Kaunas University of Technology

School of Economics and Business

Implementation of Big Data and Big Data Analytics into the Audit in Lithuania

Master's Final Degree Project

Justina Mačiulskytė

Project author

Prof. Dr. Lina Dagilienė

Supervisor

Kaunas, 2020

Kaunas University of Technology

School of Economics and Business

Implementation of Big Data and Big Data Analytics into the Audit in Lithuania

Master's Final Degree Project

Accounting and Auditing (6211LX037)

Justina Mačiulskytė

Project author

Prof. Dr. Lina Dagilienė

Supervisor

Prof. Rūta Gokienė

Reviewer

Kaunas, 2020

Kaunas University of Technology

School of Economics and Business

Justina Mačiulskytė

Implementation of Big Data and Big Data Analytics into the Audit in Lithuania

Declaration of Academic Integrity

I confirm that the final project of mine, Justina Mačiulskytė, on the topic „Implementation of Big Data and Big Data Analytics into the Audit in Lithuania“ is written completely by myself; all the provided data and research results are correct and have been obtained honestly. None of the parts of this thesis have been plagiarised from any printed, Internet-based or otherwise recorded sources. All direct and indirect quotations from external resources are indicated in the list of references. No monetary funds (unless required by Law) have been paid to anyone for any contribution to this project.

I fully and completely understand that any discovery of any manifestations/case/facts of dishonesty inevitably results in me incurring a penalty according to the procedure(s) effective at Kaunas University of Technology.

(name and surname filled in by hand) (signature)

Justina, Mačiulskytė. Implementation of Big Data and Big Data Analytics into the Audit in Lithuania. Master's Final Degree Project / supervisor prof. dr. Lina Dagilienė; School of Economics and Business, Kaunas University of Technology.

Study field and area (study field group): Business and Public Management (Accounting).

Keywords: Big Data, Big Data Analytics, audit, business entities, implementation, audit market.

Kaunas, 2020. 68 pages.

Summary

This Final Project presents the empirical research of Big Data and Big Data Analytics implementation in Lithuanian audit market. Although, Big Data is widely used in various sectors, but this topic is quite new and insufficiently studied. Big Data and its Analytics have a huge impact to business internal processes and audit procedures, as auditors are closely connected with theirs clients.

Theoretical analysis of the research is classify to two parts which consist of Big Data and Big Data Analytics analysis and audit characteristics. Firstly, Big Data definition ant it components were determined. For deeper understanding, the theoretical aspects of Big Data processing techniques Hadoop MapReduce and Apache Spark were accomplished. Moreover, the main potential Big Data adoption issues and challenges were presented in the second Master Thesis part. The second part of the theoretical analysis determined audit definition and processes. Audit process consist of audit planning, audit procedures performing and preparation of audit opinion. According the last researches an lack of analyses about Big Data and Big Data adoption in Lithuanian audit market, the research was oriented to the analysis through auditors and business entities perspectives.

In order to analyse the Big Data and Big Data analytics implementation in Lithuanian audit market, the qualitative research was prepared. Qualitative research consists of the questionnaire which was answered by the audit specialists and respondents from business entities. There are 10 questions in the questionnaire which are classify to descriptive analysis, Big Data adoption to business analysis and Big Data Analytics analysis in audit field.

Big Data and Big Data Analytics value added and opportunities provided are evaluated in different perspectives by respondents by business entities and auditors. It could be concluded that auditors need to improve their development in technologies according the clients, their needs and systems used. As the Big Data implementation is just a matter of time, the availability to analyse it could be a huge advantage in competitive audit market.

Justina, Mačiulskytė. Didžiųjų duomenų ir Didžiųjų duomenų analitikos diegimas audito sektoriuje Lietuvoje. Magistro baigiamasis projektas / vadovė Prof. dr. Lina Dagilienė; Kauno technologijos universitetas, Ekonomikos ir verslo fakultetas.
Studijų kryptis ir sritis (studijų krypčių grupė): Verslas ir viešoji vadyba (Apskaita).

Reikšminiai žodžiai: didieji duomenys, didžiųjų duomenų analitika, auditas, verslo subjektai, įdiegimas, audito rinka.
Vilnius, 2020. 68 puslapiai.

Santrauka

Baigiamasis darbas pristato Didžiųjų duomenų ir Didžiųjų duomenų analitikos įdiegimo Lietuvos audito rinkoje empirinį tyrimą. Nors Didieji duomenys yra plačiai naudojami įvairiuose sektoriuose, tačiau ši tema vis dar yra gana nauja ir nepakankamai išnagrinėta. Didieji duomenys ir jų analizė daro didžiulį poveikį verslo vidaus procesams ir audito procedūroms, kadangi auditoriai yra glaudžiai susiję su savo klientais nei jų naudojamomis sistemomis.

Teorinis tyrimas suskirstytas į dvi dalis, kurias sudaro Didžiųjų duomenų ir Didžiųjų duomenų analitikos ir audito charakteristikos. Pirmiausia buvo nustatytas Didžiųjų duomenų apibrėžimas ir jį sudarantys komponentai. Norint geriau suprasti Didžiųjų duomenų svarbą, buvo atlikta teorinė jų apdorojimo metodų „MapReduce“ ir „Apache Spark“ analizė. Taip pat, teorinėje baigiamojo darbo dalyje buvo pristatyti pagrindiniai galimi Didžiųjų duomenų įdiegimo aspektai ir iššūkiai. Antroje teorinės analizė dalyje buvo paaiškintas audito sąvokos apibrėžimas ir audito procesai. Audito procesus sudaro audito planavimas, audito procedūrų atlikimas ir audito išvados parengimas. Paskutinių tyrimų duomenimis, trūksta analizės apie Didžiųjų duomenų ir Didžiųjų duomenų analitikos pritaikymą Lietuvos audito rinkoje, tyrimas buvo orientuotas į analizę per auditorių ir verslo subjektų perspektyvas.

Siekiant išanalizuoti Didžiųjų duomenų ir Didžiųjų duomenų analitikos pritaikymą Lietuvos audito rinkoje, buvo parengtas kokybinis tyrimas. Kokybinį tyrimą sudaro klausimynas, į kurį atsakė audito specialistai ir respondentai iš verslo subjektų. Klausimyne yra 10 klausimų, kurie skirstomi į aprašomąją analizę, Didžiųjų duomenų ir Didžiųjų duomenų analitikos pritaikymą verslui ir audito rinkai.

Auditoriai ir respondentai iš verslo subjektų Didžiųjų duomenų ir Didžiųjų duomenų analitikos pridėtinę vertę ir teikiamas galimybes vertina pagal skirtinga perspektyvas. Galima daryti išvadą, kad auditoriai turi tobulinti savo technologijas, atsižvelgiant į klientus, jų poreikius ir naudojamas sistemas. Kadangi Didžiųjų duomenų įdiegimas yra tik laiko klausimas, galimybė šiuos duomenis analizuoti tai gali būti didžiulis pranašumas konkurencingoje audito rinkoje.

Contents

List of figures

List of tables

Introduction

Relevance of the topic. Big Data is characterized as large and complex data which is difficult to analyze and work using standard statistical software (Snijders, Matzat and Repis, 2012). Big Data cause a globally revolution in the 21st century. Big Data is adopted to variuos industries as helathcare, banking and insurance. One of the examples of Big Data in healthcare industry is nanomedicine, sensors and Artificial Intelligence allowance to collect data in real time (EY, 2015). Moreover, Big Data could be used in banking sectors as the tool for analysis of financial information of creditors as liquidity, as the main consequence is that loans are given to the most accurate creditors (Hassani, Huang and Silva, 2017). As insurance companies do not produce physical products, data is the most important asset to insurance industry. Companies collect financial data about its clients, actuarial data, claims data, rinks data and consumer data (Corbett, Schroeck and Schroeck , 2013). The definiotion of 'Big Data' associate with massive amount of data, new technologies, data archiving and data sources (Mukherjee and Shaw, 2016).

As business entities are audit companies' clients, Big Data revolution also make an impact to audit field. Audit is the inspection of financial statements in order to provide independent auditor opinion regarding the Company's financial statement (Ater, Gimbar, Jenkins, Saucedo and Wright, 2018). External auditors are responsible not only for financial statements testing, misstatements identification, reporting improvements and audit process. The expectations of the auditors are higher than before what means that audit professionals need to continue its process improvement. Companies invests into Big Data analytics for the development of the decision-making process and they look forward for auditors with capacities required for Big Data analysis in order to improve the effectiveness and efficiency of whole audit process (Early, 2015).

Problem. Big Data integration in the business systems could be divided into the main four phases: data generation, data acquisition, data storage and data analytics (Padgavankar and Gupta, 2014). Data analytics is the most important phase for auditors. Phases, main challenges and issues regarding Big Data implementation are widely analysed, while the auditors position in growing Big Data and Big Data analytics market is unstudied. Auditors has a significant position in technology development as Big Data implemented companies need to be audited. Expectation for auditors has increased, the clients required to audit companies to be ready for Big Data Analysis. Therefore, this master thesis will be based on analysis of the implementation level of Big Data and Big Data Analytics in Lithuanian audit market.

The main aim. To explore theoretical assumptions and opportunities of implementation Big Data and Big Data Analytics in Lithuanian audit market.

Object. Big Data and Big Data Analytics adoption in Lithuanian audit market.

Tasks:

1. To explore value creation and opportunities of Big Data and Big Data Analytics.
2. To represent theoretical concept of audit framework value creation by using Big Data Analytics in audit process.
3. To create research methodology for exploring opportunities of implementation of Big Data and Big Data Analytics in Lithuanian audit market.
4. To present opinions of auditors and business representatives regarding adoption of Big Data and Big Data Analytics in Lithuanian audit market.

The methods of the research. Systematic scientific and professional literature review. For evaluation of Big Data and Big Data Analytics implementation in Lithuanian market the qualitative research - structured expert interview was selected. For questionnaire, 10 question were prepared regarding the topic.

1. Big Data opportunities and challenges in auditing sector

Big Data and Big Data Analytics is not a new concept in the business field in the worldwide perspective. Contrariwise, the Big Data and Big Data Analytics adoption to various sectors greatly increased in the last few years. Therefore, the first part of this section will present the main aspects of its implementation in general perspective and particular industries as banking, healthcare and insurance. The second part of this section is more oriented to audit market and will present the Big Data and Big Data Analytics impact to audit companies which are closely dependent to its clients. As audit services are provided to entities which are potential users of Big Data and Big Data analytics, audit companies need to adopt new technologies to used techniques and procedures. The last part of this section will provide short review of Big Data and Big Data Analytics development in Lithuanian market.

1.1. Big Data and Big Data analytics revolution in business field

Big Data revolution expanded the whole world in the 21st century. While professionals are adopting it to different industries as healthcare, banking and insurance to help achieve the entities goals, sceptics rose the question of Big Data value to the different industries. Once investigated, that Big Data can detect diseases earlier and prevent complications in healthcare industry. Moreover, healthcare institutions collect and manage electronic records regarding patient history, previous choices, diagnosis and other. Obviously, Big Data is the beneficial tool to improve the management of the mentioned data (Fatt and Ramadas, 2018). Also, one of the examples of Big Data in healthcare industry is nanomedicine, sensors and Artificial Intelligence allow to collect data in real time. In the terms of diagnosis, it could be disclosed before the symptoms are reveal. On the other hand, in order to implement new technologies into the daily work, it could take lots of time and investitions for the result (EY, 2019).

As I mentioned above, Big Data helps to analyse the financial information of creditors as liquidity in banking sector, as the main consequence is that loans are given to the most accurate creditors. It should be noted that Big Data is also able to detect potential financial fraud risk and analyze and collect information about the bank clients and creditors (Hassani, Huang and Silva, 2017).

Big Data is the main tool to insurance companies. Its Companies does not produce physical products as manufacturing companies to sell or buy, for this reason data is the most important asset to insurance industry. Companies collect financial data about its clients, actuarial data, claims data, rinks data and consumer data. In general, insurers have large consumer and commercial consumer with large data populations (Corbett, Schroeck and Schroeck , 2013).

The term 'Big Data' associate with massive amount of data, new technologies, data archiving and data sources (Mukherjee and Shaw, 2016) The main three dimensions of challenges in data management area Volume, Variety, and Velocity. Volume refers to the large quantity of data. It should be noted that volume is dependent from size and type of data factors. What may be considered as Big Data now, may not be assume as Big Data in the future, because data capacities are growing and are able to proceed even bigger data packages. In addition, definitions of Big Data are relative to industry and type of the data. Variety is related to the structural level of data. New technologies gave an opportunity to use various types of structured, semi-structured, and unstructured data. Structured data is numerical information found in relational databases. Unstructured data refers to the visual information as text, images, audio or video information, the

main problem for mentioned visual data is its complicated analysis. Velocity consist of level at which data are gathered and speed at which should be analysed. The mentioned three V dimensions described the main characteristics of Big Data and its opportunities how to analyse it. Volume, Variety, and Velocity used together caused the deep insights regarding the processes of the entity and its position (Gandomi and Haiden, 2015).

Big Data has a huge potential and demand in the worldwide market. According to Statista, Worldwide Big Data market revenues for software and services is expected to grow from $42B in 2018 to $103B in 2027. Also, Compound Annual Growth Rate is forecasted to increase by 11.4% from $35B to $103B in period 2017 – 2027 (Statista, 2019).

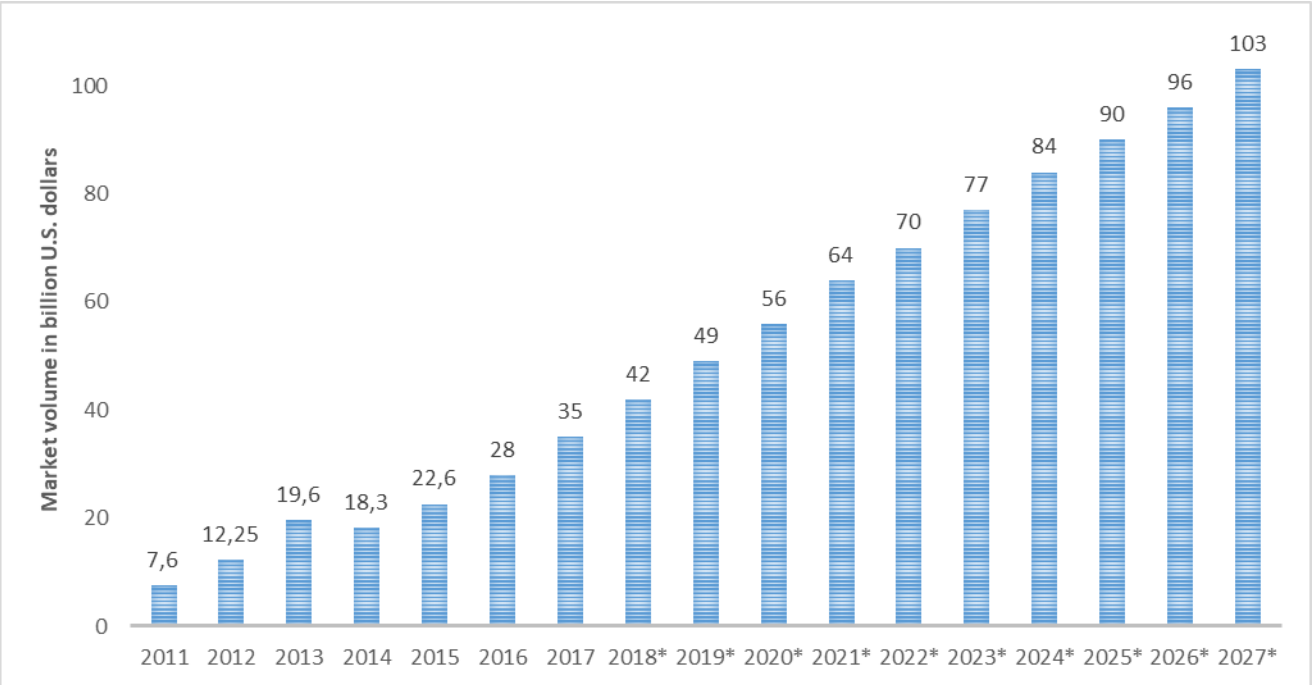

Figure 1. Big Data market size revenues forecast worldwide from 2011 to 2017 (in billion U.S. dollars) (Statista, 2019)

The importance of Big Data is the result of high demand of advanced IT professionals, data analytics field involved people with high potential, motivation and specific abilities. Data science help to investigate the information and achieve deeper insights. It should be noted that Big data analytic revolution caused the competition of various skilled professional focused not only into IT science, but also into finance, management fields. Stability and traditions are no longer the main tools of success in business. Nowadays, a business must concentrate on innovations and new technologies in order to follow the growing pace. The revolution of Big Data is challenging the business and will continue challenge the whole industries. Big Data and Big Data Analytics are the main tools which should be used for organizations to build their business strategies to follow ability and innovations into the constantly improving world. This progressive grow in the role of Big Data is discussed to be "one of the most important features of the contemporary economy and society" (Davenport and Harris, 2017).

The wide impact of Big Data on the business area included technical, managerial and social aspects. They consist of the idea how technologies should be chosen, how to limit the scope of the analysis, which should be enough for decision making, how to control a privacy and security issues. In terms

of accounting, Big Data essentially changes the understanding about accounting data and how it should be reported (Griffin and Wright, 2015).

Big Data Analysis has become one of the most important investment in various sectors as Banking and Securities, Communications, Media and Entertainment, Healthcare Providers, Retail and Wholesale trade and Insurance. It is noteworthy, that the auditing and accounting service providers are proceed with various large amount of data. Accordingly, auditing and accounting companies adopting Big Data projects to its own casual daily work.

1.2. Big Data analytics investigation into audit companies

Audit is the inspection of financial statements in order to provide independent auditor opinion regarding the Company's financial statement. In general, audit team plans audit, identified significant accounts which are the most important into Company's activities and proceed procedures in order to test provided accounting records (Ater, Gimbar, Jenkins, Saucedo and Wright, 2018). The test of records consists of classification, completeness, period, occurrence, existence and valuation tests. In the end, the Auditor sign the auditor opinion and agree with Financial Statement issued by the Company (The Institute of Internal Auditors, 2017). Auditors are regulated by local institutions, in terms of Lithuania, auditors are control by Lithuanian Chamber of Auditors.

According to ISA 250 (Revised), Consideration of Laws and Regulations in an Audit of Financial Statements the objectives of the auditor are receive enough audit evidence according, perform audit procedures to identify misstatements and to respond identified or suspected non-compliance with law and regulations (ISA 250).

External Auditors are responsible not only for financial statements testing, misstatements identification, reporting improvements and audit process. The expectations of the auditors are higher than before what means that audit professionals need to continue its process improvement. Companies invests into Big Data analytics for the development of the decision-making process and they look forward for auditors with capacities required for Big Data analysis in order to improve the effectiveness and efficiency of whole audit process (Early, 2015). On the other hand, Big Data Analytics is one of the tool to detect financial fraud (Tang and Karim, 2018).

