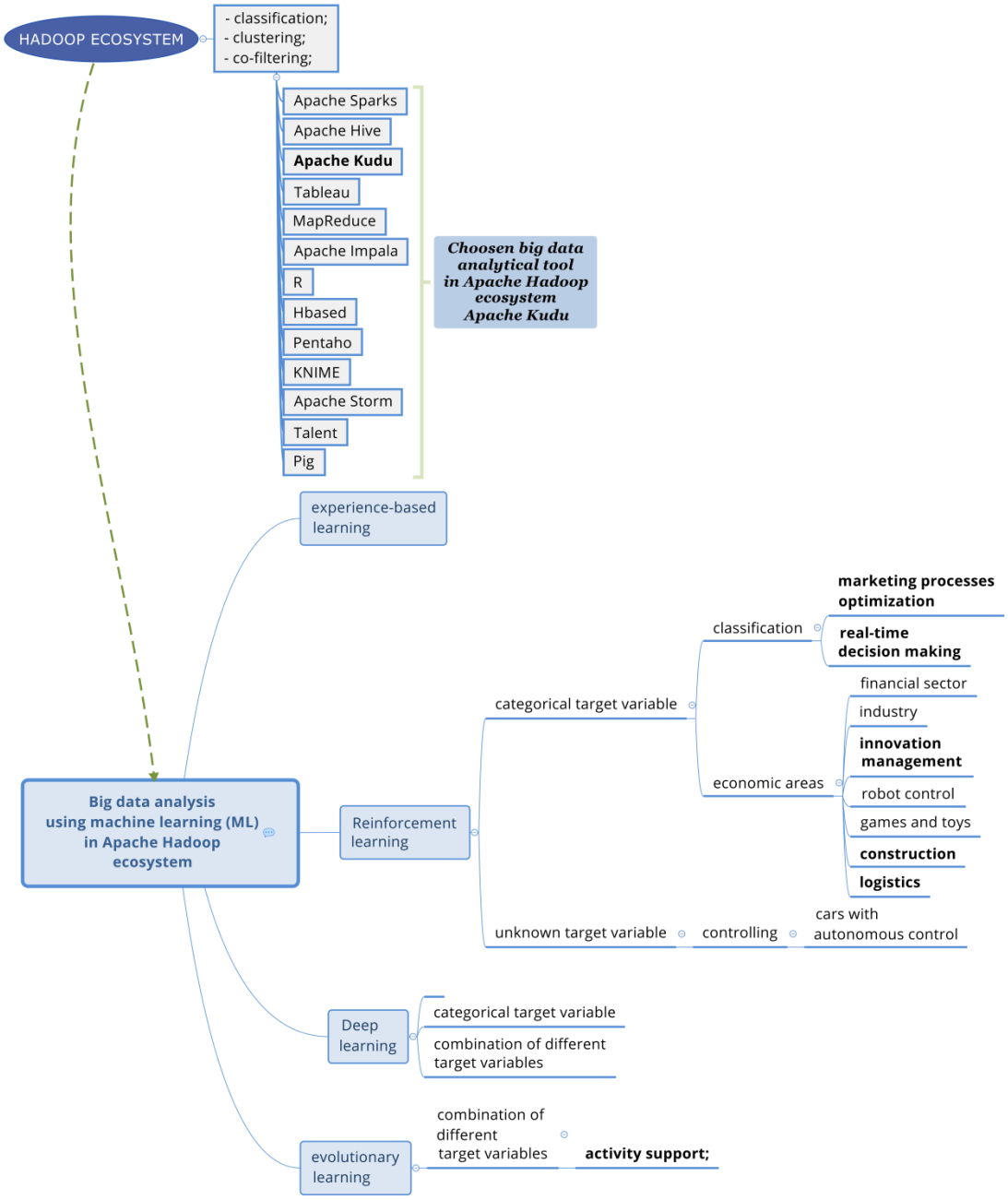


Figure 4 Big Data Tools
Source: The figure is based on a Dr. L. Mileva's research on existing in the scientific literature methods for big data analysis.

In Fig. 3 and 4 (Heidenreich, 2018) it is also present the tools for working with big data in Hadoop, through the three principles of operation: clustering, classification and co-filtering. Unlike data analysis methods, which aim are the data itself, their aggregation into data clusters and classification, the principles of the tools are based on aggregation, classification and filtering, but no longer on the data itself, but on the computer systems working with them. With Hadoop's big data analysis tools, researchers have the opportunity to design models and use specific methods to make various business decisions not only in the fields of logistics, but also in almost all other economic fields including construction. The analytical tools are presented in fig. 1 and 2:

1. Apache Sparks.
2. Apache Hive.
3. Tableau.
4. Map Reduce.
5. Apache Impala.
6. R.
7. Hbased.
8. Pentaho.
9. KNIME.
10. Apache Storm.
11. Talent.
12. Pig.

This list does not limit known analytical tools, but it is the most applicable and used in practice. Among this tools is the Apache Kudu (*Apache Kudu - Overview*, 2017) , which is characterized by the most advantages among the others and was selected for analysis in Hadoop ecosystem. Apache Kudu is a free, open source tool focused on storing large amounts of data in the Hadoop ecosystem. Its main advantage is that it is compatible with most data processing frameworks in Hadoop. It is also characterized by the speed of data transfer during storage and transfer. The advantages of Apache Kudu are base for creating a model for logistics work in making specific decisions.

3.2 Application of the Apache Kudu analytical tool in the field of logistics

Apache Kudu is a column storage manager developed for the Hadoop platform. Kudu shares the general technical features of Hadoop's ecosystem applications, running on stock hardware, being horizontally scalable, and supporting highly accessible operations.

Advantages. Kudu is characterized by fast load processing, full integration with Apache Impala, as well as high performance, so it is one of the most preferred analysis tools in Hadoop. The advantages of Kudu include:

- fast processing of OLAP loads (online analytical processing);
- Developed to integrate with Map Reduce, Spark, Flume and other components of the Hadoop ecosystem.
- Full integration with Apache Impala, making it a good, changeable alternative to using HDFS with Apache Parquet.
- strictly defined but nevertheless flexible sequence model that allows on-demand compliance requirements to be selected, including a strict sequence option.
- high productivity at simultaneous execution of consecutive and arbitrary loadings;
- easy to administer and manage via Cloudera Manager;
- high availability. Tablet servers and Masters use the Raft consensus algorithm, which provides availability as long as more replicas are available than not. Readings can be serviced by consecutive read-only tablets, even in the event of a leading tablet failure. Kudu works under certain limitations, and professionals should comply with them.

Apache Kudu could be applied properly in the field of logistics because of its ability to use appropriate solutions (Quinto, 2018). Possible logistics solutions include:

- reporting applications where the new data must be immediately available to end-users; - timeline applications that must support requests for large amounts of historical data while returning detailed requests for an individual site;
- applications that use real-time predictive decision-making models, with periodic updates of the predictive model based on all historical data; These applications are typical in the field of logistics, where it is necessary to track the movement of goods and goods from one point to another to make it possible throughout the process to track goods continuously in real time. For the purposes of the development, they have been identified some well-known problems in logistics, which would be solved with the help of analytical tool Apache Kudu. The advantages of Apache Kudu will be the base for building a model for work in construction and logistics in making specific decisions. Logistics is an area in which digital transformation is present largely.

There are a number of problems that, even when partially digitized, which despite the efforts require additional work. Many of the problems came from a lack of control (Kirova, 2009) lack of flexibility in work, and some are caused by human error. Many of the existing discrepancies are typical for the combined transport (Parvanov and Boeva, 2009) or in the transport of goods from point A to point B, when the logistics is performed

between different countries and by more than one logistics organization. Some of the problems came from poor or lack of communication between companies, as well as reluctance or inability to cooperate (sometimes due to lack of sufficient financial and human resources and information, as well as poor management). Some of them could be solved with help of Apache Kudu.

3.2.1 Some logistics problems

Some known most common issues are listed below:

- fraud related to theft of customer data;
- fraud involving theft of goods, cargo, fuel (sometimes theft is committed by outsiders, but it is also common for staff members to abuse it to profit from the sale of goods to third parties at higher prices and others.);
- problems with insurance- when there are many damaged and broken shipments, their insurance covers the value, but when there are many such events, the insurer begins to lose from them instead of winning. From there are raised the contributions because the organization becomes a risk client and in turn raises the fees for the clients in order to close the difference for the insurances. Customers become dissatisfied and start using other companies that have lower fees and better service, in which the organization loses market share. Customers are dissatisfied, as their shipments cannot arrive in time. If they are from a store, and if they are their property, they may not recover them at all (in case they are single items). For example, if a picture replicates second time it loses its true value.
- poor infrastructure of some countries on the European road network - incl. Bulgaria, Romania;
  - out-dated fleet of some organizations- leads to delays, technical problems of vehicles, the need for frequent technical maintenance, respectively, and specialists to perform it; - high unemployment;
- unqualified staff; On the one hand, employees complain about lack of work, poor working conditions, and on the other hand, employers complain about the lack of qualified staff (e.g. drivers-suppliers).
- problems with cargo storage;
- inability of the end customer to monitor the movement of his shipment;
- difficult, even impossible traceability of consignments from item A to item B.,
- due to:
- lack of sufficient information - blank data,
- lack of historical data for customers;
- lack of data caused by errors when entering them during registration (as incomplete: address, telephone number, names, etc.), which complicates the process of sending the shipment and requires additional work (call to contact the customer);
- problems with transliteration of the client's names when using more than one alphabet (for example Latin and Cyrillic);
- passing through hard-to-reach terrains;
- crossing limited routes (in connection with hostilities, disasters, accidents, crisis situations (such as the Covid-19 pandemic, the attacks in Turkey, etc.);
- passing through dangerous areas (e.g. areas inhabited by people from problem minority groups);
- remote points along the route;
- Lack of communication IT infrastructure, including scope throughout the supply chain;
  - lack of control;
  - lack of flexibility;
  - unreceived shipments;
  - unsolicited shipments;

The listed problems also require possible solutions to improve the transport and logistics of goods. This can be done by digitizing the processes in logistics organizations.

Suggestions for solutions are:

- better control;
- periodic release of control consignments that require special attention (easily damaged, fragile);
- better flexibility;
- traceability of shipments, cargo, goods in real time;
- real-time vehicle traceability;

- traceability of the personnel performing the transport activities (when the personnel is part of the human resources of the company for the determined period of time);
- predictability of various factors that could cause delays- for example, lack of storage for cargo, delays in transport, staff errors;
- avoiding busy and dangerous routes with the help of a notification system;
- finding unmarked GPS points along the route;
- tracking of consignments, when transported by two or more logistics companies, when the goods are passed from point A to point B, passing through different countries; Highlights:
- tracking staff to avoid fraud;
- cargo tracking;
- better control over shipments;
- better control of staff;
- greater flexibility;

For purposes of the article, was chosen a specific problem, which is largely present in the work of logistics organizations and the solution of which will lead to great benefits for future analysis of big data from the point of view not only of logistics companies but also customers. The main problem observed in the field of logistics is the impossibility of monitoring the transport of goods from more than one logistics organization as well as the transport from one country to another or passing through several foreign territories. Information about it is often lost when it is delivered, which makes it impossible for both the internal system and the end user to track it. This leads to other problems such as:

- inability of the customer to monitor the movement of his shipment;
- inability of the customer to receive notification of his shipment on time;

The inability to receive the shipment on time by the customer leads to:

- delay in receipt;
- charging additional fees for staying in a warehouse for more than three days;
- a fine for the office worker or courier who has failed to deliver the goods within three or more days (depending on the policy of the logistics organization); All this affects both the quality of service and customer satisfaction. It can lead to customer withdrawal, financial losses, and even the organization dropping out of the market. With regard to big data analysis, the current problem leads also to:
- inability to trace;
- loss of information when passing from one courier to another, when passing from one country to another;
- inability to perform analyses with regard to: - tracking the direction of goods and cargo;
- analyses based on time travelled;
- analyses of the movement of goods and cargoes from one country to another;

It can be concluded that all these missed opportunities stop working and make analysis impossible, through big data for business and other purposes. This disrupts work, slows down plans to attract new customers, expand the business and even vice versa can even lead to a contraction of business or dropping out of the market due to losses (Fig. 5).

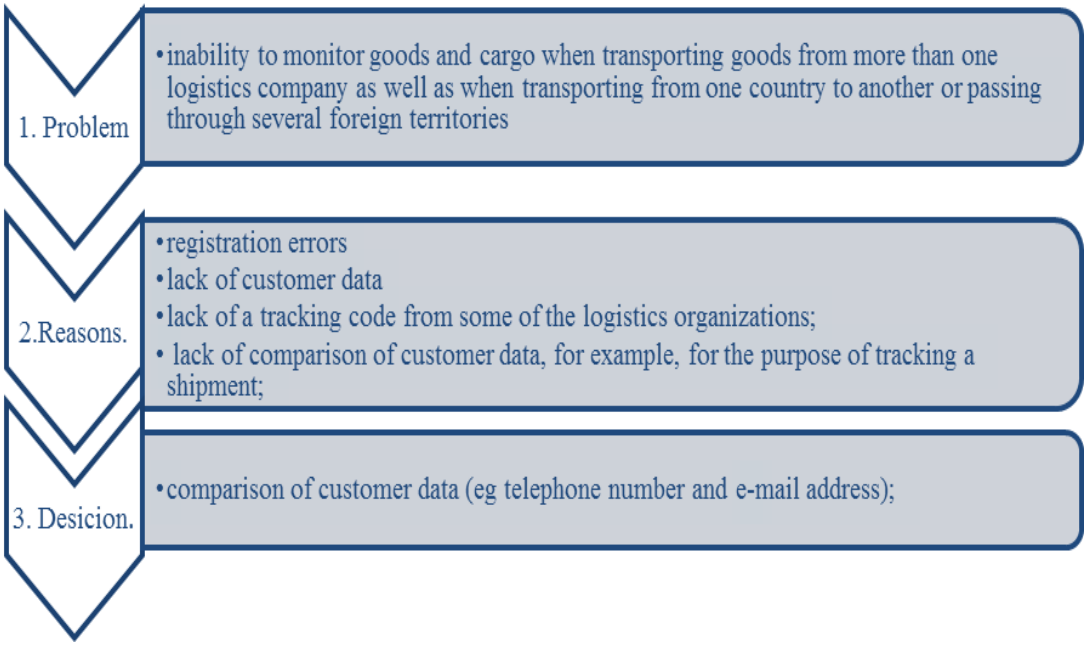


Figure 5 Decision making process in the field of logistics
Source: based on authors' research

Solving this problem will achieve many benefits and opportunities in big data analysis using the Apache Kudu tool. With its help, the relevant data first can be stored and then analysed. For this purpose, it is possible to use part of the logistics information system to solve the problem. The whole process of transferring items from one point to another goes through certain stages. The process includes not only the actual physical transport, but also the acceptance of a customer request, data processing, stay in a distribution centre (hub) and delivery. The stages can be systematized as follows (Fig. 6).