Also, KPMG and Forbes Insights performed a survey of CEOs and CFOs and other financial leaders about their opinion on the future audit regarding its quality and overall process. Majority of respondents have an opinion that auditors should use biggest samples and should use more complex technologies for data collection and analysis. Also, almost half believed that auditors should accomplish a deeper analysis in areas they are covered now. In conclusion, auditors need to be more progressive and dynamic in data analysis, risk identification and providing of insights (KPMG, 2017).

The word is changing at digital speed and became more innovative and modern, except the accounting and auditing specialists which do not accept it and continuously use old-fashioned assurance methods. The planned data ecosystem will be composed of a huge chain of interconnected data sources and lots of continuously working intelligent agents for audit procedures performing and drawing exceptions in terms of continuous audit forms. It is appropriate to assume that business assurance and reporting will became more widely information set involved partially

what is described today as non - GAAP (GAAP standing for Generally accepted Accounting Principles) measures. Majority of these measures will have to be ensured in a continuously way (European Court of auditors, 2020).

According the expectation of external auditors, more and more IT professionals are included into the team. Financial auditors focus on relevant IT systems in order to secure sensitive financial information which support their work. However, the increasing number of cybercrime activities and IT systems with lack of appropriate controls makes the audit process more difficult. If the company cannot rely on the system, then financial auditors must perform substantive testing and examine much more documents what results in ineffective audit work (Barta, 2018).

Currently, audit companies and auditors have a plenty of assignment which motivate them to implement innovative and smart tolls into the daily work. Notwithstanding, majority of audit companies have an understanding about digital tools implementation into audit procedures and whole process, but 70% of audit companies are only in the initial and primary phases by applying big data and analytics to their auditing procedures (Deloitte, 2018)

As I mentioned above, only audit market leaders are able to investigate accomplish Big Data into its services. Big 4 companies - KPMG, Ernst & Young (EY), Deloitte and PricewaterhouseCoopers (PwC) are the strongest companies into accounting and audit sector and could be called as this sector leaders. Because of size, reputation and market share these mentioned four companies dominate the field. In terms of worldwide perspective, Big 4 provides accounting, audit, tax and financial consultation services to the biggest companies in various sectors: banking, insurance, transportation, manufacturing, IT and healthcare (Statista, 2019).

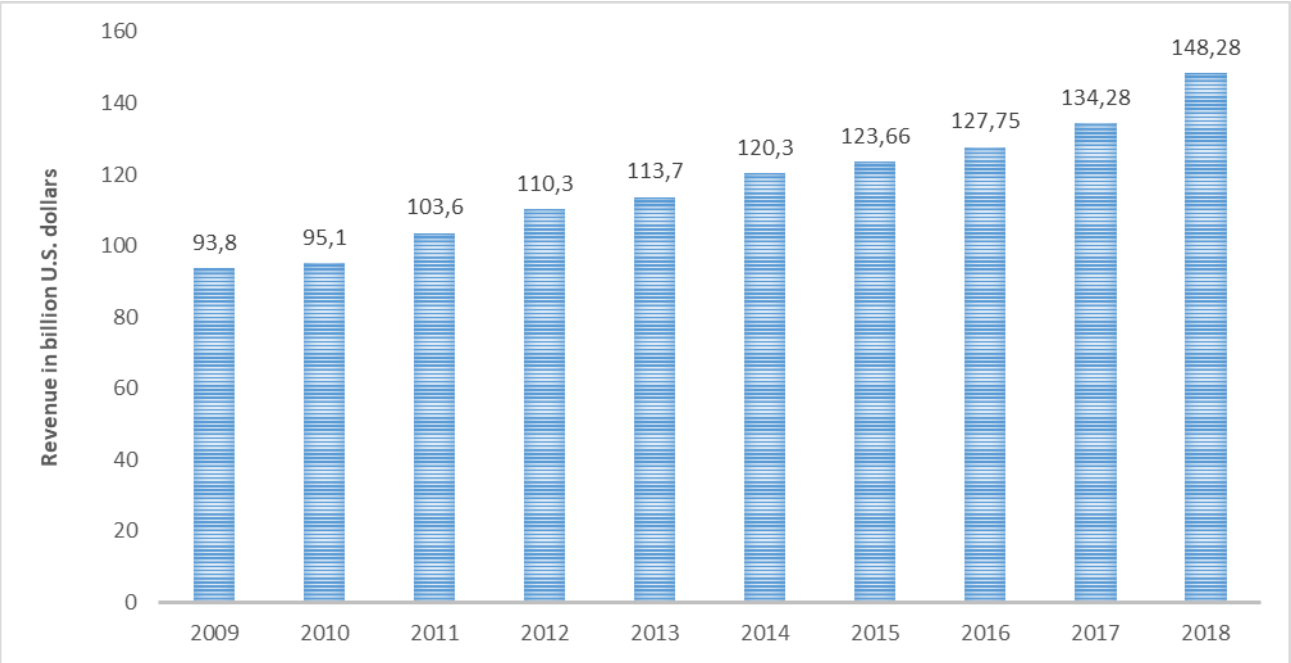

Figure 2. Combined revenue of the Big Four accounting / audit firms worldwide from 2009 to 2018 (in billion U.S. dollars) (Statista, 2019)

As we can see from the Figure 2, Big four worldwide revenue is growing in each year from 2008 till 2018, when reached 148.28 Eur in 2018. It is important to add that the mentioned companies are

searching new clients, invest into development of the services and have and ambition to grow in the future as well (Statista, 2019).

In order to understand the value of Big Data, it is important to identify the main advantages related to Big Data adopted into audit engagements. Majority of the Big Data can be analyzed as leading indicators because accounting transactions and records could be presented from one day till the more than one year. Some of accounting transactions are closely related to accounting transactions in the future as future revenues or costs. Big Data usage is a powerful tool of auditors for misstatements findings and field identification which needs corrections. In mostly cases, Big Data physically and conceptually contents are separated from accounting data. For this reason, it is more complicated to manipulate date and hide financial fraud (Early, 2015).

Big Data adoption requires new, modern and innovative technology in order to obtain relevant business insights that reveal findings related to entities' risk, profit, performance and productivity management. Accounting and external audit companies need to focused on the main aspects before the start of use Big Data. Firstly, the Companies should consider about its technology resources to capitalize on big data and analytics. Moreover, human resources should be prepared, specific competencies and high-level technical skills are necessary for appropriately leverage big data and analytics. Then, external audit companies should analyze profit, opportunities which could appear using Big Data and the changes which should be made to its strategy. In general, data capture is often the main obstacle in the big data and analytics process, therefore, improvements for IT functions are required. The last challenge for external audit companies is data protection. In the last several years, cybercrimes has increased as business used insecure data systems. As audit companies works with especially sensitive information, the investment to data protection is mandatory. For his reason, Big Data implementation to accounting and audit companies is a huge investment and achievable only to accounting and audit leaders which generates enough profit and has enough amount of the capital for audit services developments and improvements (EY, 2017).

In current audit process, the based data is related to general ledger data, but involved big data can help to obtain information of sub-ledger such as revenue or procurement-cycle data, for key business processes. It increases the volume of data for deeper analysis and insights (EY, 2015).

The value and benefits of Big Data usage into the business and audit environmental is clear. The main question is how to prepare employees and management system for new technologies and Big Data? Preparation of employees to work with Big Data should start from the first level of professionals' career. In terms of potential employees, students should be prepared for long term – career demands at University. In addition, the emphasis on first job skills should be decreased. Also, the understanding of the role value should be pointed as one of the most important subjects. The ability to ask questions can improve the quality of its knowledge and the value of data analysis. Understanding the difficulties of effective analysis, limits of analysis, value of curiosity, consideration of data sources, approaches to the analysis and evaluation the costs and benefits of the analysis are the most valuable characteristics of potential employees which could work with new technologies into the system (McKinney, Yoos and Snead, 2017).

It is important to add that Big data Analytics implementation in internal audit is also relevant process. The main challenges which meets Big Data transforming into the whole process are audit on data with different formats, audit on asynchronous data, audit on conflicting data, audit on

illegally tampered data, audit on incomplete data, audit on aggregated data, search encrypted data, audit on encrypted data (Zhang, Yang and Appelbaum, 2015).

Big Data and Big Data Analytics revolution had an influence on the whole industries, auditing is not an exception. As the expectations of audit and auditors are growing, audit leaders invest to Big Data implementation into its activities and daily work. The mentioned modern technologies provide lots of opportunities to use and analyze various data and perform deeper analysis about companies profit, management and cost system. On the other hand, the investments into Big Data required more capital and human resources and other internal modification.

1.3. Review of Big Data analytics in Lithuanian

The review of the Big Data and Big Data analytics in the worldwide perspective and audit fields have been done above in the previous problem analysis part. It important to add that new technologies and innovation are relevant aspect to Lithuania as well. First of all, the work demand requires specialists with particular skills to implement and analyze Big data into the companies. For this reason, the new and more innovative study programs have been launched into the Lithuanian universities. For instance, universities suggest particular study programs or study modules for student integration into the Big Data Analytics.

Lithuanian companies which have a Big Data in its budget plans, face the same problems as international companies around the whole world, the mentioned problems are related to data management, financial and personnel issues. It is important to mentioned, that the amount of Big Data Analytics consultation companies is increasing in Lithuanian market. For instance, various Lithuanian and International Companies create solutions for Big Data analysis, design and implement date warehousing and management systems. In Lithuania, Big Data analytics solution providers are Exadel, Baltic Amadeus, Zenitech, Inntec, Afecto and Nortal. Moreover, the mentioned companies provides solutions for financial, healthcare, retail, construction industries and others.

In Lithuania, Big Data and Big Data Analytic is quite unexplored and unstudied topic. Some of Lithuanian Companies (e.g. Financial institutions) which owns high volume and complex data already use or planning to use Big Data and Big Data Analytics in the future periods. Big Data implementation require the measured and accurate planning related to financial budget, implementation process and preparation of personnel. In terms of audit, the leaders of audit, accounting and tax services are Big 4 companies in Lithuania. The mentioned companies reached the highest part of market share and provides services to the Lithuanian business leaders.

In order to research the Big Data and Big Data analytics usage challenges into the Lithuanian audit market, the analysis should be accomplished in terms of three different perspectives: audit companies, business clients and audit regulators. The potential to use Big Data in audit companies consist of company's size, market situation, financial resources, local and international regulation system, employee's education, strategy and values of the company. It should be noted that the result of Big Data implementation into the process of audit is high quality of provided services, deeper insights during the provided whole audit procedures and efficiency internal control of the client's records and other processes. Moreover, similar analysis should be proceeded for audit clients and regulator institutions as well (Dagilienė and Klovienė, 2019).

After the resources analyzed, it could be noted that the Big Data and Big Data Analysis implementation became an important and sometimes mandatory step into the business. As audit field is closely related to its clients, any changes are make an impact to audit companies and the whole audit process as well. The main reason, challenges and issues for Big Data and Big Data Analytics implementation for Lithuanian business and audit companies will be analyzed in the following parts of the Master thesis.

2. Theoretical analysis of Big Data in auditing field

2.1. Theoretical analysis of Big Data

The second section of Master Thesis consist of theoretical analysis of Big Data into the business. In order to understand Big Data impact to Company's and its activities, the review of its definition was performed. It is important to add, that recourses are full of various Big Data definitions which are listed and shortly explained in the section below. Big Data definition is associated with high of volume data, huge data sets and complexity of usage. Then, the analysis of the main Big Data processing techniques was accomplished, as the main techniques are Hadoop MapReduce and Apache Spark, it was performed its analysis and comparison. The mentioned techniques were compared based on Processing Method, Time Efficiency, Error Recovery and Memory requirements aspects. After the comparison of Hadoop MapReduce and Apache Spark methods, one of them was suggested as the most suitable alternative for the Big Data application and processing.

The Big Data researchers (Padgavankar and Gupta, 2014) defined that Big Data could be divided into the main four phases: data generation, data acquisition, data storage and data analysis, the mentioned phases of Big Data are analyzed in the theoretical part as well. Each phase is defined and widely described in the section below. Also, the main Big Data analysis and its classifications is investigate and compared to traditional analysis. Before the Big Data implementation, it is important to identify the main issues which Company faces, as data management, privacy and security, data access and sharing of information, storage and tools and talent issues. In order to be prepared for Big Data usage, it is mandatory to know how to deal with the issues mentioned and solve the main problems.

2.1.1. Description of Big Data

Nowadays Big Data is widely analyzed in literature, but the definition of this term is rather new. In order to analyses the impact of Big Data, we need to understand the term and meaning of Big Data. In the first part of the second section I will describe different Big Data definitions explained by the different resources. The definition of the mentioned term was virtually unknown in 2010, but widely discussed later, in 2011, when it became widely touted as the latest trend. The importance of Big Data is evident as it has been mentioned in Gartner's Top 10 Strategic Technology Trends for 2013 (Savitz, 2012) and Top 10 Critical Tech Trends for the Next Five Years (Savitz, 2012).

In some cases, Revolution of Big Data is compared to cloud computing trend. The mentioned term is well known by everyone, from product vendors to large-scale outsourcing and cloud service providers. Big Data definitions have developed rapidly, when the usage of Big Data increased into various sectors (Fujitsu, 2012). The insights were found after the analysis of few different definitions of Big Data listed below:

— **Doug Laney** detailed data management as three Vs: Volume, Velocity, and Variety. This is the most respectable and well-known definition of Big Data which have been explained in 2001 (Laney, 2001).

— **A National Institute of Standards and Technology** report explained Big Data as contained from "extensive datasets - primarily in the characteristics of volume, velocity, and/or variability - that require a scalable architecture for efficient storage, manipulation, and analysis". In terms

of Big Data, it means the value of data which exceeds a petabyte – one million gigabytes (National Institute of Standards and Technology, 2015).

— **Cambridge Dictionary** described Big Data as "very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods" (Cambridge University).

— **According the CSA Big Data Working group,** the term Big Data is related to the massive amounts of digital information companies and governments which collect data about human beings and its environment (CSA Big Data working Group, 2013)

— **Data Scientists** Big Data characterized as large and complex data which became difficult to analyze and work using standard statistical software. Therefore, large and complex data required new and modern software to analyze it (Snijders, Matzat and Repis, 2012). Moreover, the size of Big Data could be called as moving target. The size of date increased from 2012 from a few dozen terabytes to many petabytes of data. Big Data is the complex of techniques and technologies that are necessary for integration and analysis for large datasets which are massive, complex and complicated for ordinary technologies (Bajaj and Ramteke, 2014).

According the mentioned citations from various articles and books, Big Data could be connected to large, massive and complex data which requires specific and modern software. In order to analyze Big Data, companies have to improve its system and its database. As we mentioned above, Doug Laney was the first who identified the main three Big Data management dimensions as volume, velocity and variety (the 3Vs) (Laney, 2001).

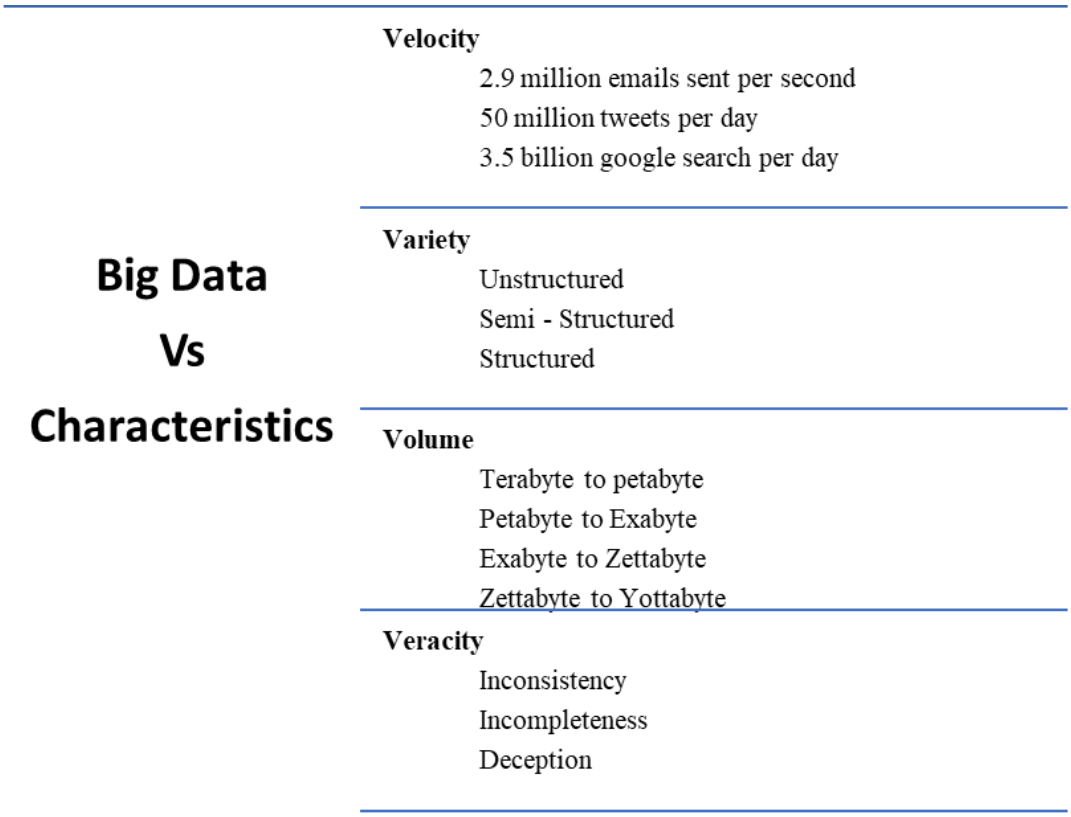

Figure 3. Four Vs of Big Data (Laney, 2001).

Since then, the list of dimensions have been expanded and other qualities as exhaustivity, fine-grained and uniquely indexical, relationality, extensionality and scalability, veracity, value and variability have been added to Big Data characteristitcs. Analysts as Gartner Characteristics of Big Data described as large Volume, fast Velocity and diverse Varierty, the mentioned combination I also called as 3Vs (Beyer and Laney, 2012). In order to gain the better knowledge of the mentioned three Vs Big Data demensions, the widely explanations and descriptions are listed below.

Volume. Volume is the characteristic which is the most related to Big Data. In general, the word Big in Big Data perspective determine the volume. In 2013, Big Data specialists predicted that according the rapidly growing of data, in nearby future, it supposed to be collected in zettabytes instead of petabytes. For instance, the social networking sites are available by themselves, when its producing data in order of terabyte. Moreover, social networking sites data reach high amount of data which is difficult to be analyzed and used by the traditional systems (Katal, Wazid and Goudar, 2013). The volume of data is growing constantly, 90% of worldwide data ever created was generated in the past 2 years (McNulty, 2014).

Velocity. In Big Data perspective, velocity is an intention that deals with the data speed which is coming from various data sources. The mentioned characteristic is not limited only to the speed of incoming data but speed at which the data flows too. For instance, the data would be continually moving from the sensor devices to the database store and this amount of data would not be called small. Therefore, traditional systems are not available to perform the analytics on the data that is continually in movement (Katal et al., 2013).

Variety. Variability of data is related to inconsistencies of the data flow. Data loads become complicated to be maintained especially with data used in social media which increased rapidly. The data variety consists of six characteristics: formats, structure, unstructured data, natural language, media, complex systems (Katal et al., 2013).

Veracity. Big Data veracity is recognized as one of the most important property for its utilization, which supplement the three previous Big Data dimensions as volume, variety and velocity. One of the reason of Veracity being in 3V, that Big Data could not be compared to a traditional data which is collected. For this reason, new processes and tools are required to use. Veracity is a data capability to make a decisions regarding the appropriate, useful and high quality data (Kaisler, Armour, Espinosa, Kaisler and Armour, 2013).

Volume, velocity, variety and veracity - three dimensions that consider the main characteristics of Big Data value and complex. On the other hand, the complexity and description of Big Data are widely than these three concepts, there are another factors and processes which are significant in Big Data comprehension. Additional dimensions are listed and defined below.

Traditional business intelligence. This characteristic contain of wide category of technologies and applications for collection, storing, analyzing, and providing access to data. Business intelligence provides relevant information which is useful for companies to make better decisions using fact based support systems. Moreover, business intelligence operate by using detailed business data collected from databases, applications and other tangible data sources. It should be noted, that business intelligence provide historical data which can be used to forecast future processes (Ohlhorst, 2013).

Data mining. This is a method of analysis when companies analyze its own data from different perspectives and then make data summary that are deemed useful. Normally, data mining is used with data at rest or with archival data. Moreover, this method techniques are oriented to data modeling and knowledge discovery to forecast processes. The main purpose of data mining is to discover new patterns from high amount of data (Ohlhorst, 2013).

Statistical applications. Statistical methods give the insights of data using algorithms with reference to statistical principles and generally focuses on data collection related to polls, census, and other static data sets. Statistical applications are able to provide sample observations which can be used to analyze populated data sets for the intension of estimating, testing, and predictive analysis. Empirical data, like surveys and experimental reporting, are the main sources for information to be analyzed (Ohlhorst, 2013).

Predictive analysis. This is a subgroup of statistical applications where data groups are analyzed to present predictions according to trends and other information collected from databases. Normally, predictive analysis is a significant tool into financial perspective where trends are used to forecast of processes. The main purpose of predictive analysis is to discover the risks and opportunities for business process, markets, and manufacturing (Ohlhorst, 2013).

According the theoretical analysis above, Big Data consists of massive quantities of data, executed in real-time which could be structured, semi-structured and unstructured. In general, Big Data is used for various range of analysis, where the result could be beneficial for management. Big Data is used in order to make better decisions, discover new patterns of activities, predict future trends and reveal the main challenges and opportunities for business process, markets and manufacturing. In general, Big Data is associated to Volume, Velocity and Variety definitions, the mentioned concepts are widely as the main for Big Data identification.

2.1.2. Big Data processing techniques

Big Data processing requires a strong preparation of whole organization, lots of resources and time. In order to process the Big Data in the organization structure, the main two methods are recommended: batch-based stored data processing and real-time data-stream processing. Hadoop MapReduce and Apache Spark are the most promising source methods which are analyzed and compared in the section below.

Hadoop MapReduce. Hadoop is abundantly accessible framework and has been used by data analysts, researchers and other specialists for more than eight years for Big Data processing process (The Apache Software Foundation Blog, 2014). Hadoop MapReduce is a great alternative for processing data which one of the requirements is one – pass computation, which means that the whole inputs should be read once. Otherwise, in multi – pass computation the data could be read more than one time, and this is not suitable for this method. Hadoop MapReduce was created by Google with two features to process large datasets, which generate high amount of various data. The mentioned two features are Map and Reduce. In general, the Map is operated to calculate the value pairs for the imputes while Reduce feature combines the results of map function into a scalar. To use the whole possible Hadoop MapReduce tools, it is necessary to convert inputs into MapReduce form. Also, it should be noted that this system is able to plan, monitor tasks and execute the failed tasks (Gu and Li, 2013).

Moreover, the MapReduce system, after each step and before the next step begins, each output of data is stored into Distributed File System (DFS), the result of mentioned process is slowly processing speed. Also, it works with the large number of clusters that are more complicated to mane and complex. On the other hand, integration of more than one method is necessary in majority of cases of big data processing. In order to perform a set of complicated tasks, the Big Data specialists should use MapReduce tools to execute them. MapReduce is used for high speed processes which need to be performed in some specific sequence (Gu and Li, 2013).

Apache Spark. As MapReduce, Apache Spark is a cluster computing system as well, which also have language – integrated application programming interface and parallel operators (Apache Storm, 2015). Apache Spark is more beneficial compared to Hadoop MapReduce and other widely used technologies. This system provides a united way how to manage processing requirements related to Big Data and how to improve the speed of applications cluster to make it faster. Hereinafter, it helps in writing applications (in scale, java or python language) with more than 75 high level operators. Moreover, Apache Spark operates SQL queries, streaming data, machine learning and data based on graph. Direct Cyclic Graphs (DCG) is widely used in Apache Spark to perform complex, multi-step data pipelines and support in-memory sharing among different jobs (Wani and Jabin, 2018).

	Hadoop MapReduce	**Apache Spark**
Processing Method	Each task or data output saved in DFS.	Tasks saved in disk after completion.
Time Efficiency	Minutes – hours.	Minutes – seconds.
Error Recovery	Uses concept of 'replication'.	Resilient distributed datasets (RDD)
Memory requirements	No issues.	More RAM instead of network and disk I/O.

Table 1. Key differences between Hadoop MapReduce and Apache Spark (Wani and Jabin, 2018)

In order to compare the Hadoop MapReduce and Apache Spark systems, the four aspects have been compared: processing method, time efficiency, error recovery and memory requirements. Each step and task performed by Hadoop MapReduce are saved into DFS, while the whole performed tasks performed by Apache Spark are saved into disks after the completion of processing. Hadoop MapReduce is a parallel data processing way to operate long running jobs that take minutes or hours to complete. On the other hand, Apache Spark is created to process real-time stream-data and SQL queries that need only few seconds to complete. In terms of error recovery, Hadoop MapReduce uses the concept of 'replication', while Apache Spark uses different data storage systems and resilient distributed datasets which are enabled to store data in servers and restore it automatically in case of failure. Hadoop MapReduce do not have any memory issues, but Apache Sparks has more memory requirements. Also, it uses more RAM, iinstead of network and disk I/O. In Conclusion, Apache Spark could be suggested as the most suitable alternative for the Big Data application and processing.

2.1.3. Phases of Big Data system

In outline, Big Data could be divided into the main three phases: data generation, data acquisition and data storage. For better understanding, data could be analyzed as a raw material, data generation and acquisitions is maintenance process, data storage is related to the possession, where data is kept. In order to understand the process, the analysis of the main phases is described below.

Data generation. In General, data generation is called as the first step of Big Data. For instance, Internet data is generated of billions of searching entries, various internet forum posts, chatting records through various channels and microblog posts and messages. The mentioned data is directly related to Internet users' life, in this case, people daily life, and have similar attributes of high value and low density. It should be noted that Internet data individually is valueless and insignificant, but when this data is used through the exploitation of accumulated Big Data, information as habits and hobbies of Internet users, become useful. The mentioned data could be used assumptions and forecast of possible customer's behavior or expectations. Moreover, when data is generated through longitudinal and/or distributed data sources, datasets become more large-scale, highly diverse, and complex. Therefore, this data sources involve data sources as sensors, videos, click streams, and all other available data sources. Nowadays, the main Big Data sources are the operation and trading information systems in organizations, logistic and sensing information in the Internet of Things, human recourses reaction and others. It is important to add, that the information exceeds the size and capacities of IT architectures and infrastructures of existing enterprises (Padgavankar and Gupta, 2014).

Data acquisition. As it was mentioned above, the second part of the Big Data system is data acquisition which includes data collection, data transmission and data pre-processing. In data acquisition process, when raw data is collected, the effective transmission mechanism should be recovered in order to be able to send it to a particular storage management program to support totally different analytical applications. When data is collected, it also could include some unnecessary and useless data, which uselessly increases storage expansion and make an impact to data analysis. For instance, high amount of unnecessary data is gathered by sensors for environment monitoring. One of the solutions for this problem could be data compression technology, which could be applied to decrease amount of redundant data. It is important to add, that pre-processing operations are mandatory to guarantee efficient data storage and exploitation (Lyko, Nitzschke and Ngomo, 2016).

Big Data Storage. As auxiliary tool of server, data storage device is used to store, manage and analyze data. The process of data storage became more important and relevant because of growth of data – Information technology companies offer huge capacity of storage in order to be more attractive and competitive. Therefore, it is a necessary to perform research on data storage and its types. Various storage systems meet the demand of massive data requirements. For instance, current massive storage technologies could be group into two groups - as Direct Attached Storage (DAS) and network storage. Also, network storage could be group into Network Attached Storage (NAS) and Storage Area Network (SAN). The classification of massive data storage is represent in Figure 4 (Padgavankar and Gupta, 2014).

In DAS, hard disks are strictly related to servers, which takes a particular amount of I/O resource which is managed by an individual application software. Therefore, DAS is available only to interconnect servers witch have a small scale. On the other hand, because of low scalability, DAS will display inappropriate efficiency when the storage capacity is increased, for instance, the upgradeability and expandability are greatly limited. In general, DAS is more used in personal computers and small-sized servers (Padgavankar and Gupta, 2014).

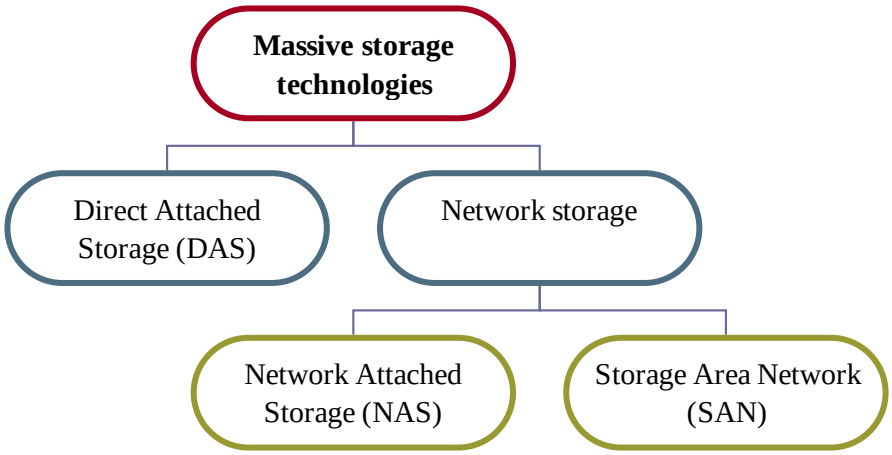

Figure 4. Classification of massive data storage (Padgavankar and Gupta, 2014).

Network storage is used to absorb network to provide it users with a union connection for data access and availability to share. It should be noted that network storage equipment consist of special data exchange equipment, disk array, tap library, and other storage media, likewise special storage software. Network storage equipment is characterized as strong expandability.

As we can see from the Figure 3, Network Storage consist of NAS and SAN. In fact, NAS is a supplementary storage equipment of a network. The mentioned equipment is directly related to a network through a hub or switch through TCP/IP protocols. Using NAS, data is actually transmitted in the form of files. In some cases, NAS is called as network-oriented, while SAN is normally designed for data storage with a scalable and intensive network. In terms of SAN, data storage management is connected to independency through a storage local area network.

From the organization perspective, in terms of a data storage system, DAS, NAS, and SAN could be divided into three parts:

- disc array, which is the base of a storage system and the fundamental guarantee for data storage.
- connection and network subsystems, which provide relation around one or more disc arrays and servers;
- storage management software, which handles data sharing, disaster recovery (Padgavankar and Gupta, 2014).

In order to adopt Big Data into the entities processes, each described phase has to be prepared. The phases could be done by the personnel of the entity or external services providers. For this reason, the business entities should pay more attention to the IT specialists competence and acquired work experience.

2.1.4. Big Data Analytics phase

As it was mentioned above, Big data is divided into main phases: data generation, data acquisition and data storage. Big Data Analytics could be called as the forth phase as it is a process which use the raw material, in our case data, to create a new value. Moreover, data analysis could be classify to traditional and Big Data Analysis. The general aspects of Big Data Analysis and both types are reviewed in the section below.

Data analysis. In principle, Big Data analysis includes analytical methods for traditional data and Big data, analytical architecture for Big data, and software used for mining and analysis. Data analysis is the last and the most significant part in the value phases of Big data, which the main aim is extract useful values and provide suggestions or decisions for organization activities. Analysis of data in different areas could be generated by different levels of potential values. Notwithstanding, data analysis is a vast filed, which constantly changes and is extremely complex. Methods, architectures and tools for big data analysis will be introduced in this section below (Agresti, 2018).

In the traditional data analysis proper statistical methods are used to analyze massive data, to focus, extract, and improve useful data which is hidden in a group of messy datasets, and to determine the inherent law of the analyzed data, for this reason the value of data is maximize. Data analysis has an important role for organization‘s analysis regarding the development plans for the future strategies, understanding customer needs ant its demand, predict market trends. Big data analysis can be consider as the analysis technique for a special kind of data (Padgavankar and Gupta, 2014).

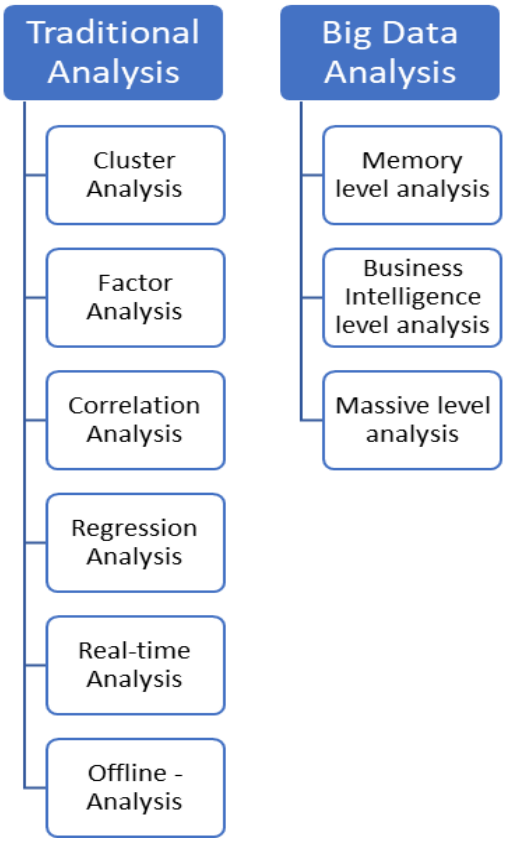

Figure 5. Classification of data analysis (Padgavankar and Gupta, 2014).

Consequently, some methods of traditional data analysis may still be used for Big data analysis. The main traditional data analysis methods are described below, part of them is from statistics and computer science.

— *Cluster Analysis.* Statistical method for grouping items, and specifically, classifying items according to particular attributes. Therefore, cluster analysis is performed in order to differentiate items with specific attributes and separate them into some groups (clusters) according to these attributes and features. Cluster analysis is an unattended study method without training data.

— *Factor Analysis.* Fundamentally, factor analysis is oriented to identification of relation between particular elements with a few elements. For instance, grouping some related variables in a factor, and then several factors are used to disclose the most information of the original data.

— *Correlation Analysis.* Analytical method to identify the law of relations, in this case, correlation, correlative dependence and common restriction between observed phenomena and accordingly conducting forecast and control. The mentioned used relations could be divided into two types:

  - function, which indicate the exact dependence relationship inter phenomena, which sometimes is called as definitive dependence relationship;

  - correlation, several dependence relations which are undetermined or inexact, and the numerical value of a variable may correspond to several numerical values of the other variable, and such numerical values present a regular fluctuation surrounding their mean values.

— *Regression Analysis.* Mathematical analysis for correlations indication between one variable and some other variables. Regression analysis recognize relationship of dependence in variables hidden by random items. Moreover, regression analysis may perform complex and undetermined correlations between variables to be simple and regular.

— *Real-time analysis.* Generally, this method is used in E-commerce and finance fields. When data continually changes, fast data analysis and required and analytical results should be received on time. This method architectures include:

  - parallel processing clusters which are used in traditional relational databases;

  - memory-based computing platforms, for instance, Greenplum from EMC and HANA from SAP are real-time analysis architectures.

— *Offline analysis.* This method is normally used for applications which are not required to provide the result on time, for instance, machine learning, statistical analysis, and recommendation algorithms. Moreover, offline analysis manage the analysis process by importing logs in the special data platform within data acquisition tools. According to big data settings, some Internet entities apply the offline analysis architecture which is based on Hadoop. The mentioned application is used to decrease the cost of data format conversion

and ameliorate the data acquisition effectiveness (M. H. Padgavankar and S. R. Gupta, 2014).

Big Data analysis is more oriented to high amount of data which is used in the organizations. Moreover, the mentioned analysis is more complex process which cover the large volume of data to make some trends, forecasts and some assumptions to make decisions. Big Data analysis could be classified into memory level analysis, Business Intelligence level analysis, and massive level analysis. The mentioned analysis methods are described below.

— *Memory-level analysis.* Generally, the method is used when the total amount of data volume is lower than the cluster's memory. Currently, the memory of server cluster exceed high number of gigabytes, while also terabytes are often used. Consequently, in order to improve the analytical efficiency, an internal database technology could be used, while hot data is store in the memory as well. Memory-level analysis is especially suitable for real-time analysis.

— Business Intelligence analysis. This method is used for the situation when the scale of data exceeds the memory level but also could be imported into the Business Intelligence analysis environment. Normally, systems which are able to support terabytes, are provided to Business Intelligence products.

— *Massive Analysis.* This method is used when the data scale wholly exceed the capacities of BI products and traditional relational databases. It should be noted that offline analysis category includes massive analysis (Padgavankar and Gupta, 2014).

According to data analysis types and classification, the Companies, which implements the Big Data into its activities should wholly change the data analysis methods and principles. For this reason, the personnel should be trained not only for Big Data usage, but analysis as well. Big Data analysis is more comprehensive compared to traditional systems and could be used for decision making process.

2.1.5. Big Data implementation issues

The process of Big Data adoption into the organizations' system is a long and difficult operation. Firstly, the organization should prepare a Big Data implementation program, where each requirement, amount of costs and involvement should be included. Secondary, the whole organization should be informed about the changes and involved personnel duties. It should be noted that the mentioned program, its costs and the time of implementation depends on organization industry, structure, size and amount of involved personnel.