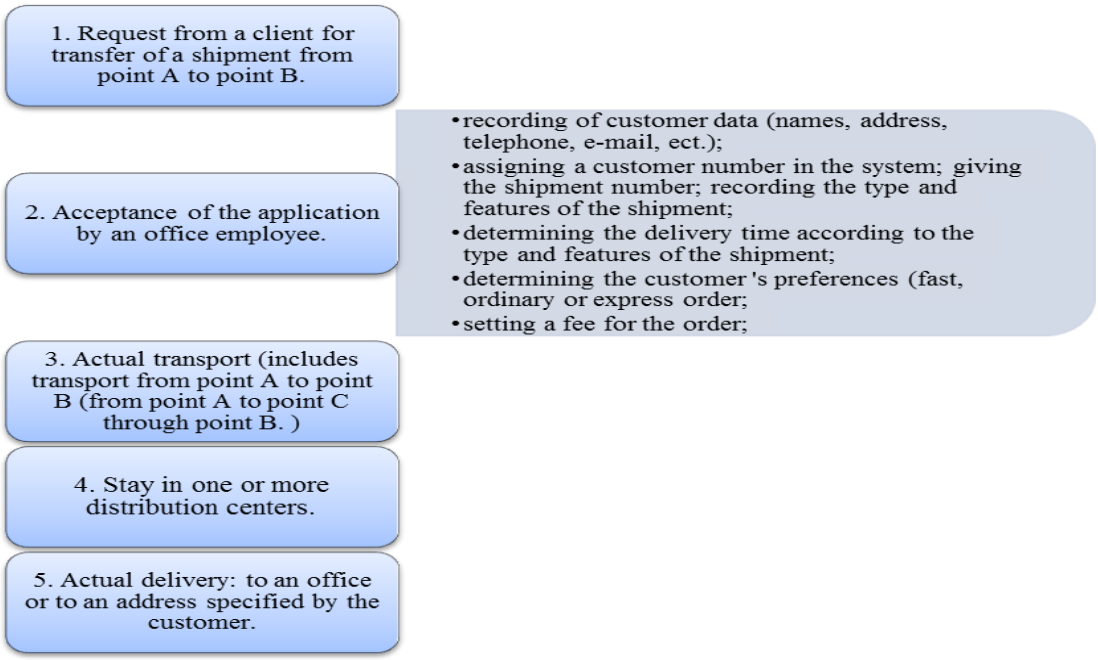


Figure 6 Shipment delivery process.
Source: based on authors' research

In order to avoid problems in tracking and to provide an opportunity for continuous monitoring, additional codes should be used in the internal system of the logistics organization to allow constant monitoring, tracking shipments (Fig. 7).

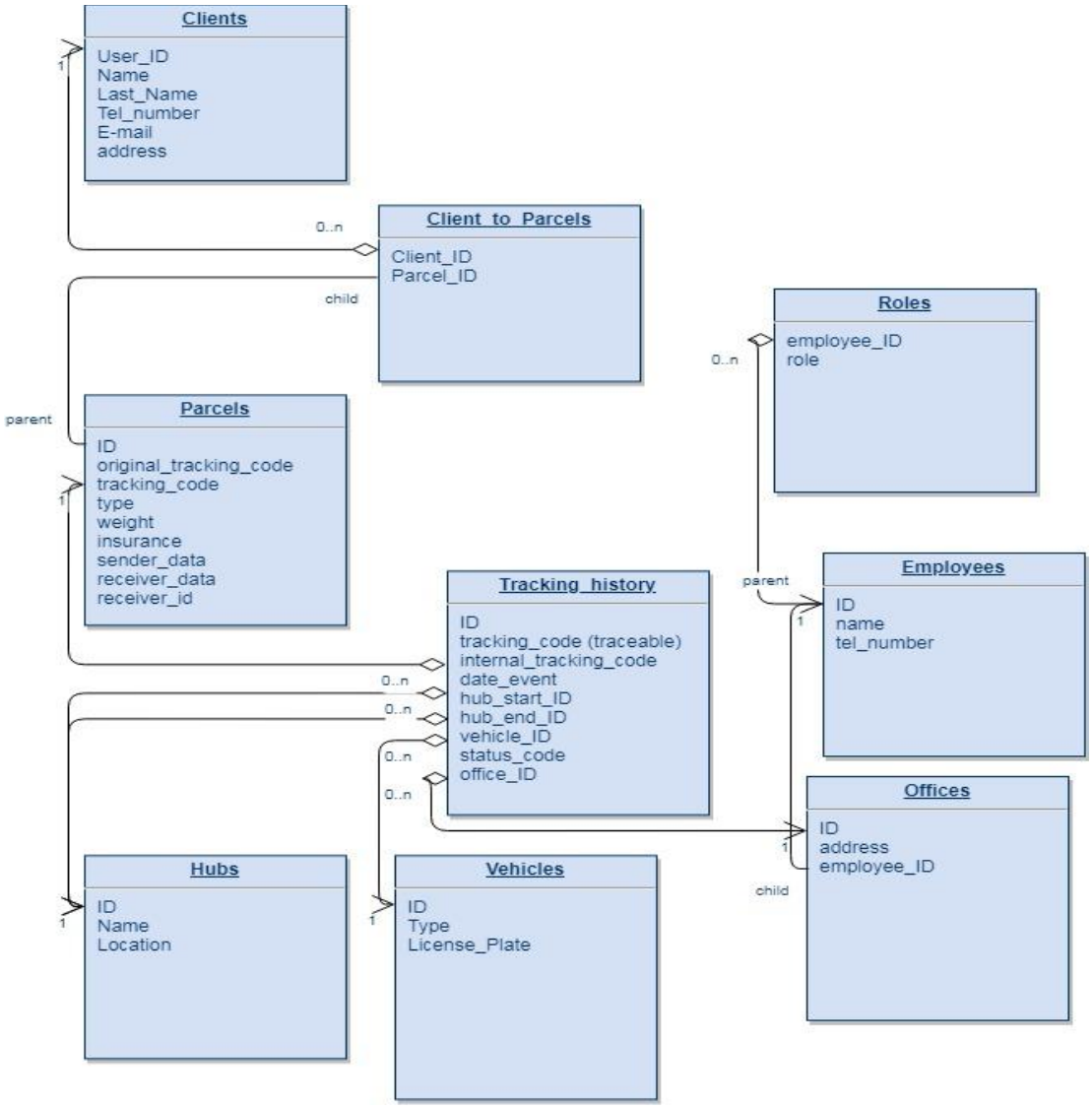


Figure 7 Model of solution by storing data in Apache Kudu for logistics area
Source figure is based on Dr. L. Mileva's research.

The presented model of solution of the defined problem (Fig. 3) is expressed by:

1) marking in the system of the original code, which is given by the first supplier in the chain;

2) search for matching customer data, data, such as phone number and e-mail address, which will prevent problems caused by transliteration of customer names, as sometimes the transport is carried out between neighbouring countries, but in different languages such as Latin and Cyrillic (within the EU);

3) assigning a new code for tracking the shipment, which would allow the customer to track his shipment, according to the old number, which the company gave him initially upon request to the sender;

The conclusions made so far are about the positive consequences of applying a prototype solution model in logistics in big data analysis in the Hadoop ecosystem using the analytical tool Apache Kudu. They present the possibilities for comprehensive and continuous monitoring of shipments, cargo and goods.

The conclusion is that no data on shipments will be lost during the entire logistics process. This is important for big data analysis, as continuous monitoring will make it possible to analyse:

- direction of transport of goods and cargoes;
- tracking the movement of consignments from one country to another;
- orders of one client;

The possibilities to track orders of a customer could lead to benefits for the organization such as offering various promotional services, concluding contracts at lower prices, a package for loyal customers and more.

- sales revenue (taking into account the movement of consignments from one country to another and vice versa;

It will be possible to assess the risk of losses in transport to destinations that carry a higher risk of damage to resources.

- transport costs;
- profitability of transport;

Except the following features there should be considered limitations when creating the model:

- economic factors;
- geographical factors;
- features of the cargo;
- paid transportation fee. In addition to the loss of data in the tracking of goods, there is another serious problem of transportation- the presence of empty shipments during transportation, which leads to losses of funds.

General problems also arise from the fact that unforeseen events can occur during transport, for which it is difficult to predict how long they would be removed, such as a flat tire, an overturned vehicle, border closures, etc. The proposed model for improving the work aims to reduce the legal burden on land through an optimization task. For the purposes of the task are described, the available criteria and they can be further divided into categories: geographical, economic and infrastructural. Additional restrictions may arise when, as full optimization is not possible in some specific cases, independent of the forecast data, for example a vehicle waiting in line due to a traffic jam of unclear nature (for example: accident, inverted vehicle, etc.), and in some cases human resources will also be used. For the purposes of designing a prototype model for analysis and decision.

4. Analysis and interpretation of the results for the field of logistics

Tables 1 to 7 examines the cases of fright transport loaded and empty divided into national and international, represented by Bulgaria and its neighbour countries. The data presenting the performed transportations in vehicle kilometres and thousand journeys. For purposes of the article were made certain calculations. The data about Romania are not available in Eurostat database so that they are not represented in this article. Croatia is presented like country nearest Bulgaria’s neighbour Serbia that is not in EU. The data are collected from Eurostat (*Eurostat*, 2021) database.

NATIONAL

Table 1

Growth rates of Loading status “Empty” (in million vehicle kilometers) by transport coverage “National” transport (in million vehicle kilometers)

Country	2019/2018	2018/2017	2017/2016
Bulgaria	0,807	0,930	1,030
Greece	0,948	1,017	0,925
Croatia	1,000	1,005	1,053

Source: The table is based on author’s calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

According to Table 1 results, there is a slight decrease in the share of empty shipments, which marks a positive trend for Bulgaria for 2019/2018 compared to 2018/2017. For neighboring Greece, the results increase as the share of vacancies also decreases, but for the last year compared to the previous 2019/2018 is higher than in Bulgaria. In comparison, Croatia maintains the position of the coefficient around one, which means that there

are no major changes in the share of empty shipments, but the values are above one, which means that empty shipments are fact (2018/2017, 2017/2016).

Table 2

Growth rates of loading status “Empty” (in thousand journeys) by transport coverage “National” transport (in thousand journeys)

Country	2019/2018	2018/2017	2017/2016
Bulgaria	0,823	0,932	1,023
Greece	1,031	0,942	0,971
Croatia	1,056	1,001	0,998

Source: The table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

The expectation is that the rates of change in the relative share of empty transportation (calculated on the base of million vehicle kilometres and thousand journeys) will be close in relation to all loads. This rule applies for the three surveyed countries (for national transport) and the three different periods. The exception is for Greece (2018/2017), where the rate of change in the relative share of empty loads (calculated on the basis of million vehicle kilometres) is over 1 and the rate of change in the relative the share of empty loads (calculated on the basis of thousand journeys) is below 1. This fact shows Greece's targeted efforts to reduce the relative share of empty loads as thousand journeys at national level.

Table 3

Ratio between loading status “Empty” (in million vehicle kilometers) divided by loading status “Total loaded and empty” by transport coverage “National” transport (in million vehicle kilometers)

Country	2019	2018	2017	2016
Bulgaria	0,139	0,384	0,372	0,375
Greece	0,377	0,394	0,387	0,392
Croatia	0,110	0,381	0,375	0,375

Source: table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

For the growth rate indicator, empty journeys are compared to the total journeys in national level, Bulgaria shows stable data for 2016-2018 and a decrease in 2019 of the empty journeys compared to the total transport. Greece has shown also decrease in the last year. Croatia shows data similar to those for Bulgaria, which shows a positive trend.

INTERNATIONAL

Table 4

Growth rates of loading status “Empty” (in million vehicle kilometers) by transport coverage “International” transport in million vehicle kilometers)

Country	2019/2018	2018/2017	2017/2016
Bulgaria	0,700	0,640	1,000
Greece	N/A	N/A	N/A
Croatia	1,084	1,021	0,979

Source: table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

For Bulgaria we have a decrease for 2018/2017 (Table. 4), with a slight increase in 2018/2019. Data for Greece at the international level are missing. For Croatia, we have a slight increase in 2018/2017 and this trend continues in 2019 in terms of empty vehicle in million vehicle kilometres in international traffic. For Croatia, the number one has been exceeded, which means that unloaded shipments have increased. Attention should be focused on international transport in order to reduce empty transportations by road.

Table 5

Growth rates of loading status “Empty” (in thousand journeys) by transport coverage “International” transport (in thousand journeys)

Country	2019/2018	2018/2017	2017/2016
Bulgaria	0,808	0,657	1,203
Greece	N/A	N/A	N/A
Croatia	0,994	1,056	0,973

Source: The table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

It is expected the rate of change in the relative share of empty loads relative to all loads (calculated on the basis of million vehicle kilometers and total journeys) to be close. The rule applies to the three countries surveyed for international transport.

Table 6

Ratio between loading status “Empty” (in million vehicle kilometers) divided by loading status “Total loaded and empty” (in million vehicle kilometers) by transport coverage “International” transport

Country	2019	2018	2017	2016
Bulgaria	0,098	0,107	0,117	0,109
Greece	0,087	N/A	N/A	0,108
Croatia	0,174	0,161	0,171	0,176

Source: The table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

The ratio between Loading status “Empty” (in million vehicle kilometers) divided by loading status “Total loaded and empty” by transport coverage “International” transport (in million vehicle kilometers) (Table 6) shows seriously low values especially for 2019 for international level. Levels for Greece and Croatia are almost the same as Bulgarian.

Table 7

Road transport year 2019

Country	National level		International level	
	empty VKM/total loaded and empty VKM	**empty TJ/total loaded and empty TJ**	**empty VKM/total loaded and empty VKM**	**empty TJ/total loaded and empty TJ**
Bulgaria	0,399	0,493	0,098	0,346
Greece	0,377	0,467	0,087	0,185
Croatia	0,375	0,453	0,174	0,359

VKM – million vehicle kilometers
TJ – thousand journeys
Source: The table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

The last table (Table 7) shows road freight transport for 2019. It makes an impression that at national level there is an increase for the three countries in number of journeys without cargo. As for the number of empty shipments to all shipments at international level, there is a high coefficient of empty shipments 0.346. At the same time as the distance traveled by the vehicles is empty- the coefficient is small 0.098, which means that the vehicles travel to very short distances empty. Statistics reports a high number of empty courses, but they are very short distances. This is due to the large number of logistics centers in Western Europe located on the out of the cities. After delivering the cargo to the address/store, the vehicle return empty in the direction of the respective logistics hub and take the next cargo. The values for all remain higher because there are not so many logistics centers in the monitored countries. Therefore, it is important to pay attention to the optimization of transport in Bulgaria and its neighboring countries.

Therefore, a model will be created that helps to overcome the problems of lack of traceability of goods throughout the supply chain and optimize shipments so that they are carried out at the lowest possible cost, with the greatest possible profit.

4. Summary and proposal for solving the problem in the field of logistics

From the conclusions made so far, it becomes clear that solving the problem of incomplete tracking of goods along the entire transport route, which leads to data loss, will contribute to solving another serious discrepancy- the problem of available empty traffic, which leads to losses of financial resources for organizations. The solution by implementing an information system in Apache Kudu, by assigning a shipment number for easier and continuous tracking, should also lead to a solution to address the problem of available empty rates for carriers. This can be done by optimizing processes, with the help of questionnaires, which will be used to assess the condition of the road. They will help you choose the right route in real time.

The data will be entered, processed and stored in Apache Kudu (as part of the Hadoop ecosystem), after the process is finished. In this way, they will be avoided possible empty courses of transportation. The proposed solution will implement through a prototype optimization model using the Apache Kudu analytics tool in the Hadoop ecosystem. This will be possible by compiling a model with help of questionnaires. The model presented finding of most profitable route at the corresponding restrictions and weights. When formulating the model, they are important dependencies on the type and type of the vehicle, and additional conditions arise (for example, the direction of the route). Therefore, they a given different weights for the same sections of the route from point A to point B. As an example, a common problem exists in transit traffic when crossing the borders of countries due to driver testing for Covid-19, hence several changes. For example, changing the quarantine period for drivers further blocks the transport of logistics companies. Within 14 (10) days, the vehicle remains blocked, with 20% of the fleet of a small company remaining unusable. For small organizations, this means that they suffer greater losses due to the same external factors, but different internal- for example, in the presence of 5-6 vehicles. For large and medium-sized companies the situation will be different, i.e. when creating an optimization model, one factor will be determining- the size of the organization. As a rule, in case of complete lockdown of a country that is in transit on a given route, the route is extends the journey and the cost of cargo transportation becomes more expensive. This leads to losses for the carrier company. Therefore, the route passes optimization through implementing the questionnaires results in the Apache Kudu analytical tool of the Hadoop ecosystem. However, with such extension of routes, other destinations, which do not lead to extension of the road or go through a longer route with lower costs, are sought. This can be expressed, for example, in lower tolls, cheaper fuel, lower prices of vignettes, reducing the cost of accommodation and day meals for drivers. In the event of a change, it may be possible to transfer additional cargo to the country through which the vehicle passes in order to bypass the closed country. Some of the proposed criteria for modelling the analysis by implementing criteria assessment questionnaires in apache Kudu for the most profitable route are presented in Table 5. Model is based on obtaining and analysing real-time information for an alternative route, at the lowest possible cost. For the purposes of constructing the column of criteria, values (Table 8) from A to D of the criteria are assigned, according to the conditions of the environment at the time of departure and later during the trip (A is the highest and D the highest (lowest)). Estimates of the criteria may change throughout the trip if any of the conditions change. The data are collected in the Apache Kudu storage with the ability to process and analyse in real time. It is important to notice that in this case an important factor is the distance travelled with a load vehicle. According to him, the model is also being built, as it became clear that the transports without cargo in many countries exceed those with cargo. The aim is to offer a solution through which to optimize the transport and the aim is to have them with cargo, so that there can be a greater profit for the logistics organization. Route calculation estimates are given as follows in Table 8.

Table 8

Questionnaire for evaluating the strength of each criterion

	Criteria	**Strength of each criterion**
	Price of transportation- per km (incl. permanent fees- (security, vignette, annual technical inspection, tax)	
	Road network condition	
	Weather conditions	
	Distance with loaded vehicle (in km)	
	Distance with empty vehicle (in km)	

Fill in A, B, C, D or E for each criterion, where “A” is the strongest influence; “E” is the weakest influence

Source: Authors’ contribution.

Table 9

Aggregated data from a conveyed survey among 5 experts

Criteria	**Aggregated data**				
	A	**B**	**C**	**D**	**E**
Price of transportation- per km (incl. permanent fees- (security, vignette, annual technical inspection, tax)	5				
Road network condition			1	4	
Weather conditions			4	1	
Distance with loaded vehicle (in km)		5			
Distance with empty vehicle (in km)					

Responses are A, B, C, D or E for each criteria, where “A” is the strongest influence; “E” is the weakest influence.

Source: Authors’ contribution

Aggregating data from 5 experts are showed in table 9. According to ranking, each evaluation of criteria could occupy a specific value in the selected range. As some factors are variable, they do not depend on the expert opinion itself, but depend on the specific environmental conditions. For example, when the country borders closed (due to Covid-19 situation), the information should be submitted by the relevant authorities, reported by the organization, entered as data in the information system and taken into account when choosing delivery routes. The same will be the case with factors such as weather road conditions. In order to assess the meteorological situation, the assessment comes from relevant specialists such as meteorological service as an expert statement, which is permanently and its assessment does not change. In advance, only the experts’ statements about this factor, re-prioritizing the criteria. This will be repeated for each factor separately. In addition, factors involving the type of vehicle, road surface and others play a major role in making a decision. All evaluated grades at precisely defined point in time will be summarized and the completed assessment will enter the company's information system. Before, entering the information system, the evaluation of the criteria must receive a numerical value. Factors are distributed according to the degree of importance and the points they have received. Experts and specialized commissions should outline the scale for assessing the factors, and appropriate software will extract the data from the various sources in order to determine the most profitable route or the route with the least losses. The recorded data will enter the Apache Hadoop system and will be processed and stored in Apache Kudu so that it can be accessed in real time.

6. ICT prototype solution with analytical tool Apache Kudu in logistics area

This chapter discusses the most important moment about working with the model.
General rules of the model.

1. Requirements for its implementation.
2. Key points.
3. Used methods.
4. Scope and elements of the model.
5. Methodology for application.
6. Logical sequence of the model evaluation.
7. Conclusions and recommendations.

For the article purposes, on the base of the made studies, will be constructed a prototype model to address specific problems, in particular the problems resolving the lack of traceability along the entire route chain and optimizing routes to minimize unloaded traffic, as a survey has shown that a large proportion of shipments in EU countries are empty. The elements of the model and its use are characterized, which are based on the analysis and assessment of the condition of the transports with the help of road transport by land. The coefficient ratio of the travelled distances with empty transportation (international to national) is calculated. Conclusions are made for the article. It has a comparison between the kilometres travelled in the EU member states. General, national and international transportations are considered. The results of the study introduce several points:

1) main problems in logistics; 2) the factors influencing the work of the logistics organizations in their part for transportation of cargoes, goods and consignments; 3) the points of dependence at which the criteria would save costs and / or bring a profit; 4) the opportunities and areas in which to look for solutions to improve the work.

The model reflects our view that built on the base of a systematic approach outlines the activities and processes that complement each other and are interconnected, through the criteria for optimization of transport to find the most profitable transport route. In addition, when considering the elements of the model, it becomes clear that performing activities of different nature in the process of transporting goods from point A to point B, the system optimization should collect and store data through the Apache storage. Kudu. This process consists of complying with various important criteria. They are located at different levels, in different activities and are performed in different periods of time. The activities related to the optimization of the logistics processes have been established. In a certain period of time, when transporting goods from point A to point B, some factors influence, and in another period - different from the first. At the same time, it is possible in some cases to perform parallel calculations when assessing the situation when deciding which route the vehicle should take in order to avoid the unloaded distance. From the analyses made it can be concluded that the travel criteria are of key importance for optimizing the implementation of logistics processes in their part when transporting goods from point A to point B.

The main principles to be followed in the model are the rules of transparency and objectivity in evaluation; observance of clear rules and systematics in the assessment of the optimization criteria for cargo transportation from item A to item B.

The requirements for the application of the model are outlined: 1) The need for specific knowledge and skills for the processes of the specialists who will deal with the development of a system for optimization of transport through the analytical tool Apache Kudu, in the Hadoop ecosystem. 2) Reliability of the information and observance of the requirements for setting the goals of the model - for clarity, specificity, reachability, measurability. 3) Precisely defined deadline, according to their importance. 4) Logical sequence of execution of the stages. The key points are:

a) The model is valid for the processes in logistics, by continuous tracking of shipments and selection of the most appropriate route, requiring an integrated systematic approach in planning, analysis and management b) The information, used in model needs to be update regularly to meet the needs of the economy and management, including planning and analysis. c) Model should be implemented in the Hadoop ecosystem through data storage in Apache Kudu d) The system approach provides serious opportunities for analogies of processes and phenomena and therefore the model discussed here can be used in various other areas (not only in logistics). Next, the methods used in the model are included: 1) systematic analysis; 2) comparative analysis; 3) content analysis; 4) method of generalization and 5) method of deduction. Sources for collecting information from: analyses of the Ministry of Transport, Information Technology and Communications, the National Statistical Institute, Eurostat; content analysis of the web pages of some logistics companies; documentation of the Kudu analytical tool in Cloudera and other studies by the author. The objectives are: 1) a proposal for the

implementation of part of the conceptual model of an information system with the possibility of assigning an additional code for tracking cargo, so as not to interrupt the traceability of the shipment 2) article of data on shipments with cargo, without cargo and total for EU countries. The information is collected through the above methods, the criteria that affect the transport of goods and consignments by land by road are differentiated.

The data processing accomplishes with the help of questionnaires cards, with the development of part of an information logistics system. This is in order to assign a code number to track the cargo along the entire route from point A to point B and rate the route criteria to find on the optimal route (with the least losses) and to implement taken decisions in big data storage system of Apache Kudu. The methodology for application of the model is divided into stages, which include analysis and remodelling (if necessary) which solves the problem of loss of data for loads in tracking from item A. to item B, through which data will be collected for shipments in Apache Kudu, especially in cases where the shipment is carried out by different companies and passes through different countries. Data analysis through the criteria is important for improving the delivery process and calculating the most profitable route with help of a criteria questionnaire assessment. The evaluation of the quality of logistics service includes results such as: 1. Development of a conceptual model of a part of an information system, by assigning a code for tracking shipments along the entire route from item A to item B by storing it in Apache Kudu. 2. Analysis and evaluation of land transport by road for EU countries with and without cargo. 3. Determining the degree of threat from passing routes without cargo, which leads to losses (transportation inside and outside the country) (considering the data of the EU member states). 4. Establishment of a prototype of a model, by answer questionnaire about finding the most optimal way according to certain criteria in order to improve the activity of the logistics organizations from the point of view of performing transportations with the least losses, caused by an empty vehicle. 5. Identify capabilities for big data analysis using the Apache Kudu analytics tool in the Hadoop ecosystem.

Elements- model for analysis of the most profitable route presents a model of a systems approach, where the elements are input, transformation processes and output. The classical model of the systems approach includes several elements; these are input, transformation and management processes, output and feedback. The model is illustrated by fig. 10.

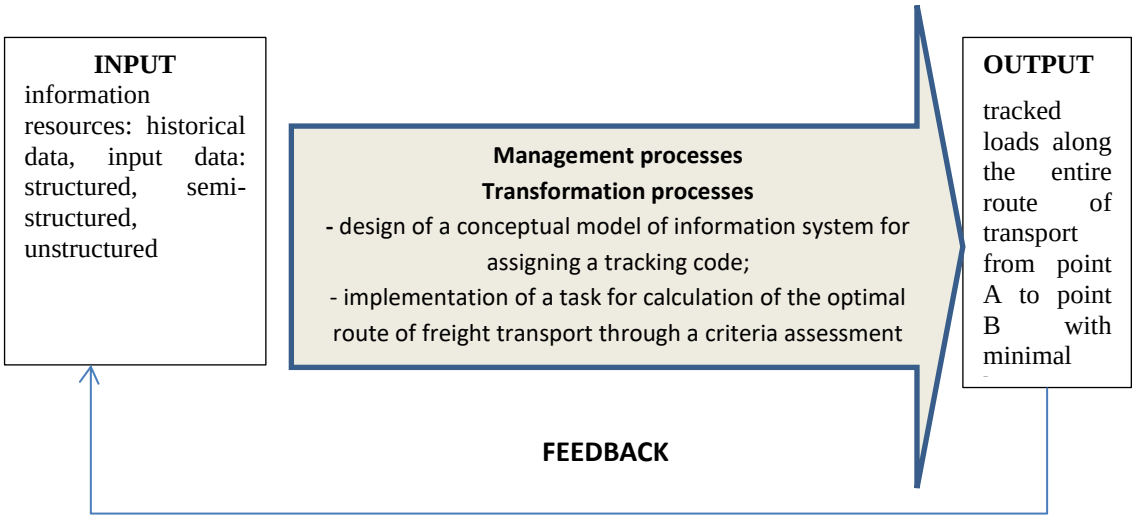


Figure 8 Prototype of a big data analysis model in Apache Kudu storage
Source: The figure is based on authors’ research

Resources that work in the operation of the system represents Model input. In this case, it is better to integrate model based on a systems approach, in order to store and process the data using the Apache Kudu analytics tool as part of the Hadoop ecosystem. Here the important resources for the model are the information resources. They are represented by historical transport data, all delivery data, routes, points through which the cargo pass, travelled distance, travelled time, delivery costs, route changes and all others that are collected and processed with traceability in real time. Then came the management processes and transformation processes,

represented by business processes. The management processes are expressed in the degree of management of the tracking information system and the optimization system. By collecting large data in the Apache Kudu storage, as part of the Hadoop ecosystem, the processes of transporting goods, cargo and shipments are tracked. The criteria characterizing the efficiency of the transport are determined and defined on the base of the research of performed transportations in and from the EU member states.

The results of the study lead to the conclusion that one of the most important criteria in transportation is the distance travelled with cargo. The studies also showed that there were too many kilometres travelled in different Member States, leading to losses. Also important for transportation are the cost of transportation and the distance travelled in kilometres. There are cases in which it is better to cover the entire distance without cargo, but they are significantly less than in other cases. Route selection criteria play a significant role in calculating the most profitable route through criteria assessment questionnaires. Calculating the optimal transport route is part of the transformation process. Other factors, such as the condition of the road surface, closure of a road from a certain route due to repairs and accidents and the meteorological situation are also observed here. It should be noticed that there are also conditions in which due to poor road and meteorological conditions there is no possibility to choose an alternative route. This reduces the choices to zero. However, most trips have options available, so optimization can be done in other cases. The result obtained after the calculation of the optimal route is in the output as an element of the model, which has tracked cargo along the entire route of transport from point A to point B with minimal losses.

Conclusion

The capabilities of big data analysis with Apache Kudu (as part of the Hadoop ecosystem) are expressed with the implementation of the model and data storage with following goals: 1) prevention of data losses due to interruption of traceability of supply codes and 2) optimizing routes, reducing costs and increasing profits for organizations in logistics. Regarding the possibilities for researching additional criteria, in the calculation of the most profitable route outline as follows:

1) the shortest route;
2) route passing on main roads; route according to the type of cargo (for special cargo);
3) route according to the type of vehicle;
4) traffic jams and repairs;
5) closed areas due to Covid 19;
6) closed areas requiring PCR or other tests (during pandemics and other crisis situations);
7) areas with overloaded traffic;
8) areas requiring increased attention due to unsuitable weather conditions.

Solving the listed problems will lead to the following benefits and opportunities for business organizations: 1. Minimization of costs for transportation on route from item A to item B without cargo. 2. Increasing the revenues from transportation along the route. 3. Reducing the risks of misuse of loads by drivers. 4. Reducing the risk of damage to cargo. 5. Reduction of costs for storage of shipments. 6. Reduce the time spent on a vehicle. 7. Greater control over shipments; 8. Greater staff control; 9. Greater flexibility; 10. Reduce the cost of food for a longer stay. The benefits of integrating the model into the Hadoop ecosystem will also be achieved with the help of the Apache Kudu analytical tool, which: 1. Ability to track in real time the cargo at any point along the route from point A to point B. 2. Ability to select compliance requirements on request, including the option of strict consistency of cargo.

At the next stage, the model can use machine learning to make predictions and analyses. To develop the formula for calculating whether the road is good or not, data collected from vehicles already on a route can be used by comparing the estimated time to travel on a route with the actual travelled time. For example, if the estimated time has been set and n serial vehicles passing this route show a longer time than the estimated one, then it follows that in the future, vehicles should not go this way. Another possibility to use machine learning and data storage in Apache Kudu is to track the average speed of vehicles kilometre by kilometre and determine the problematic sections of the road by making calculations with different coefficients the different sections again kilometre by kilometre and thus obtaining more accurate calculations. In this way, logistics organizations will add value to their activities.

Literature

Abdel-Badeeh M. Salem, S. P. (2018) ‘Developing a web-based ontology for e-business. Int J Electron Commer Stud. 2018;9(2):119–32’.

Ana-Maria Ramona, S., Marian Pompiliu, C. and Stoyanova, M. (2020) ‘Data Mining Algorithms for

Knowledge Extraction', in *Challenges and Opportunities to Develop Organizations Through Creativity, Technology and Ethics*, pp. 349–357. doi: 10.1007/978-3-030-43449-6_20.