Big Data implementation is beneficial for Company's internal processes, data analysis, storage, processing, management and maintenance. On the other hand, Big data usage is more complicated and requires to be prepare for issues related to data management, privacy and security, data access and sharing of information, storage and tools and talent issues. The main challenges and issues which could raise after Big data adoption are analyzed below.

Data management issues. Data management makes an impact to selection of data. Big Data is formed by multiple heterogeneous sources with totally different nature and representation. For this

reason, data management process requires innovative, modern, high performance and multi-dimensional management tools, otherwise, the results could be unappropriated (Wu, Zhu and Ding, 2014). As it was mentioned above, one of the main Big Data characteristics is variety, for this reason, in order to manage data with heterogeneous formats, it is required to have more sophisticated data store. It should be noted that Company's management need to have relevant, accurate and completed data in order to perform analysis of the business and make an important decisions related to strategy, missions, vision and Company main aims. Data management team perform tasks as cleaning, transforming, data clarification and validation. In terms of data management, Companies assume that the data moving from traditional platforms to the new ones, could be expensive and consuming process. As Big Data is not in the traditional form, the analysis of Big Data could be complex and difficult to receive meaningful information from it. Moreover, there is a demand to update the currently used Big Data management techniques and scalable data management tools (Wani and Jabin, 2018).

Privacy and Security. According rapidly increase of devices connected to the Internet and connected to each other, the amount of data collected, stored, and processed is growing every day, which also brings new challenges in terms of the information security (Moura and Serrao, 2015). One of the most important issue with Big Data is privacy and security that includes conceptual, technical as well as legal significance. Personal information is sensitive data of a person and could be inference of additional facts about the person when data is combine with external large data sets. It is important to add, that personal information is secretive and the people might not want Data Owner to know or any person to know their personal information (Katal et al., 2013). In Big Data analysis, there is a set of risks that must be considered. The mentioned risks include the lifecycle of information (origin, ownership and data classification), the process of data composition and collection and the security procedures deficit. After all, the Big Data security and privacy purposes are the same as for other data types – to protect its confidentiality, solidity and approachability (Moura and Serrao, 2015).

Data Access and Sharing of Information. In general, data is used to make specific decisions in time, for this reason data should be available in terms of precise, accomplished and timely approach. Therefore, the process of data management and governance became complex and need to make data open and available to government agencies in standardized approach with standardized metadata and formats thus conducting to better decision making, business intelligence and improvements of productivity. It is understandable that data sharing between companies is impossible process for need to be superior in the market. Data sharing about Company's clients and internal procedures is harmful for secrecy and competitiveness (Katal, Wazid and Goudar, 2013).

Storage Issues. Storage availability depends on the used data. The more information Company have, the more attention and resources should be taken. According the nowadays situation, lack of devices in witch data could be store is one of the most important issue related to Big Data usage and implementation (Wani and Jabin, 2018). In general, existing systems have the storage capacity where only 4 terabytes could be store per one disk. In terms of Big Data, it is normally stored in exabytes. To make the comparison clear, in order to store 1 exabyte, 25 000 disk spaces are required. One of the possible solution could be data storage into the cloud. On the other hand, the process of data transfer from multiple data source to the one and back from could to the processing point take long time (Kaisler, Armour, Espinosa, Kaisler and Armour, 2013).

Tools and Talent. One of the main problem in using and implement Big Data could be the lack of tools and trained human resources to work with big data in correct way. Regardless of the proliferation of big data and analytics training programs, this mentioned issue is still a concern (Espinosa et al., 2019). Big Data is high processing and requires complex technologies, for which high skilled specialists are not trained. There are two solutions: to invest into specialists' trainings to improve their skills and understanding about Big Data implementation and processes or to employ contractors form the third party which provide specialists which are already trained. For the business organization which are engaged with Big Data analytics, there is a huge demand for this kind of specialists who are able to support Big Data processes (Wani and Jabin, 2018).

It is noted that "big data from threat monitoring systems pose serious challenges for cyber operations because an ever growing number of devices in the system and the amount of complex monitoring data collected from critical infrastructure systems require scalable methods to capture, store, manage, and process the big data." For this reason, the suggestion eas made as "cloud computing based network monitoring and threat detection system to make critical infrastructure systems secure" (Chen, Cai and Zhan, 2016).

Based on reviewed issues above, Big Data implementation and usage is a huge challenge for Company's based on traditional IT systems. Big Data requires investments on IT systems, its storage, processing process, human resources and various tools. It is important to add, that each entity which is oriented to implement Big Data, should analyze possible problems, raised issued and analyze a solutions for them.

2.2. Theoretical analysis of Financial Audit field

The master thesis covered the two fields – Big Data and Audit – as the theoretical part above was dedicated to Big Data, the second part of theoretical part – oriented to Audit aspects. Firstly, for better understanding of audit, the description is characterized in the following section. Audit process and whole activities of audit companies is strictly supervise by the requirement and legislations which are described in the following part too. Audit process consist of audit planning, which cover calculations of materiality, audit procedures performing and preparation of audit opinion. Each mentioned process and the main aspects are reviewed and explained for better understanding of audit value to clients and audit other subjects as customers, suppliers, shareholders, stakeholders and lenders.

2.2.1. Financial Audit description

In order to understand the Audit process and procedures, it is important to realize the description of Financial audit and the main responsibilities of auditors. Audit and its process ir strictly regulate by international and local institutions. In general, the main concepts of financial statements are explained in the International Standards on Auditing (ISA) and International Financial Reporting Standards (IFRS). In the section below, the main two recourses of information are ISA and IFRS.

The purpose of financial auditors is to perform particular procedures (based on Company's activities, industry and risks) and provide auditor's opinion about financial statements (International Financial Reporting Standards). According to the ISA and paragraph 11 of ISA 200, the main objectives of the auditor are to receive reasonable assurance of financial statements without material

misstatement and to report on the financial statements, and communicate as required by the ISAs, in accordance with the auditor's findings (International Standards on Auditing (ISA), 200).

In terms of Lithuanian Financial Audit reglamentation, the audit requirements are specified in Law on the Audit of Financial Statements, issued in 1999 year, the last amendment was issued in 2016. The mentioned law determine the main definitions used in audit processes, the main principles and requirements which have to be met by the auditors and audit companies and other relevant information. According the mentiones document, Auditors report means **"**a document in which an auditor, an auditor of another Member State or an auditor of a non-Member State, in addition to other mandatory items, expresses an opinion on the audited financial statements " (Law on the Audit of Financial Statements, 1999).

One of the main aspects of auditors is their independence. The whole auditors' responsibilities and duties are based on independence, which means that they could not have any relationships with management or governance of the audited Company. Independence requires independence in mind and independence in appearance. Independence in mind means that the position of mind permits the demonstration of a conclusion without affection by objects of impact that compromise professional judgement. Moreover, independence in mind enable auditors to work with integrity, objectivity and professional skepticism. Also, independence in appearances are more dependent interpretation of others' of this independence and their faith in the auditor (ISA, 200).

Generally, financial audit is required by the potential investors and shareholders in order to be guaranteed that the information provided into financial statements represents the real situation of the Company. It should be noted that financial audit performed by the external auditors is mandatory to state and municipal enterprises, public interest entities, public limited liability companies, private limited liability companies, cooperatives, general partnerships and limited partnerships and private limited liability companies a shareholder whereof is the State and/or a municipality. Moreover, statutory audit is obligatory for private limited liability companies, cooperatives, general partnerships and limited partnerships all participants whereof are public limited liability companies or private limited liability companies when the company exceeds at least two requirements on the last day of the financial year: net turnover during a financial year – EUR 3.5 million, value of assets in the balance sheet – EUR 1.8 million, average number of payroll employees during a financial year – 50 (Law on the Audit of Financial Statements, 1999).

It should be noted that absolute accurance could not be provided for some of reasons. For instance, it is impossible to discover each human error. In oder to detect every error during the financial year, auditors should review each transaction, invoice and other supporting documents wich require a lot of time and human recources. In General, auditors prefer sampling techniques to choose and notice audit evidence which are based on statistical or other criteria, therefore there is a possibility that there could be some mistakes which are undetected. For this reason, the auditors calculate the materiality of audited organization which is as the landmarks in the audit process. Materiality is a concept which identify the significance of accounts or transactions. The audit report which is issued by the Audit Company in order to express the opinion about financial statements has an important meaning. Audit report is normally used by current or potential investors which are not directly participate in the Campany's activities, but are interested in Company's financial result. Also, audit report is used for banks or credit istitiutions when the Company has credits or other relations with banks. As a consequence, auditors are responsible for audit process to appropriately test controls,

perform procedures and other testing in order to detect financial fraud in operations and discover material misstatements (ISA, 200).

On the other hand, auditors also involve some actuarial, tax, information risk management (IRM), financial consulting specialists into the audit process. Actuarial specialists are involve only into specific engagements as insurance companies or banks for valuation of reserves, while IRM specialists are generally involve into all engagements to test IT controls and the ability to rely on information provided by the client. Financial consulting specialists are incorporiate for engagements when Company's owns real estate, the mentioned specialists perfom real estate valuation procedures for possible impairement calculation. Also, tax specialists are include in the audit team when auditors determined a risk related to tax accounts.

2.2.2. Financial Audit planning

The purpose of the auditors is to design the audit plan by applying professional judgmental and skepticism. The significant supporting document for audit planning is ISA 300 *Planning an Audit of Financial Statements.* During the planning step, the auditor's targets are to pay attention to the most important areas, recognize potential problematic areas and solve it on a timely basis and be responsible for effective and efficient organization of the engagement. Moreover, auditors are responsible of understanding of the entity, its environmental and possible risk of material misstatemen tin the financial statements. Planning step is as a guide to the auditor, it guarantees that the correct decisions are made to the specific areas (e.g. high risk, material) at the right time. The whole engagement team should be involved into the audit planning process and discussions regarding the important areas of engagement. The engagement partner is responsible for the audit plan and strategy (ISA, 300).

The relevant standard for understanding the Entity and its environment is ISA 315 *Identifying and Assessing the Risks of Material Misstatement.* In order to understand the entity and its environment including internal controls, the auditor should follow the main steps: understand the entity, its activities and identify potential risks regarding controls, understand internal control by performing design and implementation procedure of relevant internal controls to evaluate the potential risk of material misstatements, connect the risks which have been identified to what could go wrong at the assertion level and at the overall financial statement level (ISA, 315).

The process of understanding of the entity and its environment is a long, dynamic process of obtaining, gathering, updating and investigating information form the different perspectives. Risk assessment procedures to gain relevant information of the entity include inquire, observe, review, inspect and analyze, analytical procedures, benchmarks and other procedures. The mentioned procedures are listed and explained above.

- *Inquire.* Answers of provided questions from governance, managements, shareholders or others. Generally, the questions are about business activities, financial reporting or accounting policies, procedures how the company deals with unusual transactions, governance, risk management actions, IT systems, internal audit, legal issues and other relevant information.

- *Observe.* Notice company's activities and operations.

— *Review, inspect and analyze.* Use external sources of particular industry, legal or economic information (e.g. various databases, official reports about specific industry, rating agencies, economic indicators). Review internal business plans, budgets, strategies, values, internal reports regarding the Company‘s activities and plans. Analyze reports prepared by the management and governance as accounting policies, board minutes, audit committee reports.

— *Analytical procedures, benchmarks.* Analysis of both external and internal sources. Comparison of current year and previous year financial and non – financial information. Main performance indicators and activity analysis compared to similar companies.

— *Other procedures.* Visiting premises, review daily processes of the Company as entries into accounting system and other.

Prior year information as organizational structure, control analysis, management attitude and actions could be used if the information is up to date. For this reason new auditors contact previous year auditors and inquire them about processes and problems which could be applicable in the current year as well (ISA, 315).

In order to perform audit, the audit Clients have to provide financial information related to audited financial year. When audit companies have received the mentioned information, one of the first steps is calculation of the materiality. First of all, it is important to understand the definition of Materiality, Performance materiality (ISA, 315).

Financial reporting frameworks explain the concept of materiality in the frame of reference of the preparation and presentation of financial statements. Despite the fact that, financial reporting frameworks determined materiality in different ways, they normally determined that misstatements are evaluate to be material when they individually or in the aggregate could be expected to make an impact to economic decisions of users which use audited financial statements. Judgements regarding the materiality are made based on conditions which are related to size or nature of a misstatement (ISA, 320).

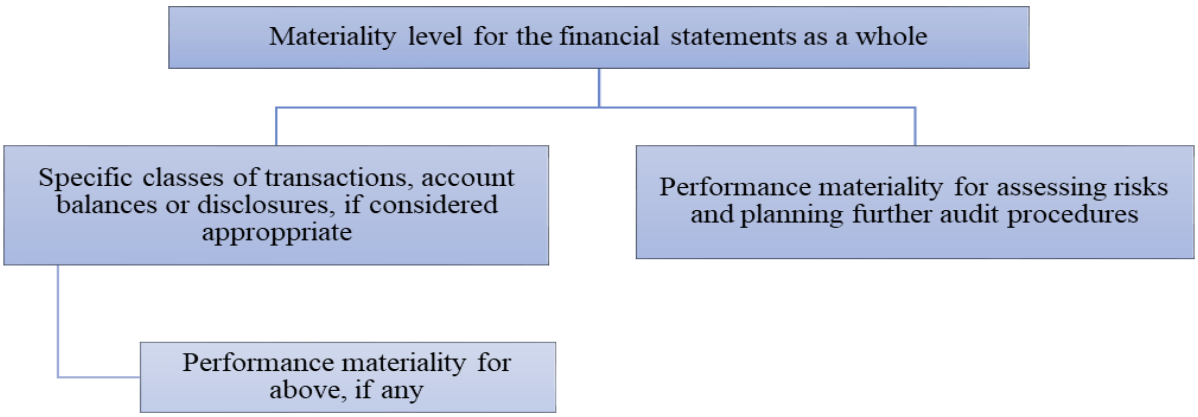

Figure 6. Levels of materiality (ISA, 320).

Performance materiality is the amount calculated by the auditor, which is less than the materiality for the financial statements. The mentioned amount is set in order to reduce the level of probability that the total amount of uncorrected misstatements exceeds materiality for the financial statements. Moreover, performance materiality make reference to the amount or several amounts selection by the auditor for specific classes of transactions, account balances or disclosures (ISA, 320).

It is important to add, that auditor must use his professional judgement to consider which accounts are material and which are not material. Auditors must have an understanding about the entity and its environmental, have enough information regarding the Company's financial results and be informed about the requirements of the users of financial statements. The determination of performance materiality is not based only on technical mathematical calculations, but also depends on the nature of the company, the past experience related to auditors (e.g. immaterial or material findings in the previous audits), use of professional judgmental and the possible misstatements in the current audit period (ISA, 320).

As it was mentioned above, materiality is one of the first step in the audit, it influences the audit planning process, audit performance (founded errors or misstatements may have an effect to the whole materiality level), make an impact to determination of audit procedures, estimation of errors identified during the audit process and measure misstatements in the financial statements.

Understanding the entity and its environmental creates ability to auditors apply professional judgement to identify what is material in the concept of the entity and its environment and control procedures. Also, professional judgement helps to classified transactions, balances and disclosures which are material and important to the audit process. It is important to add, that each entity is unique and what is material class of transaction for one entity may be immaterial to another one (ISA, 320).

2.2.3. Financial Audit process and procedures

Audit process, procedures and risks are individually and normally depending on Company's industry, activities, result, controls efficiency and other features. The first step of audit planning is identification of the significant risks and efficiency of controls, then the audit procedures are planned regarding the audit assertions. Definition of Assertions is defined in ISA 315 as "Representations by management, explicit or otherwise, that are embodied in the financial statements, as used by the auditor to consider the different types of potential misstatements that may occur ". Assertions are classified into two groups: assertions related to classes of transactions and events and connected to disclosures for the period under audit and assertions about account balances and related disclosures at the period end. Assertions are listed and analyzed in the Figure 5 and the paragraphs below (ISA, 315).

Assertions about classes of transactions and events and related disclosures for the period under audit is used for accounts of Income (Statement). Each assertion is explained as following:

— *Occurrence*. The transactions and events which have been recorded, occurred, and such transactions and events pertain to the entity. For instance, recorded sales express the goods which are sold or services which are provided to the customers. The evidence of sales

occurrence is the issued invoices to the customers, which are the appropriate evidence that sales are not overstated.

— *Completeness*. The transactions and events that have been recorded and all related disclosures that should have been included in the financial statements are already included. The mentioned assertion is mainly related to expenses or liabilities that Company could be oriented to increase its profit by the decreasing its expenses. For example, all occurred staff costs related to employed personnel have to be recorded in the accounting system.

— *Accuracy*. Each entry and other data which is to recorded transactions and events have been recorded appropriately, and related disclosures have been appropriately measured and described. One of the example could be that booked sales into booking system are correct and booked sales amounts are the same as into issued invoices.

— *Cut – off*. Transactions and events have been recorded in the correct accounting period. Cut off normally is used to review of post balance sales to make sure that sales invoices issued in the next financial year are not related to the audited financial year.

— *Classification*. Transactions and events have been recorded in the proper accounts. For instance, staff cost has been reasonable separation in operating expenses and administrative expenses.

— *Presentation*. Transactions and events are fairly aggregated or disaggregated and clearly described, and related disclosures are relevant and understandable in the context of the requirements (ISA, 315).

Assertions related to account balances and related disclosures at the period end are used for transactions and accounts of Balance Sheet. Each assertion is explained as following:

— *Existence*. Assets, liabilities and equity interests exist. Stock or asset which is recognized in the accounting system exists at the period end, for instance, the accuracy of stock could be tested during the stock count.

— *Rights and obligations*. The Company contains or controls the rights to assets, and liabilities are the obligations of the Company. It is important to ensure that the audited Company owns or controls the asset which is recognized in the financial statements. If asset or inventory consider by the audited Company on account of another Company, has not been recognized as part of inventory of the audited Company.

— *Completeness*. All assets, liabilities and equity records that should have been recorded have been recorded and all related disclosures that should have been included in the financial statements have been included there. Assets that should have been recorded have been acknowledged in the financial statements. Asset which is considered by a third party on behalf of the audited Company, should be included in the balance sheet.

— *Accuracy, valuation and allocation*. Balance sheet parts as assets, liabilities and equity records have been included in the financial statements at reasonable amounts and any valuation results or allocated adjustments have been recorded in the reasonable way as well.

— *Classification*. Assets, liabilities and equity interests have been recorded in the proper accounts as tangible and intangible assets, current and non-current assets, long term or short term liabilities and other. The mentioned assertion is important for going concern issue, when noncurrent assets have to exceed the amount of short liabilities. In this case, auditors test if the company is able to cover the whole short liabilities by using its uncurrent assets, when the company is able for this process- there is no going concern risks.

— *Presentation*. Assets, liabilities and equity records are appropriately aggregated or disaggregated and clearly described, and related disclosures are relevant and understandable in the context of the requirements of the applicable financial reporting framework (ISA, 315).