Apache Kudu - Overview (2017). Available at: https://kudu.apache.org/overview.html.

Baesens, B. (2017) 'The big effects of big data'.

Baesens, B. (2019) 'Real-world techniques for analyzing big data | SAS', *Sas*. Available at: http://www.sas.com/en_us/insights/articles/big-data/real-world-big-data-tips.html.

Calzon, B. (2021) 'Your modern business guide to data analysis methods and techiques', *Datapine*. Available at: https://www.datapine.com/blog/data-analysis-methods-and-techniques/.

Cristescu, M. P. (2019) 'Specific aspects of the optimization of the reengineering processes of the distributed information applications', in *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, pp. 627–636. doi: 10.5593/sgem2019/2.1/s07.082.

Eurostat (2021). Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en.

Heidenreich, H. (2018) 'What are the types of machine learning?' Available at: https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f.

Ileanu, B. V. *et al.* (2019) 'Intriguing behavior when testing the impact of quotation marks usage in Google search results', *Quality and Quantity*, 53(5), pp. 2507–2519. doi: 10.1007/s11135-018-0771-0.

Kirova, A. (2009) 'Innovations in the transport services market', *(in Bulgarian) Mechanics ISSN 1312-3823 Transport issue 3, 2009 Communications*, 3(3). Available at: https://mtc-aj.com/library/388.pdf.

Kuyumdzhiev, I. (2020) 'A model for timely delivery of it solutions for Bulgarian universities', in *20th International Multidisciplinary Scientific GeoConference Proceedings SGEM 2020, Informatics, Geoinformatics and Remote Sensing*, pp. 3–10. doi: 10.5593/sgem2020/2.1/s07.001.

Medvedev, A. and Sergeev, A. (2020) 'No The forecast of the methane concentration changes for the different time periods on the arctic island bely. In: AIP Conference Proceedings. 2020. p. 120019.'

Miryanov, R. and Petkov, J. (2017) 'An application of the relation between arithmetic mean and geometric mean for rational proving some inequalities', *Mathematics and informatics*, 60(4), pp. 363–369.

Nacheva, R. *et al.* (2019) 'Concept map mining approach based on the mental models retrieval', *TEM Journal*, 8(4), p. 1484. doi: 10.18421/TEM84-54.

Nacheva, R. and Sulova, S. (2020) 'Internationalization in Context of Education 4.0: AHP Ranking of Bulgarian Universities', in *ACM International Conference Proceeding Series*, pp. 278–284. doi: 10.1145/3407982.3408006.

Nacheva, R. and Sulova, S. (2021) 'Research on the Overall Attitude Towards Mobile Learning in Social Media: Emotions Mining Approach', in *Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data*. Springer, Cham, pp. 429–440. doi: 10.1007/978-3-030-65722-2_27.

Nikolaev, R., Milkova, T. and Miryanov, R. (2018) 'A new meaning of the notion "expansion of a number"', *Mathematics and informatics*, 61(6), pp. 596–602.

Parvanov, H. and Boeva, O. (2009) 'Problems of European transport integration', *(in Bulgarian)*. Available at: https://www.unwe.bg/uploads/ResearchPapers/Research Papers_vol2_2006_No2_H Parvanov, O Boeva.pdf.

Polkowski, Z. *et al.* (2020) 'Evaluation of aggregated query plans using heuristic approach', in *Proceedings of the 12th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2020*, pp. 1–4. doi: 10.1109/ECAI50035.2020.9223222.

Quinto, B. (2018) 'High Performance Data Analysis with Impala and Kudu', *Next-Generation Big Data*, pp. 101–111. doi: 10.1007/978-1-4842-3147-0_4.

Raeesi Vanani, I. and Majidian, S. (2020) 'Literature Review on Big Data Analytics Methods', *Social Media and Machine Learning*. doi: 10.5772/intechopen.86843.

Stoyanova, M. (2020) 'Good practices and recommendations for success in construction digitalization', *TEM Journal*, 9(1), pp. 42–47. doi: 10.18421/TEM91-07.

Todoranova, L. *et al.* (2020) 'A model for mobile learning integration in higher education based on students' expectations', *International Journal of Interactive Mobile Technologies*, 14(11), pp. 171–182. doi: 10.3991/ijim.v14i11.13711.

Todoranova, L. and Penchev, B. (2020) 'A Conceptual Framework for Mobile Learning Development in Higher Education', in *ACM International Conference Proceeding Series*, pp. 251–257. doi: 10.1145/3407982.3407996.

Integrating Distributed Hadoop System into the Existing Infrastructure

Stefka PETROVA[1], Liliya MILEVA[2], Pavel PETROV[3], Plamen YANKOV[4], Julian VASILEV[5]

[1] University of Economics, Varna, Bulgaria
s.petrova@ue-varna.bg

[2] University of Economics, Varna, Bulgaria
l.mileva@ue-varna.bg

[3] University of Economics, Varna, Bulgaria
petrov@ue-varna.bg

[4] University of Economics, Varna, Bulgaria
yankov.plamen@ue-varna.bg

[5] University of Economics, Varna, Bulgaria
vasilev@ue-varna.bg

Abstract. A distributed Hadoop system can integrate clusters of different organizations. The purpose of this article is to consider the options for building an architecture of a distributed Hadoop system, so that it is, on the one hand, to integrate a logically complete Hadoop system, and on the other hand – to define the conceptual framework of individual components that technically implement this. The scope of the research covers the problem of integrating remote Hadoop cluster to other one. A several findings are made and comparison between different ways of organization in data processing – multiple clusters and multi-lease are made. Both approaches have their advantages and disadvantages which could be used in practice.

Key words: Hadoop, big data, distributed systems, information systems, cluster management.

1. Introduction

Acknowledgments. This work is supported by the "Digitization of the Economy in an Environment of Big Data" Project - BG05M2OP001-1.002-0002.

The functions of a distributed Hadoop system are to integrate clusters of different organizations. Software, hardware, network, and organizational set of events are needed to realize this task. Hadoop clusters are designed to store and process large amounts of data in a distributed heterogeneous computing environment. Characteristically, they are scalable, and if they do not have enough computing resources due to the large amount of data, new nodes can be added to the cluster in order to solve the task and increase productivity. Since the same data is stored on different nodes in one cluster, in case of failure of one of the nodes the work can be continued using data from the other nodes (fig. 1). The number of copies can be set, and the default is 3.

One Hadoop cluster can connect and use data from another Hadoop cluster through different approaches.

The distributed Hadoop system, which integrates clusters of different organizations, must be connected via high-speed networks to the other components of the big data ecosystem:

a) LAN - with exchange speeds of 10 Gbps, 50 Gbps, 100 Gbps
b) MAN - with an exchange rate of 1 Gbps, 10 Gbps
c) WAN - with an exchange rate of 1 Gbps, 5 Gbps, 10 Gbps
d) LoraWAN - with exchange rate 1Mps, 10Mps, 100Mbps (fig. 2)
e) GSM network - with exchange speed 10Mps, 100Mbps, 1Gbps
f) Firewall with IPS / IDS capabilities (fig. 3)

The following specialists categories are required for a distributed Hadoop system to integrate clusters of different organizations:

a) Server hardware maintenance technicians - with tasks to keep the servers operational. Duration - until the end of the project;

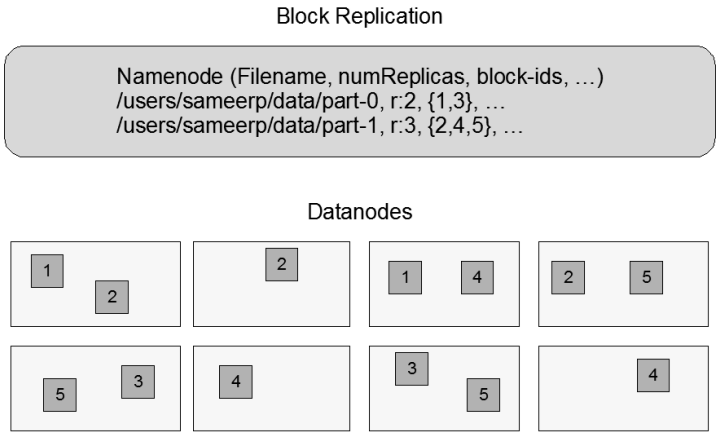


Figure 1. Replication of data blocks on DataNode servers
Source: HDFS Architecture, https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

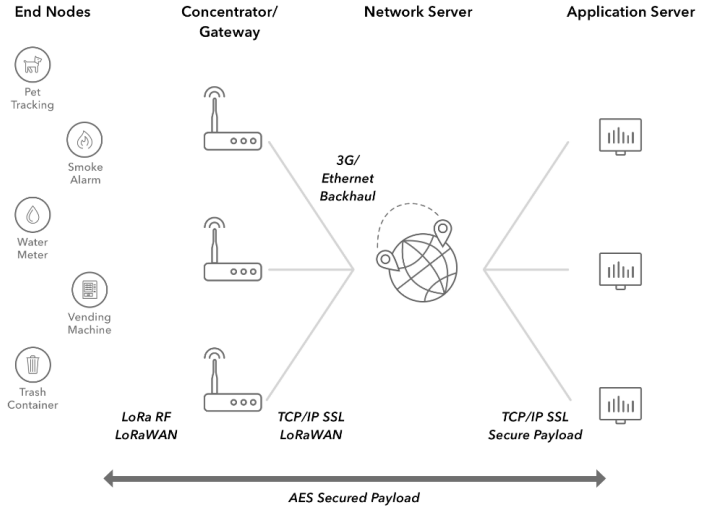


Figure 2. Using the LoRaWAN architecture as a means of communication
Source: LoRaWAN Architecture, https://www.thethingsnetwork.org/docs/lorawan/architecture.html

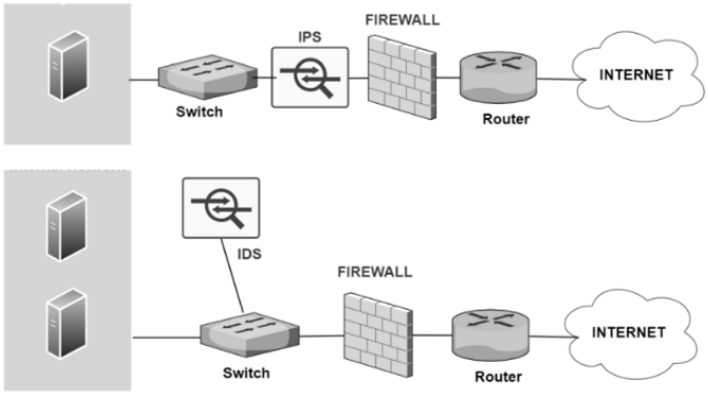


Figure 3. Using the Intrusion Prevention System (IPS) and the Intrusion Detection System (IDS) in combination with a firewall
Source: Comparison and Differences Between IPS vs IDS vs Firewall vs WAF, https://www.networkstraining.com/firewall-vs-ips-vs-ids-vs-waf/

b) Technical specialists in the hardware maintenance of the network components - with tasks to maintain in working order the network devices and those for the information security. Duration - until the end of the project;

c) Software specialists for administration of Hadoop system - with the task to be able to re-install Hadoop system, to maintain in working condition Hadoop system; to configure the respective modules - services of Hadoop system. Duration - until the end of the project;

d) Hadoop system application software specialists - with the task to be able to develop application programs using the main software components - Hadoop system services. Duration - after the launch of the Hadoop system until the end of the period of development and testing of application programs;

e) Software specialists for data analysis from Hadoop system - with the task to choose a suitable software tool for data analysis (Sulova 2019) located in Hadoop system, as well as to be able to operate with these tools (Stoyanova 2020). Duration - after the launch of the Hadoop system until the end of the period of development and testing of application programs.

2. Literature review

When integrating an external distributed Hadoop system (cluster) into another existing Hadoop infrastructure, it is possible to use different approaches (Radev 2019; Aleksandrova 2018), depending on the objectives pursued.

According to the Hadoop documentation, the operating modes of the system that are officially supported are three:

1. Local (stand-alone) mode. This is the default mode when installing on a computer, which does not use the HDFS file system, but the local file system of the computer when performing input / output operations. In this variant of using Hadoop, for the purposes of integration with existing infrastructure, it is necessary to make appropriate changes in the configuration files core-site.xml, hdfs-site.xml and mapred-site.xml in order to move to full distributed mode of operation (Apache 2020a).

2. Pseudo-distributed mode. It is a single-node cluster that uses HDFS. In this mode, both the NameNode, which contains the list of all directories and files, and the DataNode, where the data is stored, are located on the same computer. For the purposes of integration with existing infrastructure, it is necessary to make appropriate changes to the configuration files core-site.xml, hdfs-site.xml and mapred-site.xml in order to switch to fully distributed mode (Apache 2020b).

3. Fully distributed mode. It is a cluster with many nodes in which the data is distributed and processed in each of them. In this situation, the following configuration parameters are set in the main configuration files for connection with the other nodes in the cluster:

- File /etc/hadoop/conf/core-site.xml - HDFS settings;
- File /etc/hadoop/conf/hdfs-site.xml - settings for NameNode (and possibly Secondary NameNode) and multiple DataNode;
- File /etc/hadoop/conf/yarn-site.xml - settings for ResourceManager, NodeManager and History Server;
- /etc/hadoop/conf/mapred-site.xml file - MapReduce settings.

Combined big data processing in a fully distributed mode of operation of one cluster with another cluster can be performed in three main operating scenarios.

In the first scenario, the two clusters operate relatively independently of each other, whereby the data from one cluster is transferred (copied) for processing to the other, after which the result of the processing is eventually transferred back. The DistCp tool (Cloudera 2019) can be used for this purpose when transferring large volumes of data. Another option is to use the real-time HFTP tool, which allows data to be transferred between remote Hadoop clusters that use HDFS. HFTP works on HTTP and data access is read-only, i.e., post-processing write operations must be performed in the local cluster.

In the second scenario, the two clusters merge statically to work as one large cluster and the resources of one distributed Hadoop system become part of the resources of the other Hadoop system (Ryu 2018; IBM Knowledge Center 2019). Special care is required when setting network settings in terms of increasing security.

In the third scenario, the resources of the two clusters are dynamically pooled by specially designed systems that build on Hadoop (Wang et al. 2013; Jeon et al. 2014).

Each of the listed scenarios has its advantages and disadvantages, which makes it suitable for one situation or another. Each of the main scenarios may have different variations (Kuyumdzhiev 2019) related to the details of a particular implementation.

3. Main functional components of the distributed architecture

The main functional components (blocks) of the architecture of a distributed Hadoop system, which integrates clusters of different organizations, are the following:

a) Main central cluster, part of the infrastructure of the Center of Competence. It can include clusters of external organizations.

b) Additional Hadoop cluster of a research company / organization, consisting of several components of type DataNodes and type NameNode / Secondary NameNode (fig. 4). The main system software components are: Hadoop Distributed File System (HDFS) - for distributed data storage in a file system, and Hadoop YARN - for managing computing resources and their distribution among many different users.

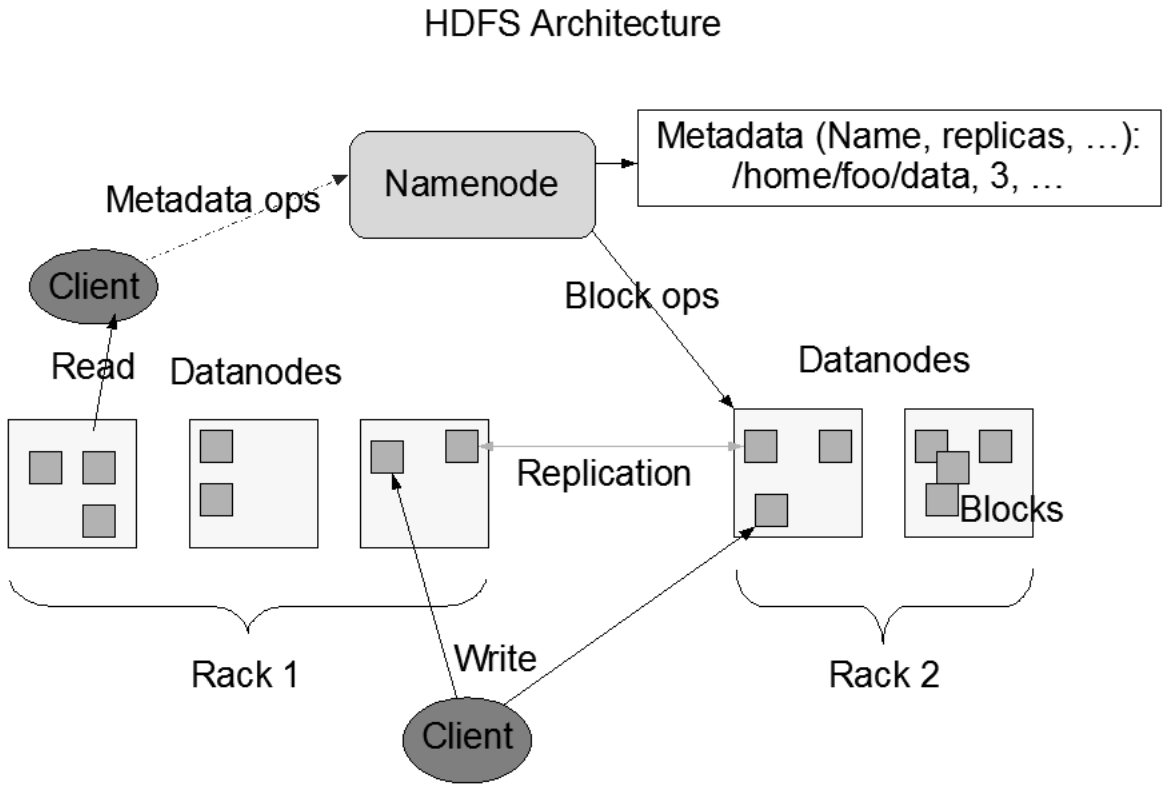


Figure 4. Replication of data blocks on different DataNode servers

Source: HDFS Architecture, https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

The scenario in which two clusters merge statically to work as a large "single" cluster and the resources of one distributed Hadoop system become part of the resources of the other Hadoop system sets certain requirements. In practice, the nodes of the new "single" Hadoop cluster are located not in one, but in several data centers, geographically located in different places. It is necessary for all nodes in the cluster to be accessible through the network, i.e. their IP addresses to be available to other nodes in the network.

There are several options for merging the networks of different geographically remote data centers:

1. To use a separate telecommunication line between the two networks for connection and thus to build a common LAN. This is a relatively expensive option in case a rent had to be paid for a leased line to a telecommunications company.

2. To use the public Internet for connection. It is possible to apply two main tactics:

a) All nodes have public IP addresses - providing a large number of IP addresses version 4 can be a problem, as their number is limited and scarce nowadays. The use of IP addresses version 6, where there are no problems with a shortage of IP addresses, is not possible at this stage, as it is not officially supported (HadoopIPv6 2019). A significant problem is the protection of each node from various attacks specific to the Internet space. The guidelines provided by Cloudera (Ahluwalia 2017) can be used for this purpose.

b) Use of virtual private networks - VPN (Virtual Private Networks). Virtual private network technology makes it possible to provide an encrypted connection between private networks by using a public network, such

as the Internet, as the transmission medium. Possible options are the use of GRE (Generic Routing Encapsulation), PPTP (Point-to-Point Tunnelling Protocol), L2F (Layer 2 Forwarding), L2TP (Layer 2 Tunnelling Protocol), IPSec (Internet Protocol Security) and MPLS (MultiProtocol Label Switching). Each of the options has its advantages and disadvantages in the implementation of the so-called. "Site-to-site VPN" and is recommended for use by various companies, such as Microsoft and Cisco.

One of the main problems in connecting several geographically remote clusters is that Hadoop generates a very large amount of traffic between nodes in the common cluster. In addition, time latency reduces the speed of data processing, as there is more waiting between the individual nodes that work together. It is very likely that time latency will be the "bottleneck" in the cluster and limit overall performance.

In the scenario where two Hadoop clusters operate relatively independently of each other, it is possible that part of the resources of one cluster will be made available for use if these resources are larger than the resources of the other cluster or need to be run in parallel. big data processing. In both cases, data from one cluster must be transferred for processing to the other and possibly the result must be transferred back (for example via the distcp tool).

The provision of some of the resources to the Hadoop cluster is known in the literature as "multitenancy" or "multi-lease". Generally speaking, this is: "a software architecture in which a single instance of an application running on a server serves multiple client organizations (so-called tenants)." (CIO Media 2014). In case of multirental, the Hadoop cluster software works simultaneously with several different configurations and data from several organizations. Each of the client organizations has access to a separate instance of the virtual application and sees only its data.

The YARN tool can be used to organize multi-tenancy (Nguyen&Won 2017). It separates resource management from big data processing workflows. This is done through separate program modules that work together - common to all ResourceManager module and a separate module for each application ApplicationMaster. The ResourceManager module allocates resources between applications by communicating with the NodeManager module, which is responsible for managing the resources of a physical machine - processor time, memory, disk space and network. The ApplicationMaster module, which is responsible for one application, declares the use of resources by ResourceManager and works with one or more NodeManager to perform work tasks (fig. 5).

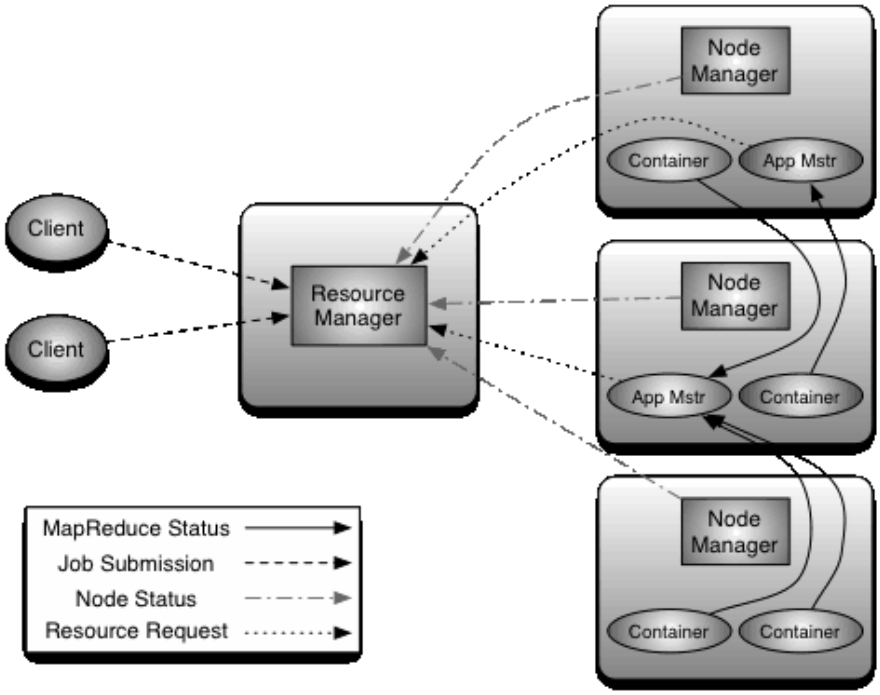


Figure 5. Interaction between ResourceManager and NodeManager
Source: https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html

ResourceManager has two main components:

1. Scheduler - is responsible for allocating resources between different running applications, subject to set restrictions. The rules under which resources are allocated can be changed or built-in allocation schemes such as

CapacityScheduler and FairScheduler can be used. The differences between the two schemes are not large and it should be borne in mind that Clodera in its product Cloudera Data Platform (CDP) only supports CapacityScheduler and offers a tool fs2cs to convert the rules from FairScheduler to CapacityScheduler (Clodera Docs 2020). For this reason, the use of CapacityScheduler is probably more promising.

2. ApplicationsManager - is responsible for accepting job requests, negotiating a container in which to run ApplicationMaster and restarting the ApplicationMaster container if necessary. In turn, ApplicationMaster is responsible for negotiating with the Scheduler of containers with appropriate resources (CPU time, memory, disk space and network) and monitors the operation of the containers.

Several factors need to be considered when choosing a multi-tenancy approach. A Hadoop cluster can generally be easily expanded with new hardware resources, but opportunities (such as power supply, air conditioning, etc.) and space (space in the server room) must be provided in advance. Therefore, it is necessary to correctly determine the initial size of the cluster, so as not to subsequently require frequent changes in terms of hardware resources. Based on the data processing needs, it is possible to choose an approach for several small clusters managed individually or one large cluster serving many users. Both approaches have their advantages and disadvantages (table 1).

Table 1.

Comparison between different ways of organization in data processing - multiple clusters and multi-lease

Way of organi-zation / Parameter	Several smaller Hadoop clusters	Large single Hadoop cluster with multi-lease
Data duplication	Specific data is stored in each cluster. Duplicate data between clusters is possible.	The data of several research companies / organizations are stored. It is possible to use a "data lake" type structure, avoiding duplication when storing shared data. Company / organization-specific data can be stored in separate directories with access control provided.
Load of hardware resources	Each cluster has its own hardware resources. Their use can be both weak and overloaded. If the cluster is designed to handle peak loads, it is very likely that hardware resources are underloaded most of the time.	Hardware resources can be utilized efficiently through YARN. By creating a schedule for using the system within a day or a longer period, a fuller use (load) of hardware resources can be achieved.
Cluster management	Different system procedures (such as upgrades and debugging) must be performed on each of the clusters separately. These increases maintenance costs compared to a centralized management option. The advantage is that problems in one of the clusters do not directly affect the work of the others.	Maintenance activities are performed centrally for the entire cluster, which reduces costs. The disadvantage is that in these periods the work on all projects can be blocked.

Source: Own observations

In case when all servers in the clusters must have public IP addresses nftables network tool and nftables.conf configuration files with the same content could be used. The settings in the configuration files block access to any server from a computer on the Internet, except those with explicitly specified IP addresses. In this case, the goal of using nftables in a Hadoop cluster is to allow access to each server in one cluster only when the network traffic is from a server in the other Hadoop cluster. Eventually only the traffic to ports 22 (SSH protocol), 80 (HTTP protocol) and 443 (HTTPS protocol) could be publicly allowed for the purpose of remote administration of the specific server and output of publicly available system information via a web server.

4. Conclusion

Several conclusions could be made as a result of this research. Integrating an external distributed Hadoop system (cluster) into another existing Hadoop infrastructure is difficult to implement because, by default, the system is most often used in a local area network that is separate from the Internet. However, it is possible to use

different approaches, depending on the objectives pursued. The operating modes of the system, according to the official documentation of Hadoop, are three: Local (standalone) mode, Pseudo-distributed mode, and fully distributed mode. The latter mode can be used to integrate distributed Hadoop system to the existing infrastructure.

When connecting several geographically remote Hadoop clusters a very large amount of traffic between nodes are generated. Moreover, the speed of data processing is reduced due to the time latency, and it is very likely that the time latency will be the "bottleneck" and will limit overall performance.

Literature

Ahluwalia, M., 2017. How to secure "Internet exposed" Apache Hadoop, https://blog.cloudera.com/how-to-secure-internet-exposed-apache-hadoop/

Aleksandrova, Y., 2018. Application of Machine Learning for Churn Prediction Based on Transactional Data (RFM Analysis). *18 International Multidisciplinary Scientific Geoconference SGEM 2018: Conference Proceedings*, 2-8 July 2018, Albena, Bulgaria: Vol. 18. Informatics, Geoinformatics and Remote Sensing. Iss. 2.1. Informatics, Sofia: STEF92 Technology Ltd., pp.125-132.

Apache, 2020a. Hadoop 3.3.0: Setting up a Single Node Cluster. Available at: https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/SingleCluster.html, accessed 20.08.2020.

Apache, 2020b. Hadoop 3.3.0: Hadoop Cluster Setup. Available at: https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/ClusterSetup.html, accessed 20.08.2020.

Apache, 2020c. Comparison and Differences Between IPS vs IDS vs Firewall vs WAF. Available at: https://www.networkstraining.com/firewall-vs-ips-vs-ids-vs-waf/

Apache, 2020d. HDFS Architecture. Available at: https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

Apache, 2020e. Apache Hadoop YARN. Available at: https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html

CIO Media, 2014. Predskazaniya za SaaS prez sledvashtata godina. Available at: https://cio.bg/management/2014/10/04/3440091_predskazaniia_za_saas_prez_sledvashtata_godina/

Cloudera Docs, 2020. Comparison of Fair Scheduler with Capacity Scheduler. Available at: https://docs.cloudera.com/runtime/7.2.1/yarn-reference/topics/yarn-FS-vs-CS.html

Cloudera, 2019. Administering HDFS, Using DistCp to Copy Files. Available at: https://docs.cloudera.com/HDPDocuments/HDP3/HDP-3.1.5/administration/content/using-distcp-to-copy-files.html, accessed 20.08.2020.

HadoopIPv6, 2019. Hadoop and IPv6. Available at: https://cwiki.apache.org/confluence/display/HADOOP2/HadoopIPv6

IBM Knowledge Center, 2019. Overview of InfoSphere Information Server on Hadoop. Available at: https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ishadoop.doc/topics/overview.html

Jeon, K., Chandrashekhara, S., Shen, F., Mehra, S., Kennedy, O., & Ko, S. Y., 2014. Pigout: Making multiple hadoop clusters work together. In *2014 IEEE International Conference on Big Data (Big Data)*, pp.100-109.

Kuyumdzhiev, I., 2019. *Metodologicheski i tehnologichni aspekti pri arhiviraneto na bazi ot danni*. Varna: Nauka i ikonomika, Bibl. Prof. TSani Kalyandzhiev; Kn.55.

Nguyen, M. C., & Won, H. S., 2017. Advanced Multitenant Hadoop in Smart Open Data Platform. In *Proceedings of the International Conference on Big Data and Internet of Thing*, pp.48-51.

Radev, M., 2019. Organizational Variants of IT Department at the University. *Information and Communication Technologies in Business and Education*: Proceedings of the International Conference Dedicated to the 50th Anniversary of the Department of Informatics, Varna: Science and Economic Publ. House, pp.284-290.

Ryu, W., 2018. Implementation of dynamic node management in Hadoop cluster. *International Conference on Electronics, Information, and Communication (ICEIC)*, Honolulu, HI, 2018, pp. 1-2, doi: 10.23919/ELINFOCOM.2018.8330612.

Stoyanova, M., 2020. Good Practices and Recommendations for Success in Construction Digitalization. *TEM Journal - Technology, Education, Management, Informatics* / Association for Information Communication Technology Education and Science, Novi Pazar, Serbia: UIKTEN - Association for Information Communication Technology Education and Science, 9(1), pp.42-47.

Sulova, S., 2019. An Approach to Storing Data Based on the Data Lake Concept to Facilitate Intelligence Data Analysis. *Serdica Journal of Computing*, Sofia: Institute of Mathematics and Informatics. BAS, 13(3-4), pp.171-182.