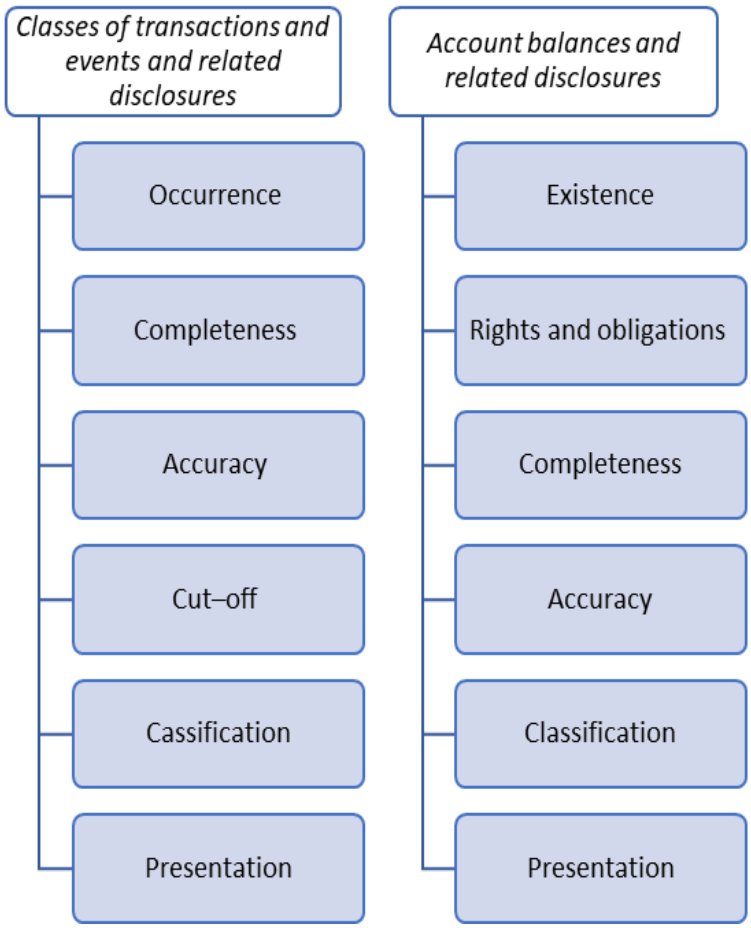

Figure 7. Classification of audit assertions (ISA, 315).

The mentioned assertions are covered during the year end audit. Generally, auditors perform interim audit of half of year, 9 or 10 months. Interim audit consists of internal control system testing, which cover documentation, control testing and evaluation. Control testing and analysis is important for audit planning, when controls are ineffective, auditors perform more particular procedures to test data. During the year end audit, the concentration is on the financial statements and the assertions about assets, liabilities and equity interests. At this step the auditors plan substantive procedures in order to guarantee that assurance has been acquired through the all assertions. Normally, the amount and type of procedures performed by the auditors are contingent on the clients size, industry and significant risks identified during the audit planning stage. For example, procedures, which will be performed in the audit of manufacturing or construction companies, are not applicable for insurance

companies or investment funds. On the other hand, significant risks which are identified in one insurance company, could be insignificant to another one and etc. Audit planning have to be performed individually for each Company.

2.2.4. Auditor's Report of Financial Statements

The next stage of the audit process is analysis of misstatements and communication with client's management. Some of misstatements could occurred because of control deficiency or accountants mistakes, for this reason, the misstatements founded are accept by the management and financial statements are correct regarding the auditors comments and performed procedures. On the other hand, some of misstatements identified by the auditors could not be admit by the management and financial statements are not corrected and represents different position as auditors analyzed. These cases are represents in Auditor's Report of Financial Statements, which could be divided into several different types. Each type of the auditor report is shortly described in the following section.

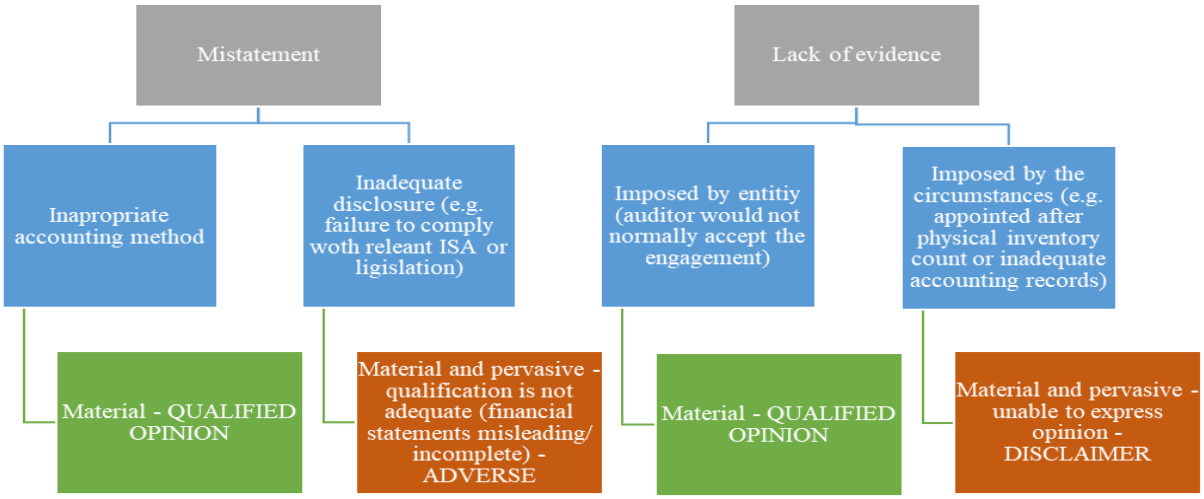

Figure 8. Misstatement and lack of evidence impact to audit opinion (ISA, 700)

Auditors opinion could be divided into two different types: modified and unmodified (clean) audit opinion. A unmodified opinion means that financial statements are prepared in accordance with accounting standards without any material misstatements. The auditor modifies its opinion when the conclusion is made that according the audit evidence received, the financial statements as whole have material misstatement or when auditor is unable to obtain enough appropriate audit evidence to make an conclusion that the financial statements as a whole does not have material misstatements (ISA, 700).

After the audit procedures, auditors are able to make conclusion regarding the process and potential auditor opinion. The first decision is make after the identification of misstatements. Auditor have to evaluate if the misstatements are material or not. The misstatements materiality is compared to calculated Company's Materiality and Performance materiality. Moreover, the auditors need to evaluate the pervasive of the misstatements in the accounts which is also evaluated by the review of audit procedures (misstatements were identified or not). The another conclusion have to made

regarding the lack of evidence. Equally, auditor is responsible for assessment of audit evidence, if the lack of evidence have an material impact or not, also the analysis of pervasive is mandatory as well. The mentioned decisions are summarized in the graph represented above.

2.2.5. Digital Financial Audit

As it was mentioned in the beginning of the second section, the theoretical part of the master thesis is divided into two main divisions: Big Data and Audit. Big Data description and the main characteristics of it, processing techniques, phases of the implementation and the most important issued have been described an analyzed in the theoretical part above. Moreover, audit description and its based regalements, audit planning process, performed procedures and the audit finalization process were reviewed as well. The last section of the theoretical part is dedicated for the analysis of Digital audit and the main circumstances which make an main impact to the traditional audit role and process.

As data and power of digitalization make an significant impact the people's personal and working lives, the auditors function is not an exception. Nowadays, Data control the personal and professional lives and force each industry to alteration of processes. Digital transformation is equally significant to the audit of public services and commercial entities as rapid digital growing is relevant to both types of entities (European Court of auditors, 2020). In order to investigate the mentioned fast – development conditions, the Association of Chartered Certified accountants (ACCA) issued a research "Audit and Technology" in June 2019. This report provided an short research of various technologies which currently make an impact to the audit profession or which it will affect in the nearly future. Moreover, the report also provides a comprehension how the development of audit environment is modeling technological development in auditing – how different technologies could be expected to affect the future processes in the companies and why it is important to audit teams. Moreover, the specialists provided the main aspects which technological are changing audit, the most important aspects are determined below.

Artificial Intelligence. AI is frequently determined as 'an evolving technology' that is providing computer systems with some additional tools which are similar to human intelligence, but it is better noticed as an umbrella term for technologies groups. The mentioned group of technologies can be merged in different methods. AI is also one of the component of a large amount of hype which is named 'humanlike intelligence', the forecast is made that it will noticeable in 2029. Some of scientist believe that AI will decrease the workforce. Because of lack of knowledge and understanding of subjects as intuition' and 'thought', it could be noted that it is unknown what is trying to be emulate. The questions are raised regarding the measurement of intelligence and emulation of quotients. The 'intelligence' in AI is frequently contained of processing power and access to data, for example, a computer is able to play a game such chess by testing the potential consequences of a move and using game history from the previous movement between the game dimensions. For the ability to analyze and apply the historical data, AI is useful for people: it could analyze the whole population of particular data and characterize the patters or deviations. During the AI usage in audit field, the auditors could make an decisions and have insights without technical work. For this reason, auditors could pay more attention and time to tasks which requires professional judgmental (ACCA, 2019).

Robotic process automation (RPA). People are often confused in terms of RPA and AI determination and facilities. RPA is as 'robots' software which is able to make tasks in repeatedly way. RPA is widely used when data needs to be transferred from one financial process to another one or when multiple sources information required to be find. For this reason, in some cases RPA is called as 'swivel chair automation'. Some of personnel tasks are repetitive, mundane, takes lots of time and the final task could be prepared with some human errors. Normally, the mentioned tasks are more technical and does not require some specific skills or knowledge. The simple example of this task could be sales and purchase invoices booking to the accounting system. The best solution for the monotonous and technical tasks are RPA implementation into the system. Analyzing the mentioned example, the robot would take the information from the paper invoice and book into the accounting system. RPA is responsible for data moving from one system to another one, it is not able to analyze data or make some decisions. The main advantages of RPA are that robots are able to work periodically without breaks, holidays, free time, shortly, robots are able to performed the particular tasks 24 hours per day and 7 days per week. Also, robots are faster, more accurate and able to be changed. On the other hand, there are some issues and questions regarding the accountability and RPA right of possession and data security which is processed during the tasks (ACCA, 2019).

Data analytics. Various data analytical tools have long be been applied to the data analysis process. Some audit companies are already using some data analytical tools for data analysis, testing of transactions, data sampling, forecasting and other. The main advantage of data analytics in audit – it admit auditors to use the whole population's transaction in the testing procedures. Moreover, when auditors are testing some data as journal entries by using data analytics tools, they are required to consider the data completeness and accuracy questions. Auditors are oriented to AI and machine learning algorithms application to improve the quality of testing, prediction and increase the availability of fraud detection into engagements. Companies storage their data into various different public and/or proprietary sources, the cloud – based applications prove an opportunity to combine data from various sources (ACCA, 2019).

Machine Learning (ML). ML is closely related to AI as it is a branch of the AI. Based on ACCA's report "Machine Learning: More Science than Fiction", if the increased in the volume of financial transactions would not be correctly managed, it could became a serious a threat for the accountants position (ACCA, 2019). For auditors usage of ML is related to sample they analyze and its capacity to prepare it in the representative way of the population and make conclusions which can be generalized beyond the sample. ML apply statistical analyses for prediction and decision making process of a high historical dataset. An example of ML is credit analysis before the decision for loans requests. For instance, accounting software company named Xero has adopted ML for preparation of coding decision for invoices. ML suggested to the customers high level of accuracy and quick adoption to particular process. Moreover, ML 'forecasts' could be used not only for future predictions, but for past as well. ML has clear applications for risk management, the detection of fraud and inaccuracy which are done by comparing historical data with current data. The mentioned facilities are the main advantage for risk assessment preparation (ACCA, 2019).

Audit environment is closely related to each clients and theirs used systems and tools. As more and more companies uses various technologies, the auditors are required to change and improve their position. Auditors need to adopt the developments in their client's delivery models and understand how technologies are used by the clients. It is important to add that data analytics is the most mature

of the technologies which is used by most audit companies. On the other hand, machine learning is quite new and not widely adopted to audit engagements practically. In general, technology is a subject which is beneficial for audit shift from a retrospective view to one which is prospective. In order to be step by step with the clients, auditors have to be able to adapt to the changes in business model and systems. Audit field is engaged in early stage with AI using and required more preparation for understanding of AI advantages to the sector and its implementation. It is important to add, that auditors also have a human relationships with the clients, which is the main factor when clients are choosing the relevant audit company, the human relationship could not be replaces by the technology systems. Moreover, auditors have to be more open for innovative technologies and changes which are mandatory in order to provide high quality services (ACCA, 2019).

The majority of audit tasks are prepared by the audit specialists, in order to manage the technology development, audit specialists are required to improve their competencies and skills in the nearly future. IT skills are beneficial for audit procedure and whole process, but according the periodically technology movement forward, the regularly improvement of the mentioned skills is mandatory process. The main skill for auditors need to be provided is flexibility to adapt to technology development and client's environment which continuously changing each day. Audit employers should be oriented into investments to employees trainings, new hires with particulate IT skills for new roles.

ACCA prepared a research named "Professional Accountants – the Future: Drivers of Change and Future Skills" about the required skills to auditors in the future and how they could change. The research characterized a group of quotients which are needed now and will be required in the future for financial reporting function. Quotients for the future auditor are technical skills and ethics, intelligence, creative, digital. Emotional, vision and experience. Technical skills and ethics – the skills and abilities for audit performing regarding the standard by the applying the highest quality of integrity, independence and skepticism. Intelligence means the ability to acquire and use knowledge for decision making, consideration of significant risks. Creative is about the ability to apply knowledge in unknown and new situation, make connections and develop new ideas. Digital means that auditor needs to be aware about the application used in digital technologies, capabilities, practices and strategies. Emotional is the skill to identify the own emotions and those which are felt by others, and adopt them to performing tasks, also regulate and manage them. Vision means to be able to foresee future trends accurately by extrapolating present trends and facts and fill the gaps using the knowledge and innovatively ideas. Experience – collected skills to realize customer needs, expectations, meet desired outcomes and create value (ACCA, 2016).

Auditors future is significant related to technology changes in current word. As the expectation for auditors is growing, auditors are required to improve their technology skills and be step by step with the clients' needs. Future auditors need to be technologically skilled with project management skills and oriented to adopt technology development changes. Auditors role is changing, requirements and expectation for the future auditors are growing, for this reason, they need to be flexible open for innovations.

2.3. Big Data and Big Data Analytics integration in Lithuanian market

Big Data accomplishment into business processes is a challenge not only of employees of the business but external auditors as well. A further issue is that external audit is based on various

standards that were issued some years ago when big data was not used. For this reason, auditors are required to consider their used tools and processes for audit procedures. The audit of the future could be different from nowadays as auditors will be capable to use massive data sets to deeper understanding of business, its activities and the main risks. On the other hand, only audit market leaders are oriented into Big Data audit processes which clients uses high volume of data.

The minority of researchers worked on analysis based on Big Data and Big Data analytics in audit market. However, the mentioned topic is relevant and important in the current period, when new technologies are widely used. As Big Data and Big Data Analytics provides deeper information for audit procedures performed and insight regarding the decision making, it might became as competitive advantage of the Audit Company or whole audit industry.

According the ACCA (ACCA, 2020), the latest new technologies development cause significant benefits for audit specialists, as current technology revolution is the reason for the demand of highly qualified auditors. current technology revolution is based on rapidly increase in volume of data, corrections in business models, automation processes adoption and broad-minded approach to audit. The mentioned evolution require auditors to be technologically prepared for success audit services.

ACCA analysed and prepared a review of each technology development which is widely used and could make and impact to auditors daily process. The researchers provide the key messages regarding the Big Data and Big Data Analytics implementation in business and audit markets. Firstly, auditors are required to implement the changes in business models. Data Analytics is the most mature and widely used by the business entities in this technology development century. Audit specialists are in very early stage with Artificial Intelligence usage and has not involved it as deeply as could. Auditors are required to be more flexible for changes into the technologies (ACCA, 2020).

There are many researches about Big Data and Big Data implementation in audit market from the global perspective, but there is lack of reasearches about Big Data and Big Dat analytics adoption in Lithuanian audit market. Researchers L. Dagilienė and L. Klovienė, analyzed the main aspects for motivation to use Big Data and Big Data Analytics in external auditing. The research is prepared through the three different perspectives – audit companies, business clients and audit regulators. Motivation were analyzed by using following aspects: Current practices, Institutional factors, Strategy related factors, Resources related factors, Company factors, Internal control, Audit process, Quality and Outcome. Research results provide audit companies' main purpose to use Big Data Analytics and improve the understanding of Big Data and Big Data Analytics value created in business entities. Moreover, audit companies need to improve their tools in terms of Big Data and Big Data Analytics for audit quality improvement and audit efficiency which could be as the result for competitive audit fee (Dagilienė and Klovienė, 2019).

Finally, based on theoretical framework and prior researches, audit sector is closely dependent to other sectors in which their clients work. For this reason, Big Data implementation into the business entities make an impact to audit sector companies either. For instant, when audit clients implement Big Data into their processes, Big Data analytics practices need to be adopt to traditional audit procedures performed by the external auditors. Therefore, after reviewing the scientific literature and prior researches, the main challenges and impact of Big Data and Big Data Analytics implementation in Lithuanian audit market will be evaluated in the last Master thesis section.

3. Research methodology

Problem of the research. Big Data and Big Data analytics are adopting to different companies in various industries. The impact of data and digital processing is widely researched, the main challenges and opportunities in Big Data implementation is well known. It was observed that there is gap of empirical research about Big Data and Big Data Analytics impact to audit companies in Lithuania.

Object of the research. Big Data and Big Data analytics implementation in audit sector in Lithuania

Purpose of the research. Empirical test of the impact of Big Data and Big Data Analytics implementation in audit field and disclosure of potential position of audit firms in the future.

Method of the research. As Big Data and Big Data Analytics is not widely used in Lithuanian audit market, for opinions collection qualitative research - structured expert interview - was selected.

Structured expert interview:

Type of research. Qualitative.

Population of research. Auditors working in Big 4 companies in Lithuania and IT specialists employed in big or middle size business entities in Lithuania.

Sample of research. During the research it was inquired 11 audit specialists employed in one of the Big Data company in Lithuania and 12 specialists employed in big or middle size business entities in Lithuania. It was expected to inquire more respondents, the ability for inquires and opinions collection was limited for quarantine declared for COVID-19.

Sample research method. Probabilistic, not accidental.

Data collection method. Online questionnaire (interview).

Data analysis method. Qualitative partial content data analysis.

Research Questionnaire. During the semi-structured interview, respondents were inquired to answer 10 question regarding the topic of research. The questionnaire consist of three sections: the first one is related to identification of respondents and their competence, the second one is more oriented into Big Data implementation in business entities and the third one focus on Big Data into audit perspective.

The first part of questions is related to respondents, their industry they work in, volume of work experience and position of their role. Also, how many experience they have and what kind of position they held. Respondents are able to choose the industry their work in, the possible industries to choose are Transportation and Warehousing, Telecommunications, Information Services and Data Processing, Finance and Audit, Real Estate, Rental and Leasing, Legal Services, Construction and Other. Also, respondents are able to identify their own industry if no one listed is not applicable.

The second question is regarding the respondents work experience, how many years they work in particular sector they have selected in the first question, the available answers are Under 1 year, from 1 to 3 years, from 3 to 7 years, from 7 to 10 years and Over 10 years. The third question is related to respondents role, the available answers are Manager (Business unit or department), Senior Specialist/ Senior Team member, Specialist/Team member and Other, Also, respondents are able to identify their own role if no one above is applivable. The last two questions are more related to variation of questions. It may be tendency that respondents with more work experience in higher role could have different opinion then respondents who have less work experience and hold the position of beginner level.

The second part of the questionnaire is oriented to Big Data implementation in the business field. The first question is related to sectors which could be the most oriented into Big Data implementation in Lithuania. The respondents have an opportunity to choose more than one sector for this question. The following sectors are provided as possible answers: Construction, Manufacturing, Wholesale and Retail, Transportation and Warehousing, Telecommunications, Information Services and Data Processing, Finance and Insurance, Real Estate, Rental and Leasing, Education, Health Care and Social Assistance, Hotel and Food Services, Government and Public Administration. At least, respondents can identify their own industry or industries for this question.

The second question of this part is connected to the main reasons why companies use Big Data, the possible reasons are that companies use high amount of data, demand of automatically processes, for greater data accuracy, for decision making advantages, need wide data analysis and other. In questionnaire I have prepared a question to respondents regarding the areas in which organizations seek to implement Big Data, the following fields are provided as possible answers: eCommerce, eBusiness, Online Operations, Direct and online marketing, Fraud detection / management, Customer and market analysis, Customer service, Supply Chain management and logistics, Information Technology, Finance and administration, Human Resources and people development, Risk management and Other. Also, respondents are able to identify areas by their own, as answers for this question. The last question of this part is about challenges for the business companies in Big Data implementation process, the respondents are request to rank the following challenges in order of importance with #1 being the most important object to #6 being the least important object, the possible challenges are Data management issues, Data security issues, Storage, Lack of human resources, Lack of technical tools, Financial budget. The mentioned challenges are already deeper analysed and explained in Mater thesis second theoretical part.

The last part of the questionnaire is more related to audit perspective in the Master Thesis main topic. Questions are based on changes in audit sector which could be as the result of Big Data implementation in business. The first question is about of the main barriers to successful integration of Big Data into the audit field. Respondents are request to rank the following obstacles in order of importance with #1 being the most important object to #4 being the least important object. The ranked barriers are Data capture, Lack of human resources, Lack of technical tools, Breaches in audit standards and regulations. The second question of this questionnaire part is dedicated to advantages of Big Data analysis for audit. Respondents are request to rank the following items in order of importance with #1 being the most important object to #6 being the least important object. The following answers are provided for ranking: Capability to analyse sub-ledger information, Sample -based analysis expansion, More relevant business insights, Better understanding the entity and its environment, Identification of controls efficiency and Significant risk identification of

particular account. Finally, the last question is general open question for the respondent's insights and opinion regarding the future of Big Data implementation into the business and its audit process. It is important to add that all questions are mandatory and respondents are not able to skip one of the question.

In order to collect opinions, the mentioned questionnaire was sent to auditors and personnel who are working in business entities. The majority of respondents from audit field are working in Big Four Companies, as it was mentioned Big Four companies are the leaders and their clients are potential for Big Data implementation in their systems.

After the responses collection, the opinions were analyzed in the following part of Master thesis. On the other hand, majority of respondents from business were from big or medium size companies. The respondents were chosen that mentioned companies are the most oriented into Big Data implementation. Big and medium size companies have high volume of data and transactions, have enough human resources and financial reserves for the new applications. Also, the answers by the mentioned respondents would be the most appropriate for current Lithuanian position in the Big Data market.

Time and place of the research. The collection of respondents opinions were in the period March – April 2020. Questionnaire was filled and questions were answered remotely through Google online platform for surveys.

Issues in the opinions collection process. As World Health Organization declared the coronavirus pandemic on 11 March 2020, Lithuanian Government made a decision to declared the quarantine in Lithuania from 16 March of 2020 till 30 March of 2020.

Issues in the opinions collection process. As World Health Organization declared the coronavirus pandemic on 11 March 2020, Lithuanian Government made a decision to declared the quarantine in Lithuania from 16 March of 2020 till 30 March of 2020. The quarantine was announced when collection of questionnaire answers was started. The Covid – 19 quarantine limited the communication with respondents as some of them changed the working time or had limited time for additional tasks.

Ethic of the Research. Respondents were informed about the questionnaire purpose and that collected answers and opinions will be used in the master's research part.

4. Research findings and discussion

According theoretical analysis, Big Data implementation and its analysis is quite new and unexplored in Lithuanian market. In General, Big Data is implemented in huge companies which uses high volume of data, various data sets (structured, semi – structured and unstructured data) which are collected in diverse sources and are in different sizes from terabytes to zettabytes. Moreover, analysis of Big Data provide an opportunity to management, internal and external auditors and business users to make better and faster decisions. It is important to add that Big Data implementation requires the whole Company involvement is this process. For this reason, the essential changes could be a significant challenge for old business which have strict traditions.

In order to understand the possibility to implement Big Data into business in Lithuania, the review of Lithuanian business market will be represent in the Fourth Master thesis part. Moreover, the main opportunities, expectations for the next years and main risks occureed will be review for better understanding of business and economic situation in Lithuania. Lithuania was successful and attractive for new Investments and start-ups, especially in fintech sector. The Central Bank of Lithuania created good conditions for Fintech development in the country. On the other hand, Lithuanian economic situation is effected by the global pandemic declared on 12 March of 2020 and quarantine which limited business activities and their economic life.

As auditors position is important into Big Data implementation in Lithuanians business, the short review of Lithuanian audit regulations will be prepared in the following part of Master Thesis. Audit Companies, auditors and the whole audit process is strictly observed by the local institutions. Also, International companies are strictly observed by the International institutions which are responsible for the quality of provided services.

In the following part of master thesis the research of Big Data implementation is represented. As Big Data is a new topic in Lithuanian market, the availability of quantitative data is inaccessible. For this reason, survey was prepared for collection of opinions of business members and auditors. The mentioned questionnaire consist of six multiple, three ranking and one open question. After collection of answers the review of collected opinions will be analysed and reviewed. The analysis is prepared from two different perspectives – from the business entities and auditors positions. Also, in the last part of Master thesis research I will provide my insights regarding this topic.

4.1. Review of Lithuanian business market

In general, Lithuanian economic situation is used to get better, because the country has fully recovered after 2008 crisis. In accordance with numbers, the Gross Domestic Product (hereinafter – GDP) of Lithuania has grown by 3.9% in 2019 compared to the previous 2018 year (Economy, 2019). The growing GDP is the result of increased investments into production capacity and development, which provide more opportunities to the companies in new markets and ensure sustainable development. Significant growth was affected by the domestic demand as well. Nowadays, investments based on constructions of residential and non-residential buildings and engineering structures have increased and the value added created in the mentioned sector accounted for 16% of overall Lithuanian economic growth. On the other hand, the GDP growth is expected to fall down in the next 2020 year. Likewise, wages were rapidly growing in 2019. Annual wages increase more than 9% in private sector. The actual situation of labor market is closely related to labor scarcity. Nevertheless, the number of workplaces is growing and employers

requires only high skilled and motivated personnel. For this reason, the demand of workers is in growth and workers has an opportunity to expect higher wages (Bank of Lithuania , 2019).

Lithuania became attractive to International Companies: Moody's Corporation selected Vilnius, the Capital of Lithuania, for its new office in Europe and started its activities here in 2019. Moody's in engaged in credit ratings, research and analysis tools that promote to transparent and integrated financial markets. Also, Continental has built its factory in Kaunas, Lithuania and started manufacturing processes in October of 2019. Continental is one of the leader of manufacturing companies which produces production for the automotive and transportation industries.

Also, the end of 2018 was success for Lithuanian economical as well, Centric IT Solutions and Hella Lithuania has started their businesses in Lithuania. Centric IT Solutions is one of the largest IT Companies in Neverlands which provides IT solution services to its clients. Hella is manufacturing company which develops and produce lighting and electronic details and systems for the automotive industry. International Companies which have started their activities in Lithuania has brought not only investments to the country, but also new workplaces which provide great opportunity and challenge for Lithuanian specialists to became a part of international Companies.

FDi Intelligence made and research and indicated which cities are the most attractive for start-up and received the most tech start-up relative to their population in year from 2016 till 2018. The mentioned research was published in Financial Times (published on 17 October 2019). Vilnius, the Capital of Lithuania, was in the first place in fDi's first Tech Start-up Foreign Direct Investments (hereinafter – FDI) Attraction Index and the winner of research prepared by the fDi Intelligence. For the results of the reaserch see table below (FDi Intellligence, 2019).

Rank City	City	Number of projects	Number of projects/ 100,000
1.	Vilnius	7	1.006
2.	London	110	0.875
3.	Amsterdam	11	0.802
4.	Helsinki	12	0.797
5.	Singapore	42	0.731
6.	The Hague	6	0.691
7.	Dublin	9	0.646
8.	Belfast	5	0.633
9.	Dubai	21	0.615
10.	Duseldorf	9	0.581

Table 2. FDI projects in 2016 – 2018 (FDi Intellligence, 2019).

The first place of this analysis means that Vilnius received the highest amount of foreign direct investment projects from start-ups in IT services sector during the period from 2016 till 2018. Tech Start-up FDI Attraction Index 2019 is calculated regarding the number of attracted FDI projects and population of the city. In this case, Vilnius attracted 1.006 FDI projects per 100,000 residentials.

As it was mentioned above, Lithuanian economy has grew the last several years, the GDP rate was constantly increasing. Also, Lithuania became an attractive location for international companies as Moody's Corporation, Continental, Centric IT Solutions and Hella, which choose Lithuania for their new offices and manufacturing factories. Furthermore, the last several years were successful

for Fintech sector in Lithuania, the main trends had a straight impact to Fintech position in Lithuania: the emergence of qualified workforce and market promotion by the Bank of Lithuania.

On the other hand, specialists' made expectations for the whole Lithuanian market and economic situation could not be reached in 2020. After the coronavirus expansion in Europe and other Continent, World Health Organization declared the coronavirus pandemic on 11 March 2020. From 12 March of 2020, kindergartens, schools, universities and other education institutions were closed in Lithuania. Also, from 13 March of 2020, the restaurants, bars and other food providers were also closed for risk to distribute the possible virus. The government made a decision to declared the quarantine in Lithuania from 16 March of 2020 till 30 March of 2020, after the quarantine, it was extended for two additional weeks. During the quarantine, only food stories and food providers as restaurants and cafes continuous their activities. It is important to add that restaurants and cafes could provide food only for take away approach.

Furthermore, the movement between countries was limited during the quarantine period, for this reason, the activities of airline Companies has significant decrease, mostly of airline companies fired their employees. The quarantine had an impact to the whole economic situation in the country, the majority of Companies suspended their activities. Suspension of economic activities was unexpected decision and most of economic entities especially small business were not prepared for this situation. The quarantine had an impact to residents as well, some of them lost their workplaces and source of finance. For this reason, it could be concluded, that the economic growth will not reached the expected volume and impacted economic will required some to recover after the financial suspension which were in quarantine period. For this reason, coronavirus pandemic and quarantine in Lithuania could make an impact to the Big Data implementation in Lithuania as well. Companies effected by the financial crisis may avoid additional costs. On the other hand, only huge and massive data owned Companies are oriented to implement Big Data in their systems. In my opinion, huge Lithuanian companies have enough reserves and are prepared for economic stagnation or reduction of income.

4.2. Lithuanian audit market review

Lithuanian Chamber of Auditors (hereinafter – LAR) is a public institution which collaborate all qualified auditors in the Country. The mission of LAR is to unite the all auditors in Lithuania, coordinate their activities, represent and defend their interests, to exercise their professional self-management and to fulfill other public interests. LAR responsibilities are required by the Law on the Audit of Financial Statements which is issued by the Republic of Lithuania (Lithuanian Chamber of Auditors, 2019).

According to LAR Annual activity report for the period 1 September of 2018 and 31 August of 2019, the organization unity 349 auditors in the beginning of period and 352 – in the end of the mentioned. Period. Moreover, there are number of 172 audit companies as at date 2020.03.30 in Lithuania.

The peak of increase of audit companies was in 2002 – 2003, when the amount of audit companies has reached 224. After 2003, the number of the audit companies progressively decreased till 2010. The mentioned reduction of audit companies occurred because of high competitive and strict regulation which are applied to audit companies. Moreover, Financial Crisis in 2008 – 2009 had an impact to audit companies as well, some of small or medium companies were not able to continue

their activities, for this reason, audit companies lost the clients. After 2010, the amount of companies also has decreased till 2018 and reached 173 companies in the whole country. The change of number of audit companies in Lithuania in the period from 2000 till 2018 is represented in figure below.

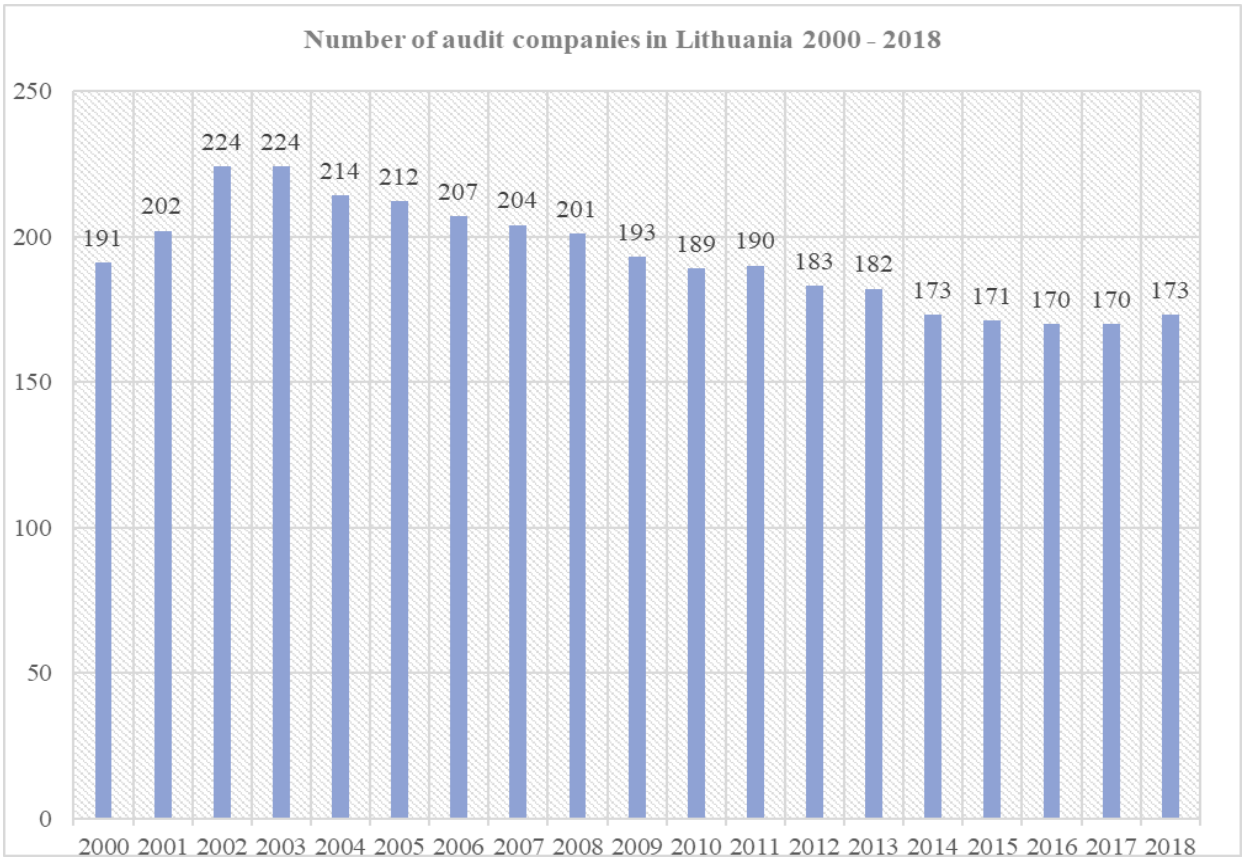

Figure 9. Number of audit companies in Lithuania in the period from 2000 till 2018 (Lithuanian Chamber of Auditors , 2019).

Lithuanian audit market leaders are Big four companies - Ernst & Young Baltic, UAB, KPMG Baltics UAB, PricewaterhouseCoopers, UAB and Deloitte Lietuva, UAB. The mentioned four companies audit and provide tax, accounting and financial consulting services to the biggest Lithuanian and International Capital companies in the whole country. The mentioned three Companies are engaged in audit and assurance, tax and legal, advisory and enterprise services. Also, Big four companies are the main competitors in the local and global audit market.

From the local perspective, EY employed the higher amount of employees in Lithuania compared to other big four companies. The company has 363 specialists for now and have opened two offices in Vilnius and Kaunas. PricewaterhouseCoopers employed about 182 employees and has also two offices in Lithuania – in Vilnius and Klaipėda. Moreover, KPMG has about 171 specialists and also has offices in Vilnius an Klaipėda. As PricewaterhouseCoopers has opened office in Klaipeda only in the last year, KMPG and PricewaterhouseCoopers became the main competitors in this Lithuanian region. At least, Deloitte has about 85 specialists and has office only in Vilnius.

According to data provided by Statista, Deloitte is a leader in terms of revenue in 2019. In Americas region Deloitte earned 24.2 billion U.S dollars, in EMEA – 15 billion U.S dollars and in Asia Pacific region – 7.1 billion U.S dollars. Regarding the revenues the second Company is

PricewaterhouseCoopers which reached 17.8 billion U.S dollars in Americas region, 16.7 billion U.S dollars in EMEA and 7.95 billion U.S dollars in Asia Pacific. EY and KPMG are the last ones according the revenues in 2019. For the revenues comparison see figure above (Statista, 2019).