The Things Network, 2019. LoRaWAN Architecture. Available at: https://www.thethingsnetwork.org/docs/lorawan/architecture.html

Wang, L., Tao, J., Ranjan, R., Marten, H., Streit, A., Chen, J., & Chen, D., 2013. G-Hadoop: MapReduce across distributed data centers for data-intensive computing. *Future Generation Computer Systems*, *29*(3), pp.739-750.

Значение на дисциплината „Стокознание“ в обучението на студенти от различните професионални направления в България

Радослав РАДЕВ[1]

[1] Икономически университет - Варна, България
r.radev@ue-varna.bg

Резюме. Целта на настоящата разработка е да се проучи значението на дисциплината „Стокознание в обучението на студенти от различните професионални направления в България. С извършеното проучване е доказано значението на дисциплината „Стокознание“, в различните й разновидности за обучението и изграждането на специалисти от различни професионални направления в някои университети в България в областта на икономиката, администрацията и управлението, туризма и хранителните технологии, за които е от особена важност знанието за стоките. Основен център на стокознанието в България е катедра „Стокознание“ и единствената по рода си в страната специалност “Стокознание и митническа дейност“ към Икономически университет Варна. Настоящото изследване може да бъде добра основа за проучване на мнението на студентите, които изучават дисциплината „Стокознание“ относно ценността на придобитите знания, умения и компетенции и реалното им приложение в практиката. Освен това може да се проучи мнението и на различни работодатели относно необходмостта от кадри със знания за стоките. Получените резултати дават важна информация, която ще е необходимо да се вземе предвид при бъдещото развитие и усъвършенстване на обхвата на дисциплината „Стокознание“.

Ключови думи: Стокознание, дисциплината „Стокознание“, обучение на студенти

1. Въведение

Възникването и развитието на стоковедната наука е обусловено от развитието на стоковото производство и обръщение, от усъвършенстване на производствените, договорните отношения и потребителската култура. Необходимостта от знания за свойствата и полезността на стоките възниква още в древността, като по това време те са имали изцяло практическа насоченост и емпиричен характер. Особено бързо стоковедната наука се развива в началото на XIX век, когато стокообменът излиза от своите национални рамки и се развива международната търговия.

Няколко са етапите в историческото развитие на знанията за стоките:

- *Античен етап* – от този етап датират интересни сведения, свързани със същността, свойствата, съхранението на стоките в трудовете на различни автори от древността.
- *Средновековен етап* – по това време бурно развитие бележат естествените науки (биология, химия, физика), а това допринася за развитието на знанията за стоките. С развитието на производството се увеличава количеството и асортимента на стоките, развива се бурно търговията. През този етап възниква необходимостта от бързи и надеждни методи за окачествяване на стоките.
- *Епохата на просвещението* – формира се научната терминология и се поставят научните основи на стокознанието. През 1594 г. в университета в Падуа (Италия) се поставят основите на стокознанието като учебна дисциплина и се създава катедра по стокознание на растителните, животинските и фармацевтичните материали, а през 1739 г. е издаден един от първите учебници по стокознание – „Въведение в стокознанието или учение за важни чуждестранни стоки”. В този период стокознанието се въвежда като самостоятелна учебна дисциплина в учебните планове на редица университети.
- *Епоха на промишлената революция* – през 30-те години на XX век се диференцират знанията за хранителните и нехранителните стоки. През този етап от развитието на стокознанието описателният характер на научните изследвания се заменя със задълбочен изследователски подход относно формирането на качеството на стоките и влиянието на отделните фактори. От началото на 60-те години на XX век започва системното издаване на учебна, справочна и научно-практическа литература в областта на стокознанието.

➢ *Съвременен етап* - на настоящия етап стокознанието има два неразривно обвързани аспекта – научен и практико-приложен. Стоковедната наука се отличава със своя интердисциплинарен характер и с интегрирането на научни знания от различни области. Изграден е единен подход за изследване на хранителните и нехранителни стоки, базиран върху теоретична рамка, анализираща основните характеристики на всички видове стоки (Stoykova et. al., 2020).

Стокознанието в България

На 14 май 1920 г. е открито първото Висше търговско училище във Варна, днес Икономически университет – Варна. За основоположник на научното стокознание в България се смята неговият създател, първи ректор и професор по стокознание – професор Цани Каляндjиев. Той има редица изследвания за спецификата и състава на стоките, възможностите за фалшифицирането им и др. През 1922 г. той издава първият учебник по стокознание, предназначен за висши училища (Stoykova et. al., 2020).

Специалност „Стокознание” е традиционна за университета и уникална за страната. Създаването, развитието и утвърждаването на катедра и специалност „Стокознание” е неизменно свързано с историята на университета. Висшето търговско училище е първото, което въвежда стокознанието като учебна дисциплина във висшето образование в България. Социално-икономическите условия, потребността от научни изследвания и кадри с висше икономическо образование в областта на качеството на стоките, наличието на подготвени преподаватели, натрупаният опит от преподаването на дисциплината „Стокознание” и далновидността на проф. Цани Каляндjиев са основните фактори за откриването на специалност „Стокознание”. Това решение се взема на 19 октомври 1948 г. с протокол №68/19.10.1948 г. на ФС на факултета за стопански и социални науки към Държавния университет „Св. Кирил Славянобългарски”. Разработва се първият учебен план на специалността, като се използва опитът на западноевропейските страни и Русия. Катедра „Стокознание” към ИУ – Варна не е променяла името си от своето създаване. От 1948 до 2015 г. специалността запазва наименованието си „Стокознание”, а през ноември 2015 г., с решение на Академичния съвет на ИУ – Варна, е преименувана на „Стокознание и митническа дейност”. Днес тя е в състава на Стопански факултет към ИУ – Варна, който е единственото висше училище в България, обучаващо бакалаври по специалността (Stoykova, 2018).

Едновременно със специалността се създава и катедра „Стокознание”, която става първият академичен център в страната за подготовка на кадри с висше образование и научноизследователска дейност по стокознание. За първи ръководител в периода 1948-1949 г. е избран проф. Димитър Ганев. По късно се създава катедра „Стокознание” в ИУ „Карл Маркс” в София (сега УНСС), където се преподава дисциплината „Стокознание”, а след това тази дисциплина е застъпена в учебния план и за икономистите от Стопанска академия – Свищов (Stoykova, 2018).

Над 70 години катедра „Стокознание” следва актуалните икономически промени и предизвикателства, осъществява и ръководи обучение на специалисти в областта на стоковото познание, развива и задълбочава научните и приложните изследвания в областта на качеството и стоковия мениджмънт, предоставя решения на актуални проблеми на националния и международния бизнес, работи в сътрудничество с наши и чуждестранни учени в перспективни направления на стоковедната наука (Stoykova, 2018).

Развитието и устояването във времето на специалност „Стокознание”, „Стокознание и митническа дейност“ и катедра „Стокознание” се обуславя от факторите: качествен и висококвалифициран преподавателски състав; гъвкаво адаптирани актуални учебни планове и програми; дългогодишно създавана и модернизирана материално-техническа база; актуални научни изследвания; успешно завършили специалисти с многостранна реализация във всички области на жизнения цикъл на стоките (Stoykova, 2018).

Обучението на специалисти с висша квалификация в областта на стоковедната наука в ИУ – Варна е уникално за България. То се осъществява в съответствие с изменящите се нормативни изисквания, икономическите условия и потребностите на бизнеса. Развитието на специалността е свързано с периодични промени в учебните планове и програми, които многократно са актуализирани. Възпитаниците на катедрата работят на ръководни длъжности в производствени, търговски, логистични, туристически, маркетингови, консултантски и контролни фирми; в митниците и регионалните митнически дирекции; в организации за контрол и оценка на качеството; в сертифициращи фирми на системи и продукти; като експерти по надзор на пазар и защита на потребителите; мениджъри и отговорници по качеството; инспектори в контролни фирми; експерти в технически комитети по стандартизация; борсови стокови агенти; митнически посредници; ръководители на отдели по иновационен мениджмънт (Stoykova, 2018).

През своята над 70-годишна история катедра „Стокознание” се развива динамично в перспективни направления съобразно съвременните пазарни условия. Тя предлага богато научно наследство в областта

на стоковото познание, актуални и значими за практиката изследвания, прилага съвременни методи и средства за решаване актуални проблеми на пазарната икономика и изграждане доверие в качеството и безопасността на стоките. Осигурява висококвалифицирани кадри подготвени в областта на качеството на стоките и защита на потребителите. Актуалността и нарастващата необходимост от знания за стоките, изучаването на богат набор от стоковедни дисциплини е предпоставка за подготовка на кадри от катедра „Стокознание”, което определя нейната значимост и перспективност (Stoykova, 2018).

Редица са авторите (Donchev, 1998; Zhelyazkova, 2008a; Zhelyazkova, 2008b; Zhelyazkova et. al.,, 2006; Kozhukharov, 2013; Stefanov, 1995; Stefanov, 2008; Stoykova, 2018; Stoykova et. al., 2020 и др.), които проучват създаването, развитието и перспективите на катедра „Стокознание“ и специалност „Стокознание“, а от 2015 г. специалност „Стокознание и митническа дейност“, но по-малко е проучвано (Pashova, 2009) значението на дисциплината „Стокознание“ при обучение на студенти в различните професионални направления в България.

Целта на настоящата разработка е да се проучи значението на дисциплината “Стокознание” в обучението на студенти от различните професионални направления в България.

2. Проучване на обучението на студенти по дисциплината „Стокознание“ от различните професионални направления в университетите в България

Проучването относно обучението на студенти по дисциплината „Стокознание“ в различните професионални направления е осъществено чрез задълбочен преглед на учебните програми на интернет страниците на висшите училища в България в периода януари-февруари 2021 г.

В рамките на изследването е направено уточнението, че под дисциплината „Стокознание“ се има предвид т.нар общо „Стокознание“, а под дисциплината „Стокознание“ в различните ѝ разновидности се имат предвид всички останали стоковедни дисциплини в тяхното разнообразие.

В няколко таблици е представено обучението на студентите по дисциплината „Стокознание“ в различните професионални направления в университетите в България.

Таблица 1

Дисциплината „Стокознание“ включена в учебните планове на специалностите в професионално направление 3.7 Администрация и управление, ОКС бакалавър, квалификация - мениджър

Висше училище	Факултет	Катедра	Специалност	Дисциплина
ВВМУ „Н.Й.Вапцаров“- Варна	Факултет Навигационен	Мениджмънт и логистика	Логистика	Стокознание, задължителна
НВУ „Васил Левски“- Велико Търново	Факултет Общовойскови	Логистика на сигурността	Гражданска специалност “Стопанска логистика”	Стокознание, задължителна
НВУ „Васил Левски“- Велико Търново	Факултет Общовойскови	Логистика на сигурността	Гражданска специалност “Логистика на сигурността”	Стокознание, задължителна
Лесотехнически университет - София	Факултет Стопанско управление	Икономика и управление на ресурсите и природополването	Стопанско управление	Стокознание, задължителна

Източник: Собствено проучване и Blagoeva, 2017

В таблица 1 е отразено, че обучението на студенти по дисциплината „Стокознание“ в професионално направление 3.7. Администрация и управление в ОКС бакалавър се извършва в три университета в България - ВВМУ „Н.Й.Вапцаров“- Варна, НВУ „Васил Левски“- Велико Търново и Лесотехнически университет – София. Прави впечатление, че дисциплината „Стокознание“ е задължителна, а освен това завършва с изпит при всички специалности. В посочените университети дисциплината „Стокознание“ е преподавана пред студенти от специалностите „Логистика”, гражданска специалност „Стопанска логистика”, гражданска специалност „Логистика на сигурността” и „Стопанско управление”.

В таблица 2 е представено обучението на студенти от професионално направление 3.8 Икономика по дисциплината „Стокознание“ в различните ѝ разновидности, включени в учебните планове на някои специалности в ИУ-Варна, Стопански факултет.

Таблица 2

Дисциплината „Стокознание“ в различните ѝ разновидности, включена в учебните планове на специалностите в професионално направление 3.8 Икономика в ИУ-Варна, Стопански факултет, ОКС бакалавър, квалификация - икономист

Катедра	Специалност	Дисциплина
Стокознание	Стокознание и митническа дейност	Основи на стокознанието на промишлените стоки, задължителна
Стокознание	Стокознание и митническа дейност	Стокознание на битовите стоки, задължителна
Стокознание	Стокознание и митническа дейност	Стокознание на текстилните и обувни стоки, задължителна
Стокознание	Стокознание и митническа дейност	Основи на стокознанието на хранително-вкусовите стоки, задължителна
Стокознание	Стокознание и митническа дейност	Стокознание на храните от животински произход, задължителна
Стокознание	Стокознание и митническа дейност	Стокознание на храните от растителен произход, задължителна
Икономика и управление на търговията	Икономика и търговия	Стокознание, задължителна
Индустриален бизнес и логистика	Логистика	Стокознание, избираема

Източник: Собствено проучване

От таблицата (табл. 2) е видно, че в ИУ-Варна, Стопански факултет дисциплината „Стокознание“ се изучава в различните ѝ разновидности в областта на хранителните и промишлени стоки. Тази констатация и разнообразието от стоковедни дисциплини се дължи на факта, че в ИУ-Варна е единствената и уникална за страната ни специалност „Стокознание и митническа дейност“. Освен това дисциплината „Стокознание“ се изучава и от специалностите „Икономика и търговия“ и „Логистика“. Всички дисциплини в областта на стокознанието в ИУ-Варна завършват с изпит.

Таблица 3

Дисциплината „Стокознание“, включена в учебните планове на специалностите в професионално направление 3.8 Икономика в УНСС – София, ОКС бакалавър, квалификация – бакалавър по икономика

Факултет	Катедра	Специалност	Дисциплина	Контрол
Бизнес факултет	Икономика на природните ресурси	Аграрен бизнес	Стокознание, избираема	изпит
Бизнес факултет	Икономика на природните ресурси	Екоикономика	Стокознание, избираема	
Факултет Икономика на инфраструктурата	Икономика на търговията	Икономика на търговията	Стокознание, избираема	
Факултет Икономика на инфраструктурата	Логистика и верига на доставките	Бизнес логистика	Стокознание, избираема	текуща оценка
Факултет Икономика на инфраструктурата	Икономика на туризма	Икономика на туризма	Стокознание, избираема	

Източник: Собствено проучване

От таблица 3 е видно, че в УНСС-София дисциплината „Стокознание“ присъства в учебните планове на различни специалности „Икономика на търговията“, „Аграрен бизнес“, „Екоикономика“, „Бизнес логистика“, „Икономика на туризма“. При специалностите „Бизнес логистика“, „Икономика на туризма“ завършването на курса е с текуща оценка, а при останалите специалности е с изпит.

В таблица 4 е отразено обучението на студенти от професионално направление 3.9 Туризъм по дисциплината „Стокознание на храни и напитки“ в УХТ-Пловдив и по дисциплината „Стокознание“ в Колеж по туризъм – Варна. Обучението се осъществява пред специалностите „Хотелиерство и ресторантьорство“, „Туризъм“ и „Мениджмънт на хотели и ресторанти“ и завършва с полагане на изпит, като дисциплините са задължителни.