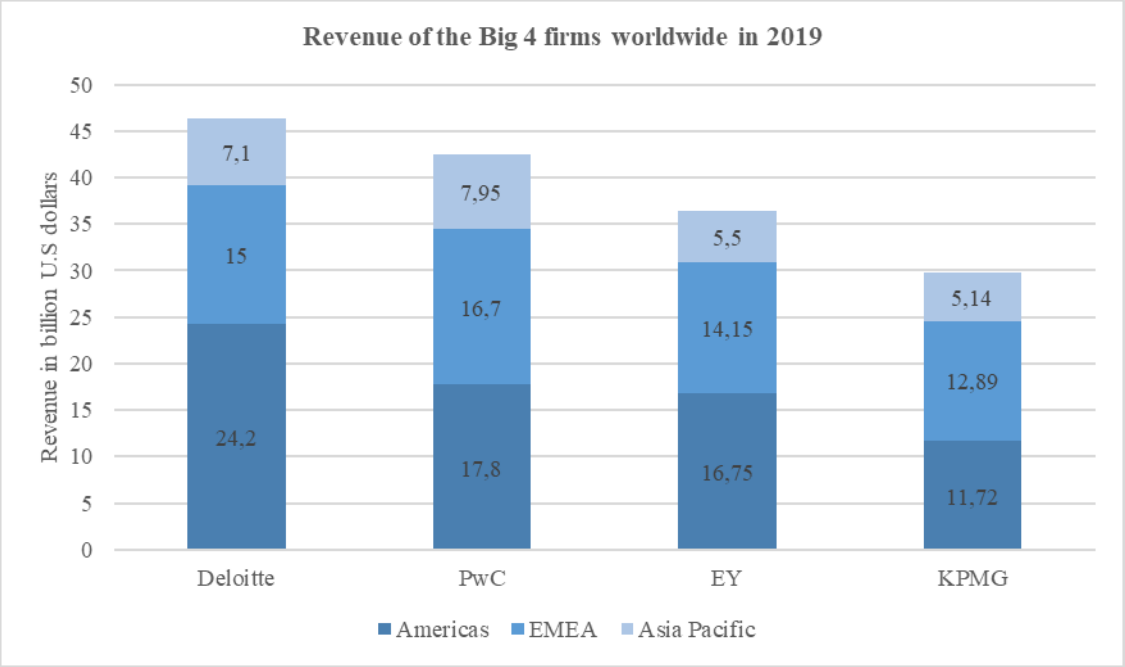

Figure 10. Revenue of the Big Four firms worldwide in 2019, by geographical region (Statista, 2019)

As it was mentioned, audit process and signed independent auditors' opinions are strictly control by local institutions in Lithuania. Also, Big Four Companies are the leader of Lithuanian audit market and audit the biggest Lithuanian Companies. In terms of revenues in the worldwide perspective Deloitte reached the higher amount of revenues in all regions compared to other Big Four companies. As the mentioned companies are the leaders, they are most oriented into the challenges for Big Data implementation in the client's systems.

4.3. Results of expert interviews

In the following section of the Master thesis, the answers collected from the respondents will be presented. The following part is divided into three sections. The first section is oriented to descriptive analysis which consist of representation of answers related to respondents, their position, sector they work in and experience. The second section is based on answers regarding Big Data implementation into the business entities. While the last section is about opinions regarding Big Data Analytics adoption to audit sector.

4.3.1.Descriptive analysis

The answers were received from 23 respondents, the majority of respondents are working in Big Four Companies and big or small size business entities. The survey was prepared and respondents were inquire in the period from March till April 2020. As quarantine for COVID-19 was declared in Lithuania, the answers of questionnaire were collected by remote method using online Google survey application. In the following part of the Master Thesis the descriptive analysis will be presented.

As the topic is modern and unstudied, the empirical test was prepared and the 23 respondents were inquired. According the results, 11 (48%) responds are working in Finance Services and Audit industry and other 12 (52%) respondents – are working in other industries in Lithuanian market. The respondents classification was prepared by the sector, respondents which are working in the Finance Services and Audit industry – are auditors, while respondents from the other sectors – specialist from business entities.

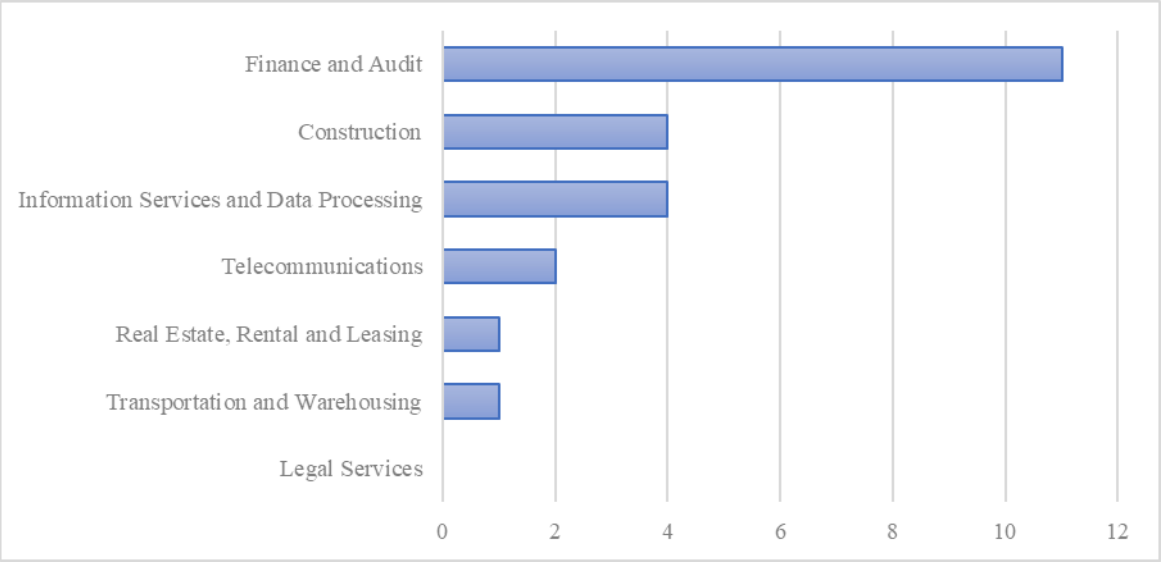

Figure 11. Industries where respondents work

More precisely, four respondents work in Information Services and Data Processing field, also four respondents are from Construction industry while two respondents are from Telecommunications. Only one respondent work in Transportation and Warehousing and Real Estate, Rental and Leasing sectors. On the other hand, there were no respondents working in Legal services or other industry. The distribution of respondents is described in the figure 11, represented above.

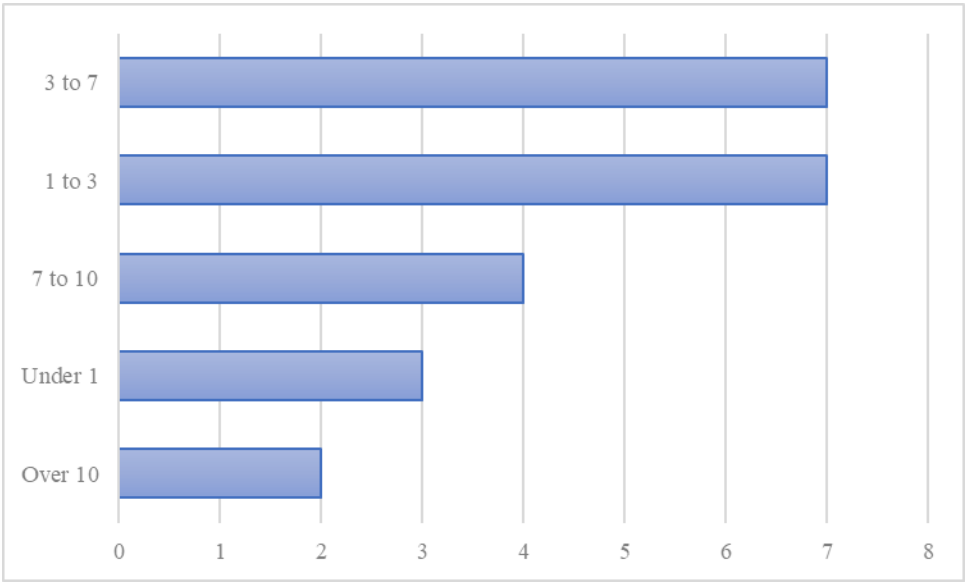

Figure 12. Respondent's work experience

The majority of respondents have work experience from 1 to 3 and from 3 to 7 years, 7 respondents are included in each range. Additionally, four respondents have work experience from 7 to 10 years,

three respondents under one year and only two respondent over 10 years. The detailed respondents work experience representation is provided in figure above.

The third question is closely related to the work experience as the role is usually closely related to employment history. According the questionnaire results, the majority of respondents, eleven of them are working in Senior Specialist or Senior Team member position, eight respondents are as Specialists or Team members and only four have achieved Manager position. Regarding the work experience results and positions, it could be notice that respondents are experienced and are able to provide widely and appropriate opinion regarding Big Data implementation into the business field and Big Data auditing process.

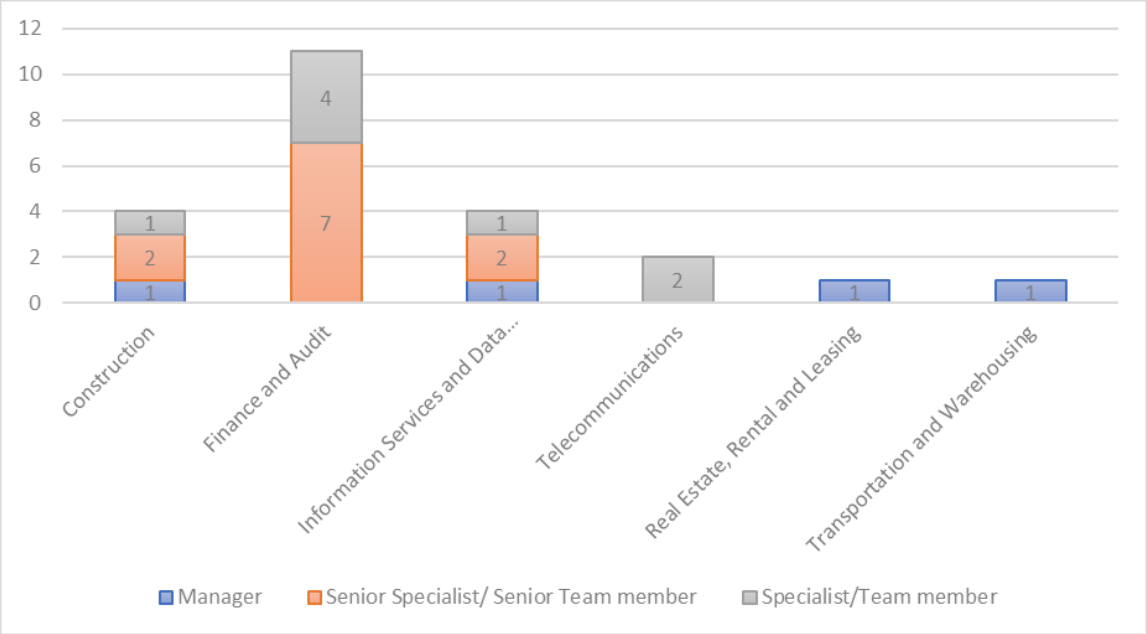

Figure 13. Respondent's work position and industry

Moreover, two from four respondents working in Construction industry held the Senior Specialist or Senior Team member position, while other two – Manager and Specialist or Team member positions. Majority of respondents working in Information Services and Data Processing sector held the Senior Specialist or Senior Team member positions, while other two – Manager and Specialist or Team member position. Both respondents from Telecommunication industry are working as Specialists or Team members in their companies. The left two respondents from Real Estate, Rental and Leasing and Transportation and Warehousing industries are working as Managers.

As it was mentioned below, 11 respondents are working in audit companies, the majority of them - 7 respondents held the Senior Specialist or Senior Team member position, while the other 4 are Specialists or Team members. The distribution of respondents' positions and industries are described in the figure above. After the descriptive analysis it could be concluded that respondents are experienced and qualified to provide appropriate opinion regarding the Big Data and Big Data Analytics adoption to Lithuanian business and audit market.

4.3.2.Big Data implementation in business entities

The following part of the Master Thesis is dedicated for answers related to Big Data implementation in Lithuanian business market. In order to understand the position of respondents from the business entities and audit industry, the analysis of answers will be prepared by testing responses of the mentioned groups separated.

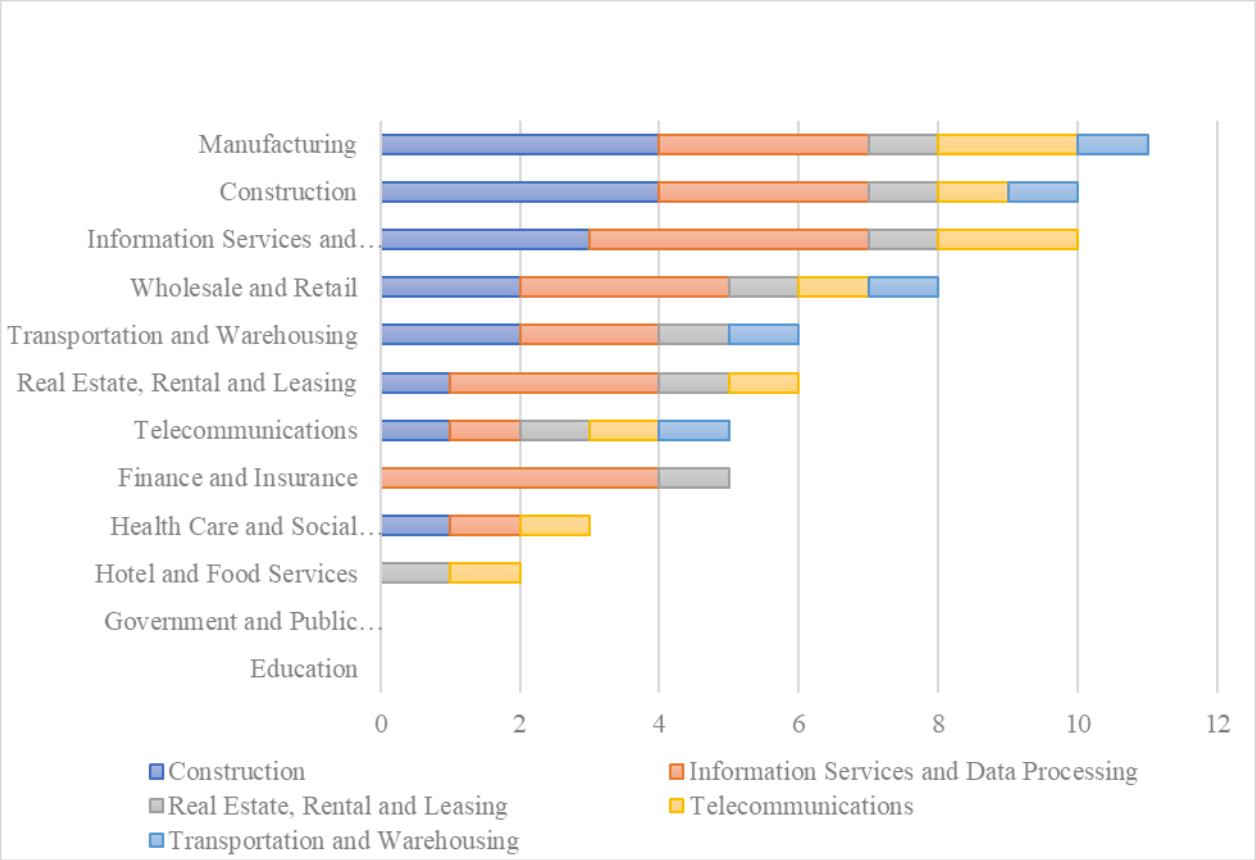

Figure 14. Sectors oriented into Big Data implementation by the business entities

According the next questionnaire answers, the majority of business entities respondents have an opinion, that Manufacturing, Construction and Information Services and Data Processing are the most oriented sectors into Big Data implementation in Lithuania. According the picture above, 11 respondents from 12 have an opinion, that Manufacturing is the most oriented to Big Data implementation into its sector, while 10 respondents also selected Construction and Information Services and Data Processing. On the other hand, specialists from business entities have an opinion that Government and Public Administration and Education sectors are behind the demand of Big Data implementation into their sectors. The distribution of opinions regarding the sectors is presented in the figure 14.

Auditors position regarding the sectors, which are more oriented to implement Big Data is totally different compared to opinions of respondents from business entities. According the audit specialists, Finance and Insurance, Information Services and Data Processing, Wholesale and Retail and Telecommunication sectors were identified as most oriented to Big Data implementation. The answers received from the auditors are identified in the figure below.

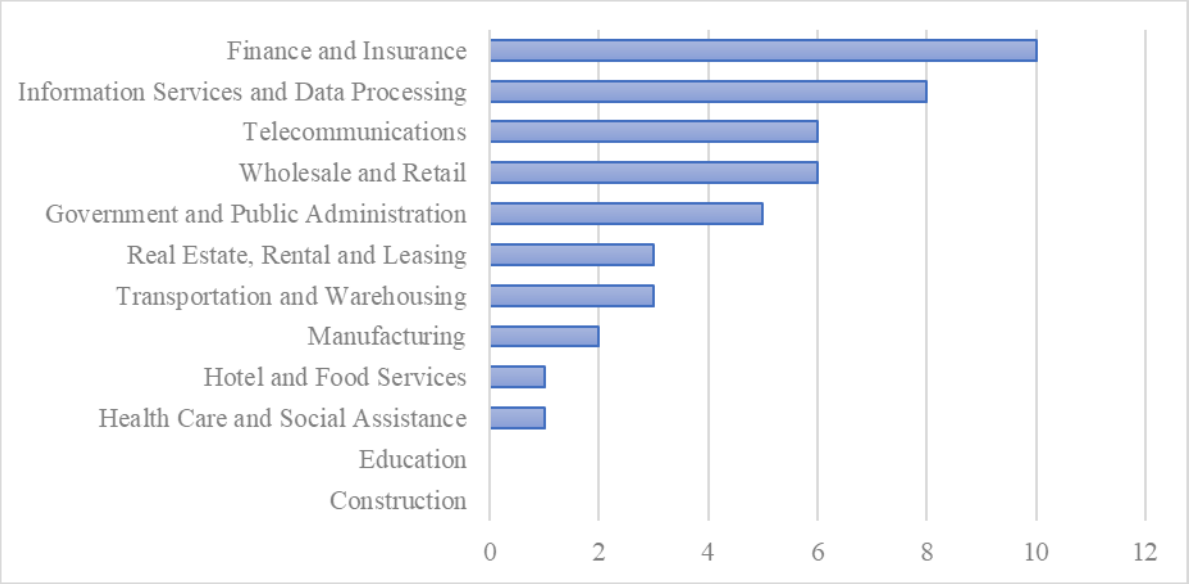

Figure 15. Sectors oriented into Big Data implementation by the auditors

The majority of audit respondents (10 out of 12 specialists) have an opinion that Finance and Insurance sector is oriented to adopt Big Data into its activities, while only 5 business entities respondents selected this sector as the most oriented. Moreover, 8 audit respondents selected Information Services and Data Processing, the result is quite similar to business entities answers as 10 respondents also selected this sector as the most oriented to adopt Big Data. The result of answers regarding the Wholesale and Retail, Telecommunication, Health Care and Social Assistance and Hotel and Food Services industries are rather similar as well.