Таблица 4

Дисциплината „Стокознание“ в различните ѝ разновидности, включена в учебните планове на специалностите в професионално направление 3.9 Туризъм, ОКС бакалавър (вкл. професионален бакалавър)

Висше училище	Факултет	Катедра	Специалност	Квалификация	Дисциплина
УХТ-Пловдив	Стопански факултет	Хранене и туризъм	Хотелиерство и ресторантьорство	управител в хотелиерството и ресторантьорството	Стокознание на храни и напитки, задължителна
УХТ-Пловдив	Стопански факултет	Хранене и туризъм	Туризъм	бакалавър по туризъм	Стокознание на храни и напитки, задължителна
Колеж по туризъм - Варна	-	Туризъм	Мениджмънт на хотели и ресторанти	професионален бакалавър	Стокознание, задължителна

Източник: Собствено проучване

В таблица 5 са представени данни за обучението на студенти в професионално направление 5.12 Хранителни технологии, които изучават дисциплината „Стокознание“ в различните ѝ разновидности, в УХТ-Пловдив.

Таблица 5

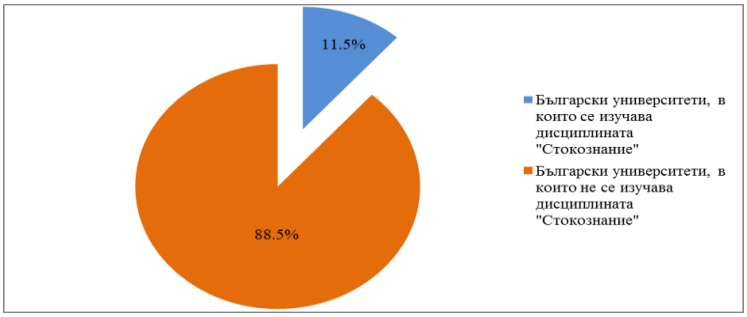
Дисциплината „Стокознание“ в различните ѝ разновидности, включена в учебните планове на специалностите в професионално направление 5.12 Хранителни технологии в УХТ-Пловдив, ОКС бакалавър

Факултет	Катедра	Специалност	Квалификация	Дисциплина
Стопански факултет	Хранене и туризъм	Хранене и туризъм	инженер-технолог	Стокознание на храни и напитки, задължителна
Стопански факултет	Хранене и туризъм	Храни, хранене и диететика		Стокознание на храни и напитки, задължителна
Стопански факултет	Хранене и туризъм	Кетъринг	инженер-технолог по кетъринг	Стокознание на храни и напитки, задължителна
Технологичен факултет	Технология на тютюна, захарта, растителните и етерични масла	Технология на тютюна и тютюневите изделия	инженер	Стокознание на тютюна, задължителна
Технологичен факултет	Технология на зърнените, фуражните, хлебните и сладкарските продукти и Технология на тютюна, захарта, растителните и етерични масла	Технология и дизайн на храни и ароматно-вкусови продукти		Стокознание на тютюна, задължителна

Източник: Собствено проучване

От таблицата (табл. 5) е видно, че дисциплината „Стокознание на храни и напитки“ е задължителна за студентите от специалностите „Хранене и туризъм“, „Храни, хранене и диететика“ и „Кетъринг“ и дисциплината „Стокознание на тютюна“, която е задължителна за студените от специалностите „Технология на тютюна и тютюневите изделия“ и „Технология и дизайн на храни и ароматно-вкусови продукти“ в УХТ-Пловдив. Посочените специалности завършват курса си на обучение с изпит.

На фигура 1 е представено изучаването на дисциплината „Стокознание“ в българските университети.



Фигура 1. Относителен дял на университетите в България, в които се изучава и в които не се изучава дисциплината „Стокознание“ в различните ѝ разновидности
Източник: Собствено проучване

От фигура 1 е видно, че в 11.5% от университети в България се изучава дисциплината „Стокознание“ в различните ѝ разновидности, а в 88.5% от тях не се преподава дисциплината. Въпреки, че дисциплината „Стокознание“ се преподава в едва 6 университета в страната, тя се изучава в различни професионални направления, което е доказателство за значението на дисциплината при формирането на знания, умения и компетенции при изграждането на различни специалисти.

В таблица 6 е представено изучаването на дисциплината „Стокознание“ в различните ѝ разновидности във висшите училища в България.

Таблица 6

Разновидности на дисциплината „Стокознание“ във висшите училища в България

Висше училище	Брой дисциплини
Икономически университет – Варна (ИУ-Варна)	8
Университет за хранителни технологии – Пловдив (УХТ-Пловдив)	7
Университет за национално и световно стопанство – София (УНСС-София)	5
НВУ „Васил Левски“- Велико Търново	2
Лесотехнически университет - София	1
ВВМУ„Н.Й.Вапцаров“- Варна	1
Колеж по туризъм - Варна	1
Общо	25

Източник: Собствено проучване

От таблицата (табл. 6) е установено, че най-висок брой (8бр.) на изучаване на дисциплината „Стокознание“ в различните ѝ разновидности се среща в ИУ-Варна. Причина за полученият резултат е, че в университета се обучават студенти в единствената и уникална за страната специалност „Стокознание и митническа дейност“, като в учебната програма са включени 6 специални стоковедни дисциплини (таблица 2) – „Основи на стокознанието на промишлените стоки“, „Стокознание на битовите стоки“, „Стокознание на текстилните и обувни стоки“, „Основи на стокознанието на хранително-вкусовите стоки“, „Стокознание на храните от животински произход“, „Стокознание на храните от растителен произход“. Освен това дисциплината „Стокознание“ в ИУ-Варна се изучава от специалностите „Икономика и търговия“ и „Логистика“ (таблица 2). В УХТ-Пловдив (7бр.) също се изучава в голяма степен стокознанието (таблици 4 и 5) в две основни дисциплини „Стокознание на храни и напитки“ пред специалностите – „Хотелиерство и ресторантьорство“, „Туризъм“, „Хранене и туризъм“, „Храни, хранене и диететика“, „Кетъринг“ и „Стокознание на тютюна“ пред специалностите – „Технология на тютюна и тютюневите изделия“ и „Технология и дизайн на храни и ароматно-вкусови продукти“. Останалите университети са подредени въз основа на резултатите в низходящ ред УНСС – София (5бр.), НВУ „Васил Левски“- Велико Търново (2бр.), Лесотехнически университет – София (1бр.), ВВМУ „Н.Й.Вапцаров“- Варна (1бр.), Колеж по туризъм – Варна (1бр.), а наименованието на дисциплината е „Стокознание.“

В таблица 7 са представени професионалните направления, в които се изучава дисциплината „Стокознание“ в различните ѝ разновидности.

Таблица 7

Разновидности на дисциплината „Стокознание“ във висшите училища в България по професионални направления

Професионални направления	Брой дисциплини
3.8. Икономика	13
5.12 Хранителни технологии	5
3.7 Администрация и управление	4
3.9 Туризъм	3
Общо	25

Източник: Собствено проучване

От таблица 7 е видно, че обучението по дисциплината „Стокознание“ в различните ѝ разновидности най-много се изучава в професионално направление 3.8 Икономика (13бр.), следвано от 5.12 Хранителни технологии (5бр.), 3.7 Администрация и управление (4бр.) и 3.9 Туризъм (3бр.). Посочените резултати доказват интердисциплинарния характер на дисциплината „Стокознание“, която се преподава на студенти от различни професионални направления и формира техните знания и умения, свързани с качеството, безопасността, стандартизацията на хранителни и промишлени стоки.

На таблица 8 са представени специалностите, в които се изучава дисциплината „Стокознание“ в различните ѝ разновидности в България

Таблица 8

Специалности, в които се изучава дисциплината „Стокознание“ в различните ѝ разновидности във висшите училища в България

Специалности	Брой дисциплини
Стокознание и митническа дейност	6
Логистика	5
Туризъм	5
Икономика, търговия	3
Храни, хранене, диететика и кетъринг	2
Хранителни технологии	2
Аграрен бизнес	1
Стопанско управление	1
Общо	25

Източник: Собствено проучване

От таблица 8 е видно, че с най-голям дял на изучаваните дисциплини в областта на стокознанието е специалност „Стокознание и митническа дейност“ с 6бр., следвани от изучаването на дисциплината „Стокознание“ в областта на логистиката 5бр. и в областта на туризма 5бр., в областта на икономиката, търговията 3бр., в областта на храните, храненето, диететиката и кетъринга 2бр. и хранителни технологии 2бр., а с най-малък дял са изучаващите дисциплината „Стокознание“ в специалностите „Аграрен бизнес“ 1бр. и „Стопанско управление“ 1бр. Полученият висок брой (табл. 8) при специалност „Стокознание и митническа дейност“ се дължи на разнобразието от стоковедни дисциплини в учебния план на единствената и уникална по своята същност специалност в нашата страна. Специалностите „Логистика“ и „Туризъм“ също са със сравнително висок резултат с по 5бр., тъй като предвид спецификата на тези специалности е от особено значение за студентите да изучават дисциплината „Стокознание“. Със средни стойности (3бр.) са специалностите в областта на икономиката и търговията, за които също е от особено важно значение да изучават дисциплината „Стокознание“.

В таблица 9 са представени специалностите, в които се изучава дисциплината „Стокознание“ в различните ѝ разновидности по професионални направления във висшите училища в България.

Таблица 9

Специалности, в които се изучава дисциплината „Стокознание“ в различните ѝ разновидности по професионални направления във висшите училища в България

	Професионални направления				
	ПН 5.12	ПН 3.9	ПН 3.8	ПН 3.7	Общо
Специалности					
Икономика, търговия			3		3
Логистика			2	3	5
Туризъм	1	3	1		5
Храни, хранене, диететика и кетъринг	2				2
Хранителни технологии	2				2
Аграрен бизнес			1		1
Стопанско управление				1	1
Стокознание и митническа дейност			1		1
Общо	5	3	8	4	20

ПН 5.12 Професионално направление „Хранителни технологии“;
ПН 3.9 Професионално направление „Туризъм“;
ПН 3.8 Професионално направление „Икономика“;
ПН 3.7 Професионално направление „Администрация и управление“
Източник: Собствено проучване

От таблицата (табл. 9) е установено, че в 20 специалности в различни университети в България се изучава дисциплината „Стокознание“ в различните ѝ разновидности, като най-висок е делът на специалностите (8бр.), които са в направление 3.8 Икономика, следвани от 5.12 Хранителни технологии (5бр.), 3.7 Администрация и управление (4бр.) и 3.9 Туризъм (3бр.).

Таблица 10

Наименование на дисциплината „Стокознание“, в различните ѝ разновидности по професионални направления във висшите училища в България

	Професионални направления				
	ПН 5.12	ПН 3.9	ПН 3.8.	ПН 3.7	Общо
Дисциплини					
Стокознание		1	7	4	12
Основи на стокознанието на промишлените стоки		предложение на автора	1		1
Стокознание на битовите стоки		предложение на автора	1		1
Стокознание на текстилните и обувни стоки		предложение на автора	1		1
Основи на стокознанието на хранително-вкусовите стоки			1		1
Стокознание на храните от животински произход			1		1
Стокознание на храните от растителен произход			1		1
Стокознание на храни и напитки	3	2			5
Стокознание на тютюна	2				2
Общо	5	3	13	4	25

Източник: Собствено проучване

В най-много от случаите (табл. 10) дисциплината „Стокознание“ в различните ѝ разновидности се среща в направления 3.8 Икономика (13бр.), следвани от 5.12 Хранителни технологии (5бр.), 3.7 Администрация и управление (4бр.) и 3.9 Туризъм (3бр.). По-голямото разнообразие от дисциплината „Стокознание“ се среща в направление 3.8 Икономика, тъй като в ИУ-Варна е единствената за страната специалност „Стокознание и митническа дейност“, в която се изучават шест стоковедни дисциплини – три стоковедни дисциплини в областта на промишленото стокознание и три стоковедни дисциплини в областта на хранителното стокознание.

От таблица 10 е видно, че най-често срещаното наименование на дисциплината е „Стокознание“ (12бр.), следвано от „Стокознание на храни и напитки“ (5бр.), „Стокознание на тютюна“ (2бр.), „Основи на стокознанието на промишлените стоки“ (1бр.), „Стокознание на битовите стоки“ (1бр.), „Стокознание на текстилните и обувни стоки“ (1бр.), „Основи на стокознанието на хранително-вкусовите стоки“ (1бр.), „Стокознание на храните от животински произход“ (1бр.), „Стокознание на храните от растителен произход“ (1бр.). Наименованието на дисциплината „Стокознание“ е най-срещано, тъй като в по-голяма част от случаите дисциплината се преподава, като т.нар. общо „Стокознание“, като основната цел е да подпомогне изграждането на специалисти в различни области и направления на знанието 3.8. Икономика (7бр.) 3.7 Администрация и управление (4бр.), 3.9 Туризъм (1бр.), където са необходими стоковедни знания, умения и компетенции. Дисциплината „Стокознание на храни и напитки“ е профилирана в даване знания и развиване на умения в направление 5.12 Хранителни технологии (3бр.) и 3.9 Туризъм (2бр.). Дисциплината „Стокознание на тютюна“ е фокусирана върху изграждането на специалисти със знания, умения и компетенции в областта на тютюна в направление 5.12 Хранителни технологии (2бр.). Останалите шест дисциплини в областта на стокознанието (промишлено и хранително) се изучават в направление 3.8. Икономика.

В таблица 10 е използван и методът на правоъгълника (Vasilev & Kehayova-Stoycheva, 2017), с помощта на който могат да се намерят възможности за приложение на дисциплината „Стокознание“ в различните ѝ разновидности по професионални направления. Дисциплините „Основи на стокознанието на промишлените стоки“, „Стокознание на битовите стоки“ и „Стокознание на текстилните и обувни стоки“ биха могли да се обединят и преподават в професионално направление 3.9 „Туризъм“, като част от дисциплина с наименование „Стокознание“. При обучението на студенти в професионално направление 3.9 „Туризъм“ се преподават дисциплините „Стокознание“ и „Стокознание на храни и напитки“, при които акцента е върху качеството на храните. В стоковедните дисциплини, в които се изучават промишлени стоки се придобиват знания, умения и компетенции за различни стокови групи – тъкани, готово облекло, кожи, обувки, стъклени стоки, керамични стоки, метални стоки, битови електрически стоки, дървени стоки, мебели, пластмасови стоки, стоки /средства/ за пране и измиване, хартия и картон, което може да бъде от изключително полза за студенти, които се обучават в професионално направление 3.9 „Туризъм“, тъй като бъдещите специалисти ще имат досег с посочените стокови групи. Предвид спецификата на дисциплините чрез метода на правоъгълника не са установени други възможности за преподаване на дисциплината „Стокознание“ в различните ѝ разновидности в посочените професионалните направления.

От проведеното проучване е установено, че в рамките на изследването преобладава задължителния характер на дисциплината „Стокознание“ със 76%, а като избираема дисциплина „Стокознание“ се изучава в 24%. Получените резултати се дължат на факта, че в изчисленията са включени задължителните шест стоковедни дисциплини, които се изучават в специалност „Стокознание и митническа дейност“ в ИУ-Варна, но и при изключването им дисциплината „Стокознание“ със задължителен характер преобладава (68.42%), в сравнение с дисциплината „Стокознание“ (31.58%), която се изучава избираемо. Посочените резултати са доказателство, че по-голяма част от отделните катедри в различните университети в България, в които се преподават стоковедни дисциплини са осъзнали необходимостта от задължително изучаване на дисциплината „Стокознание“, която формира знания за стоките на студентите и поради тази причина тя преобладаващо е включена в учебните планове, като задължителна дисциплина. Освен това в 92% от случаите проучваните дисциплини „Стокознание“ в различните ѝ разновидности завършват с изпит, а при 8% завършват с текуща оценка.

3. Значение на дисциплината „Стокознание“ при обучението на студенти от различните професионални направления в университетите в България

Въз основа на представените таблици и фигури в разработката биха могли да се обобщят следните основни констатации:

- обучението на студенти по дисциплината „Стокознание“ в професионално направление 3.7. Администрация и управление се извършва в три университета в България - ВВМУ „Н.Й.Вапцаров“- Варна, НВУ „Васил Левски“- Велико Търново и Лесотехнически университет – София;
- обучението на студенти в ИУ-Варна в професионално направление 3.8 Икономика по дисциплината „Стокознание“ в различните ѝ разновидности се извършва в областта на хранителните и промишлени стоки. Тази констатация и разнообразието от стоковедни дисциплини се дължи на факта, че в ИУ-Варна е единствената и уникална за страната ни специалност „Стокознание и митническа дейност“. Освен това в ИУ-Варна дисциплината „Стокознание“ се изучава и от специалностите „Икономика и търговия“ и „Логистика“;

- в УНСС-София дисциплината „Стокознание“ влиза в учебния план на различни специалности „Икономика на търговията“, „Аграрен бизнес“, „Екоикономика“, „Бизнес логистика“, „Икономика на туризма“, които са в професионално направление 3.8 Икономика;
- обучението на студенти от професионално направление 3.9 Туризъм по дисциплината „Стокознание на храни и напитки“ в УХТ-Пловдив и по дисциплината „Стокознание“ в Колеж по туризъм – Варна се осъществява пред специалностите „Хотелиерство и ресторантьорство“, „Туризъм“ и „Мениджмънт на хотели и ресторанти“;
- дисциплината „Стокознание на храни и напитки“ е задължителна за студентите от специалностите „Хранене и туризъм“, „Храни, хранене и диететика“ и „Кетъринг“ и дисциплината „Стокознание на тютюна“, която е задължителна за студентие от специалностите „Технология на тютюна и тютюневите изделия“ и „Технология и дизайн на храни и ароматно-вкусови продукти“ в УХТ-Пловдив в професионално направление 5.12 Хранителни технологии;
- в 11.5% от университети в България се изучава дисциплината „Стокознание“ в различните ѝ разновидности, а в 88.5% от тях не се преподава дисциплината. Въпреки, че дисциплината „Стокознание“ се преподава в едва 6 университета в страната, тя се изучава в различни професионални направления, което е доказателство за значението на дисциплината при формирането на знания, умения и компетентции при изграждането на различни специалисти;
- с най-висок брой (8бр.) на изучаване и най-голямо разнообразие на дисциплината „Стокознание“ в различните ѝ разновидности се среща в ИУ-Варна, поради факта, че в университета се обучават студенти в единствената и уникална за страната специалност „Стокознание и митническа дейност“, като в учебната ѝ програма са включени шест специални стоковедни дисциплини. Освен това дисциплината „Стокознание“ в ИУ-Варна се изучава от специалностите „Икономика и търговия“ и „Логистика“. В УХТ-Пловдив (7бр.) също се изучава в голяма степен стокознанието в две основни дисциплини „Стокознание на храни и напитки“ пред специалностите – „Хотелиерство и ресторантьорство“, „Туризъм“, „Хранене и туризъм“, „Храни, хранене и диететика“, „Кетъринг“ и „Стокознание на тютюна“ пред специалностите – „Технология на тютюна и тютюневите изделия“ и „Технология и дизайн на храни и ароматно-вкусови продукти“. В останалите университети подредени въз основа на резултатите в низходящ ред УНСС – София (5бр.), НВУ „Васил Левски“- Велико Търново (2бр.), Лесотехнически университет – София (1бр.), ВВМУ „Н.Й.Вапцаров“- Варна (1бр.), Колеж по туризъм – Варна (1бр.), а наименованието на дисциплината е „Стокознание“;
- обучението по дисциплината „Стокознание“ в различните ѝ разновидности се изучава в професионално направление 3.8 Икономика (13бр.), следвано от 5.12 Хранителни технологии (5бр.), 3.7 Администрация и управление (4бр.) и 3.9 Туризъм (3бр.). Посочените резултати доказват интердициплинарния характер на дисциплината „Стокознание“, която се преподава на студенти от различни професионални направления и формира техните знания и умения, свързани с качеството, безопасността, стандартизацията на хранителни и промишлени стоки;
- с най-голям дял на изучаваните дисциплини в областта на стокознанието е специалност „Стокознание и митническа дейност“ с 6бр., следвани от изучаването на дисциплината „Стокознание“ в областта на логистиката 5бр. и в областта на туризма 5бр., в областта на икономиката, търговията 3бр., в областта на храните, храненето, диететиката и кетъринга 2бр. и хранителни технологии (Технология на тютюна и тютюневите изделия и Технология и дизайн на храни и ароматно-вкусови продукти) 2бр., а с най-малък дял са изучаващите дисциплината „Стокознание“ в специалностите „Аграрен бизнес“ 1бр. и „Стопанско управление“ 1бр. Полученият висок брой (табл. 8) при специалност „Стокознание и митническа дейност“ се дължи на разнообразието от стоковедни дисциплини в учебния план на единствената и уникална по своята същност специалност в нашата страна. Специалностите „Логистика“ и „Туризъм“ също са със сравнително висок резултат с по 5бр., тъй като предвид спецификата на тези специалности е от особено значение за студентите да изучават дисциплината „Стокознание“. Със средни стойности (3бр.) са специалностите в областта на икономиката и търговията, за които също е от особено важно значение да изучават дисциплината „Стокознание“;
- в 20 специалности в различни университети в България се изучава дисциплината „Стокознание“ в различните ѝ разновидности, като най-висок е делът на специалностите (8бр.), които са в направление 3.8 Икономика, следвани от 5.12 Хранителни технологии (5бр.), 3.7 Администрация и управление (4бр.) и 3.9 Туризъм (3бр.). В най-много от случаите дисциплината „Стокознание“ в различните ѝ разновидности се среща в направления 3.8 Икономика (13бр.), следвани от 5.12 Хранителни технологии (5бр.), 3.7 Администрация и управление (4бр.) и 3.9 Туризъм (3бр.). По-голямото разнообразие от дисциплината „Стокознание“ се среща в направление 3.8 Икономика, тъй

като в ИУ-Варна е единствената за страната специалност „Стокознание и митническа дейност“, в която се изучават шест стоковедни дисциплини;

- най-често срещаното наименование на дисциплината е „Стокознание“ (12бр.), следвано от „Стокознание на храни и напитки“ (5бр.), „Стокознание на тютюна“ (2бр.), „Основи на стокознанието на промишлените стоки“ (1бр.), „Стокознание на битовите стоки“ (1бр.), „Стокознание на текстилните и обувни стоки“ (1бр.), „Основи на стокознанието на хранително-вкусовите стоки“ (1бр.), „Стокознание на храните от животински произход“ (1бр.), „Стокознание на храните от растителен произход“ (1бр.). Наименованието на дисциплината „Стокознание“ е най-срещано, тъй като в по-голяма част от случаите дисциплината се преподава, като т.нар. общо „Стокознание“, като основната цел е да подпомогне изграждането на специалисти в различни области и направления на знанието 3.8. Икономика (7бр.) 3.7 Администрация и управление (4бр.), 3.9 Туризъм (1бр.), където са необходими стоковедни знания, умения и компетенции. Дисциплината „Стокознание на храни и напитки“ е профилирана в даване знание и развиване на умения в направление 5.12 Хранителни технологии (3бр.) и 3.9 Туризъм (1бр.). Дисциплината „Стокознание на тютюна“ е фокусирана върху изграждането на специалисти със знания, умения и компетенции в областта на тютюна направление 5.12 Хранителни технологии (2бр.). Останалите шест дисциплини в областта на стокознанието (промишлено и хранително) се изучават в направление 3.8. Икономика;
- с помощта на метода на правоъгълника е намерена възможност за приложение в професионално направление 3.9 „Туризъм“ на дисциплината „Стокознание“, в която да бъдат включени стокови групи по промишлени стоки (тъкани, готово облекло, кожи, обувки, стъклени стоки, керамични стоки, метални стоки, битови електрически стоки, дървени стоки, мебели, пластмасови стоки, стоки /средства/ за пране и измиване, хартия и картон). Посоченото може да бъде от изключителна полза за студенти, които се обучават в професионално направление 3.9 „Туризъм“, тъй като бъдещите специалисти ще имат досег с посочените стокови групи.
- преобладава задължителния характер на дисциплината „Стокознание“ в различните ѝ разновидности със 76%, а като избираема дисциплина „Стокознание“ се изучава в 24%. Посочените резултати са доказателство, че по-голяма част от отделните катедри в различните университети в България са осъзнали необходимостта от задължително изучаване на дисциплината „Стокознание“, която формира знания за стоките на студентите и поради тази причина тя преобладаващо е включена в учебните планове, като задължителна дисциплина;
- в 92% от случаите проучваните дисциплини „Стокознание“ в различните ѝ разновидности завършват с изпит, а при 8% завършват с текуща оценка.

Представените резултати са доказателство за огромното значение на дисциплината „Стокознание“ при изграждането на специалисти със знания за свойствата и полезността на стоките от различни специалности и университети в България. Универсалният и интердисциплинарен характер на дисциплината „Стокознание“ (изучаване на хранително и промишлено стокознание) в различните ѝ разновидности я прави единствена и уникална за страната ни. Основен център на стокознанието в България е катедра „Стокознание“ и единствената по рода си в страната специалност “Стокознание и митническа дейност“ към ИУ-Варна. Дисциплината „Стокознание“ се преподава в шест висши училища пред различни специалности, което е доказателство за нейната необходимост по отношение на изграждане на специалисти със знания, умения и компетенции за стоките в областта на икономиката, логистиката, търговията, аграрния бизнес, стопанското управление, туризма, хранителните технологии, храните, храненето, диететиката и кетъринга.

Заключение

С извършеното проучване е доказано значението на дисциплината „Стокознание“ в различните ѝ разновидности за обучението и изграждането на специалисти от различни професионални направления в някои университети в България в областта на икономиката, администрацията и управлението, туризма и хранителните технологии, за които е от особена важност знанието за стоките. Основен център на стокознанието в България е катедра „Стокознание“ и единствената по рода си в страната специалност “Стокознание и митническа дейност“ към ИУ-Варна.

Настоящото изследване може да бъде добра основа за бъдещи проучвания в различни направления:

- проучване на мнението на студентите, които изучават дисциплината „Стокознание“ в различните ѝ разновидности относно ценността на придобитите знания, умения и компетенции и реалното им приложение в практиката;

- проучване на мнението на студенти и преподаватели относно възможността за обучение в професионално направление 3.9 „Туризъм“ по дисциплината „Стокознание“ в съдържанието на която да е включено изучаване на качеството на промишлените стоки;
- проучване мнението на различни работодатели относно необходимостта от кадри със знания за стоките;
- провеждане на кръгли маси, конференции, семинари с участието на преподаватели, които преподават дисциплината „Стокознание“ в различните й разновидности и студенти, които изучават стоковедните дисциплини и работодатели, които са заинтересовани от кадри със знания за стоките, на които да бъдат обсъждани състоянието, тенденции и перспективи за развитие на дисциплината „Стокознание“;
- съставяне на национална карта на дисциплината „Стокознание“, в която да бъдат включени всички университети, факултети, професионални направления, специалности в които се преподава дисциплината „Стокознание“;
- проучване на състоянието на дисциплината „Стокознание“ в Европа и света.

Получените резултати ще дадат важна информация, която ще е необходимо да се вземе предвид при бъдещото развитие и усъвършенстване на обхвата на дисциплината „Стокознание“ в България.

Literature / Използвана литература

Blagoeva, S., 2017. Current problems of the development of logistics in Bulgaria (in Bulgarian). Modern logistics – business and education*: Conference proceedings International scientific and practical conference*, Varna : Publishing house "Science and Economy", I, 23 - 38.

Donchev, H., 1998. 50 years of specialty "Commodity Science" (in Bulgarian). *"Izvestyia", Journal of University of Economics - Varna*, Varna: University publishing house „Science and Economics, 4, 86 - 96.

Zhelyazkova, M., 2008a. Sixty years of specialty "Commodity Science" (in Bulgarian). Course "Commodity Science" : Traditions and perspectives *: Conference proceedings, dedicated to the sixty years since the establishment of the specialty*, Varna : Publishing house "Science and Economy", 7 - 14.

Zhelyazkova, M., 2008b. Sixty years of specialty "Commodity Science" and Department of "Commodity Science" (in Bulgarian). *"Izvestyia", Journal of University of Economics - Varna*, 4, 112 - 117.

Zhelyazkova, M., Uzunova, I. & Zlateva, D., 2006. University education in specialty "Commodity Science" - relevance and european perspectives (in Bulgarian*). Varna:Conference proceedings “Quality of goods - problems and prospects”* - Varna, 300 - 308.

Kozhukharov, H., 2013. More about the commodity science as cource, specialty and science. Commodity science – traditions and relevance (in Bulgarian). *XII scientific conference with international participation : Celebrating sixty-five years specialty “Commodity science” and department “Commodity science”*, Varna: Graphic, 155 - 160.

Pashova, S. 2009. Need for the study of commodity science by professionals in tourism (in Bulgarian). *Conference proceedings International scientific and practical conference. Tourism and the sustainable development of society*. Varna : Publishing house "Science and Economy", 455-463.

Stefanov, D., 1995. Seventy-five years of study in commodity science at the University of Economics – Varna (in Bulgarian). *"Izvestyia", Journal of University of Economics - Varna*, Varna : University publishing house „Science and Economics, 2, 123 - 134.

Stefanov, D., 2008. Sixty years of the Department of “Commodity Science” : Organizer of the education in the specialty Commodity Science at the University of Economics, Varna (in Bulgarian). *Specialty "Commodity Science" : Traditions and perspectives : Conference proceedings, dedicated to the sixty years since the establishment of the specialty, Varna* : Publishing house "Science and Economy", 15 - 28.

Stoykova, T., 2018. Short history of the “Commodity science” department and “Commodity science” course degree : 70 years (1948 - 2018) (in Bulgarian). Varna : Publishing house "Science and Economy", 102.

Stoykova, T., Zlateva, D. & Pashova, S., 2020. Commodity science in modern market conditions (in Bulgarian). *Economic science, education and the real economy: development and interactions in the digital age. Jubilee International Scientific Conference dedicated to the 100th anniversary of the University of Economics – Varna*:Vol. IV, Varna : Publishing house "Science and Economy", 4, 427 – 439.

Vasilev, J. & Kehayova-Stoycheva, M., 2017. Sales analysis by the rectangle method. *Leonardo Electronic Journal of Practices and Technologies*, http://lejpt.academicdirect.org/A30/149_160.pdf, 2017, 30, 149-160.

Importance of the discipline Commodity Science in the education of students from different professional fields in Bulgaria

Radoslav RADEV[1]

[1] University of Economics - Varna, Bulgaria
r.radev@ue-varna.bg

Abstract. The aim of this research is to study the importance of the discipline "Commodity Science" in the education of students from different professional fields in Bulgaria. The study shows the importance of the discipline "Commodity Science", in its various forms for the training and development of specialists from different professional fields in some universities in Bulgaria in the field of economics, administration and management, tourism and food technology, for which it is particularly importance knowledge of goods. The main center of commodity science in Bulgaria is the Department of Commodity Science and the only specialty of its kind in the country "Commodities Science and Customs Activity" at the University of Economics Varna. The present study can be a good basis for researching the opinion of students studying the discipline "Commodity Science" about the value of acquired knowledge, skills and competencies and their actual application in practice. In addition, the opinion of various employers about the need for staff with knowledge of goods can be surveyed. The obtained results provide information that will be important for the future development and improvement of the scope of the discipline "Commodity Science".

Key words: Commodity science, the discipline "Commodity science", student training.