It is important to add that audit specialists (5 from 11) selected Government and Public Administration sector as oriented one, while none of the business entities respondents selected the mentioned sector. On the other hand, none of auditors have selected Construction sector, while the majority of business entities respondents selected the mentioned sector (10 from 12 respondents). The similar situation is with Manufacturing service as only 2 audit specialists dedicated it as potential for Big Data usage, while according the business entities respondents, the mentioned sector is the most oriented from the whole listed sectors.

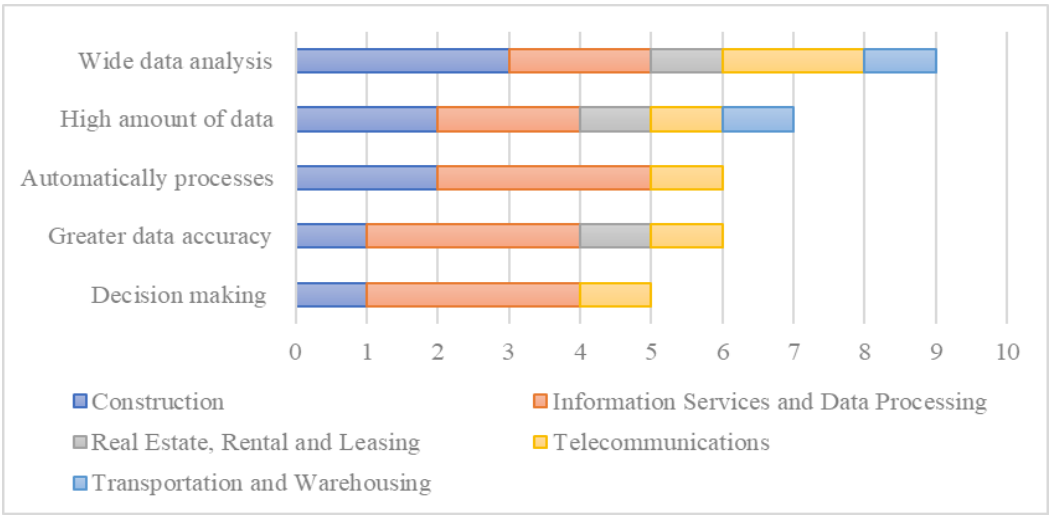

Figure 16. Reasons to use Big Data by the business entities

The following question of the survey was regarding the reasons for business companies to use Big Data. Answers and opinions by the business entities respondents were rather similar, the respondents identified that Wide data analysis and High amount of data are the most important reasons of Big Data usage. Majority of business entities experts, more princely 9 from 12 respondents specified Wide data analysis as the main reason for business companies to use Big Data. Equally, 7 from 12 respondents also selected high amount of data as the most relevant as well. On the other hand, the other reasons have reached majority of opinions too – 6 business entities respondents selected Greater data accuracy and Automatically processes, while Decision making is at the last position with 5 business entities respondents opinions. The survey answers regarding each industry are presented in the figure above.

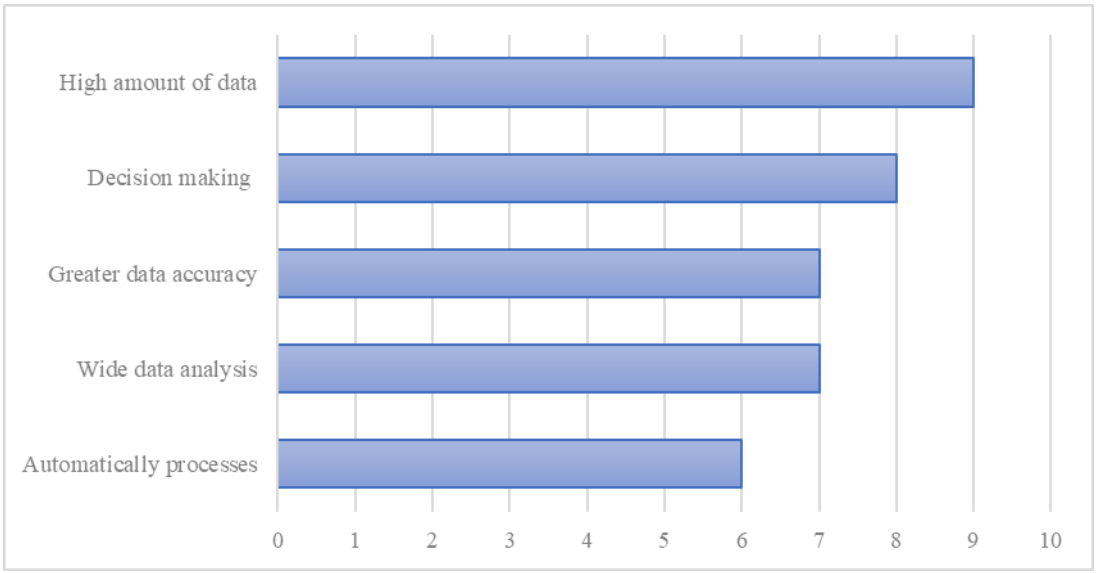

Figure 17. Reasons to use Big Data by the auditors

The audit respondents' position regarding the reasons to use Big Data differs from the respondents from business entities. In the audit respondents opinion, the main reason to use Big Data is High amount of data as 9 respondents from 11 selected this subject. Moreover, according audit specialists, Decision making is dedicated to the second position as according business entities specialist, the mentioned reason was in the last position. The audit respondents answers about other reasons to use Big Data were quite similar to business entities respondents answers.

As the sectors which are the most oriented into Big Data implementation and the reasons of the mentioned process are determined, the next question is associated to the areas in which organizations seek to implement Big Data applications and processes. The respondents were asked to select the areas in which organizations seek to adopt Big Data applications. The potential areas were listed: eCommerce, eBusiness, Online Operations, Direct and online marketing, Fraud detection / management, Customer and market analysis, Customer service, Supply Chain management and logistics, Information Technology, Finance and administration, HR and people development and Risk management. Answers collected from the business entities are provided in the figure below.

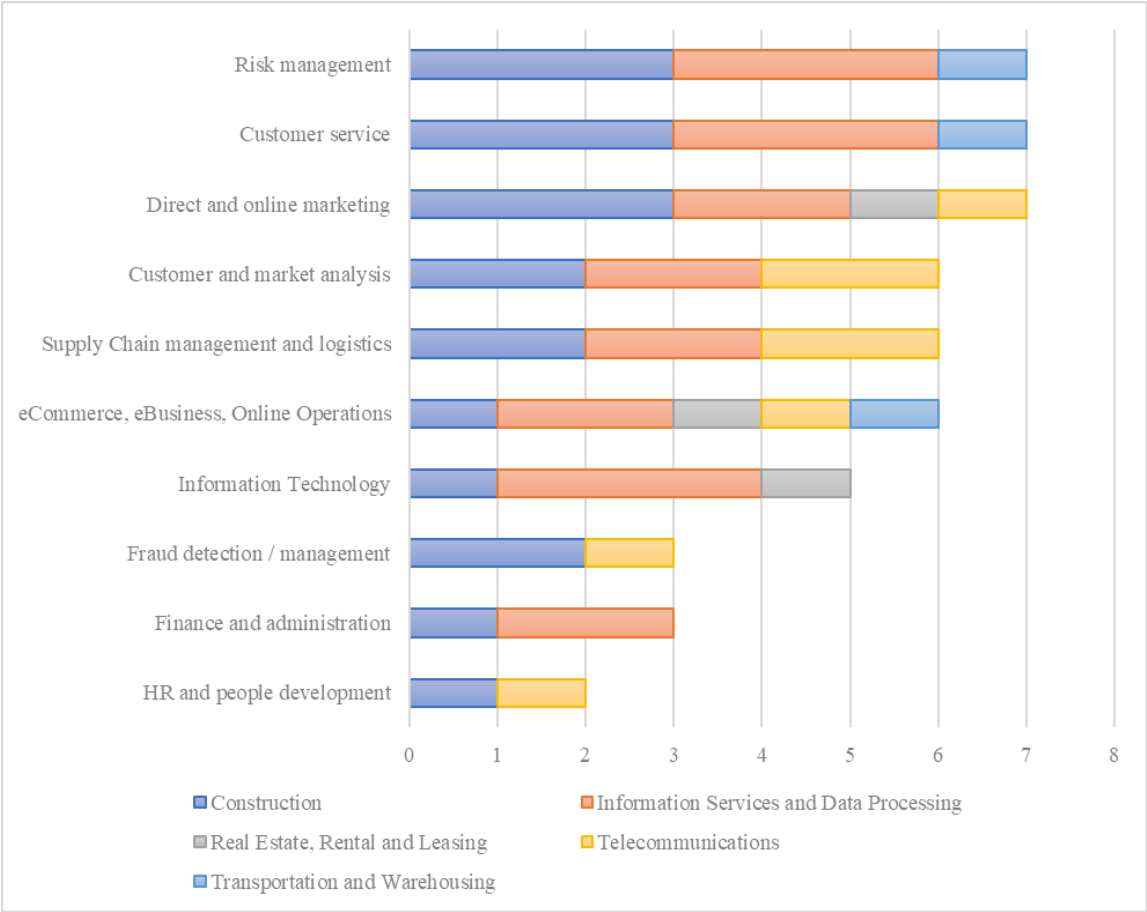

Figure 18. Areas where companys seek to adopt Big Data applications by the business entities respondents

The majority of respondents (7 respondents from 12) from business entities identified that Direct and online marketing, Customer service and Risk management are the main areas where companies seek to implement Big Data. Moreover, 6 specialists also selected Customer and market analysis, eCommerce, eBusiness, Online Operations and Supply Chain management and logistics, while 5 of them – selected Information Technology. The minority of respondents have an opinion that business entities are oriented to adopt Big Data into Finance and administration, Fraud detection and management and HR and people development fields.

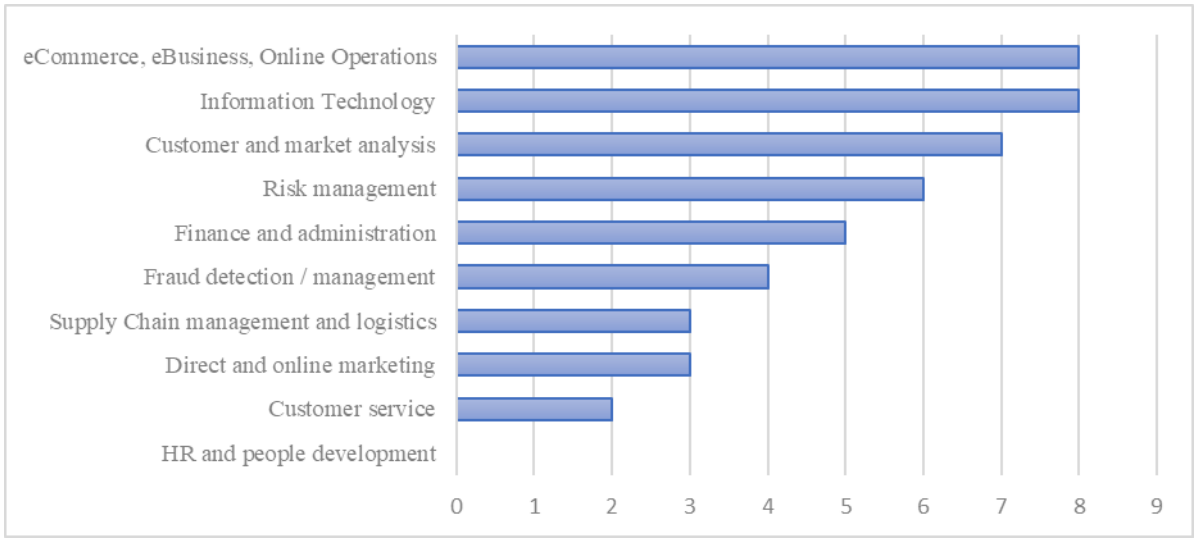

Figure 19. Areas where companies seek to adopt Big Data applications by the auditors

Auditors answers about the areas where companies seek to implement Big Data applications are rather similar than business entities. According the answers provided, the majority of auditors have an opinion that eCommerce, eBusiness, Online Operations, Information Technology and Customer and market analysis are the main fields where companies seek to implement Big Data. It is important to add, that the auditors opinion defers from business entities respondents position, as audit specialists dedicated the last positions to Customer service and Direct and online marketing, while respondents from the business entities appointed the mentioned fields to the first positions.

During the theoretical part the essential challenges for the business companies in Big Data implementation process were analyzed and explained. The mentioned challenges are Data management issues, Data security issues, Storage, Lack of human resources, Lack of technical tools and Financial budget. In order to classify the challenges regarding the priority, respondents were asked to rank the mentioned items in order of importance with #1 being the most important object to #6 being the least important object. Answers collected from the business entities are provided in the table below.

	1	2	3	4	5	6	Total	Score
Storage	5	4	0	2	0	1	12	9.50
	42%	33%	0%	17%	0%	8%		
Data management issues	4	2	2	1	1	2	12	8.17
	33%	17%	17%	8%	8%	17%		
Data security issues	0	3	4	3	1	1	12	7.17
	0%	25%	33%	25%	8%	8%		
Financial budget	2	0	3	3	2	2	12	6.50
	17%	0%	25%	25%	17%	17%		
Lack of human resources	1	2	1	0	6	2	12	5.67
	8%	17%	8%	0%	50%	17%		
Lack of technical tools	0	1	2	3	2	4	12	5.00
	0%	8%	17%	25%	17%	33%		

Table 3. Challenges for the business companies in Big Data implementation by the business entities

According to the answers, the respondents have an opinion that Storage issues are the main challenge for business entities, as summary of answers represents, 5 respondents (42%) ranked the mentioned challenge as the most important, 4 responds (33%) – identified it as the second, none of the respondents – as a third one. In the end, Storage issues were evaluated by the 9.5 score which is the highest one from the whole challenges. The second issue regarding the importance is Data management as 4 respondents (33%) identified it as the most important, while 2 respondents (17%) dedicated it to the second and third places. The final score is 8.17 which is the second one from the population analysed. Moreover, Data security issues are in the third place in the ranking table, as none of the respondents identified in it the first position from the issues, while 3 respondents (25%) dedicated it to the second position, 4 respondents (33%) – third position and also 3 respondents (25%) - to the fourth position. The final rank score of the mentioned issue is 7.7. Other issues as financial budget, lack of human resources and lack of technical tools are ranked in the last three positions regarding the significance.

	1	2	3	4	5	6	Total	Score
Data security issues	4	2	2	2	1	0	11	8.33
	36%	18%	18%	18%	9%	0%		
Data management issues	2	1	4	3	1	0	11	7.33
	18%	9%	36%	27%	9%	0%		
Lack of technical tools	4	0	2	2	1	2	11	7.00
	36%	0%	18%	18%	9%	18%		
Financial budget	0	5	0	3	2	1	11	6.50
	0%	45%	0%	27%	18%	9%		
Storage	0	3	0	1	4	3	11	4.83
	0%	27%	0%	9%	36%	27%		
Lack of human resources	1	1	2	0	2	5	11	4.67
	9%	9%	18%	0%	18%	45%		

Table 4. Challenges for the business companies in Big Data implementation by the auditors

According to the answers bu auditors, the experts have an opinion that Data security issues are the main challenge for business entities, regarding the table above, 4 audit respondents (36%) ranked the mentioned challenge as the most important, 2 audit respondents (18%) – identified it as the second position, the same amount of audit respondents ranked the mentioned challenge to third and fourth positions, while none of the respondents – as a sixth one. Summarizing the results, Data security issues was evaluated by the 8.33 score which is the highest one from the whole challenges. To compared with business entities respondents, the Data security issues was dedicated to the third position. Majority of respondents additionally commented and based their opinion on argument that cyberattacks and data breaches became a widely prevalent problem. Companies collected and store information about their activities, customer and clients in data bases, for this reason, the cyberattacks could make an impact to trust by the customers, suppliers and clients. The second issue regarding the importance is Data management as 2 respondents (18%) identified it as the most important, while 1 respondentt (9%) dedicated it to the second place, while 4 respondents (36%) - third place. The final score is 7.33 which have taken the second position. It is important to add that business entities respondents identified this issue to the second position as well.

Furthermore, Lack of technical tools is ranked as the third challenge regarding the importance of Big Data adoption. Analysing the results, 4 audit respondents (36%) have an opinion that the mentioned issue is in the first position, one of them dedicated it to the second place, while 2 audit respondents (18%) ranked it to the third and fourth places. The final score is 7.00. Lack of technical tools was identified in the last position regarding the business entities respondents. Financial budget is ranked as the fourth challenge regarding the significance of Big Data implementation. It can be seen from the table below that none of audit respondents dedicated it to the first and third positions, while 5 audit respondents (45%) ranked it to the second position, 3 respondents (27%) – fourth and 2 respondents (18%) to the fifth position. The final score is 6.50. Financial budget was ranked to the fourth place by business entities respondents as well. Furthermore, Storage was identified as fifth challenge regarding the importance. Neither of audit respondents dedicated it to the first and third positions, while 3 audit respondents (27%) ranked it to the second position, 1 respondent (9%) – fourth one and 4 respondents (26%) to the fifth position. The final score is 4.83. Storage was ranked to the first position by business entities respondents. Lack of human resources is identified to the last position – sixth one. The majority of audit respondents identified it to the sixth position.

For this reason, the final score is 4.67. The mentioned challenge was identified in the similar position by business entities respondents as well – the fifth one.

4.3.3.Big Data and Big data analytics in audit

The following section is oriented to Big Data and Big Data Analytics implementation in audit field. In order to understand the position of respondents from two different fields - the business entities and audit industry, the analysis will be prepared by testing answers of the mentioned groups separated.

The following question was more oriented into auditors and their challenges for Big Data auditing. In order to select the most important barriers to successful integration of Big Data into the audit, the respondents were asked to rank the following items in order of importance with #1 being the most important object to #4 being the least important object. The barriers for ranking were Data capture, Lack of human resources, Lack of technical tools and Breaches in audit standards and regulations. According the answers, Data capture was the barriers identified as the most relevant for successful Big Data implementation in audit field. Answers collected from the respondents are provided in the table below.

	1	2	3	4	Total	Score
Data capture	6	3	3	0	12	9.75
	50%	25%	25%	0%		
Lack of human resources	4	1	3	4	12	7.25
	33%	8%	25%	33%		
Lack of technical tools	0	6	2	4	12	6.50
	0%	50%	17%	33%		
Breaches in audit standards and regulations	2	2	4	4	12	6.50
	17%	17%	33%	33%		

Table 5. Barriers to successful integration of Big Data into the audit (business entities respondents)

Majority of business entities respondents identified Data Capture as the most important barrier for successful integration of Big Data in audit market. It could be seen from the table above, that 6 respondents (50%) ranked the mentioned subject as the first one - as the most essential, 3 respondents (25%) as the second one, and third one, while none of them identified it to the last position. The final score is 9.75 which is the highest one. Moreover, Lack of human resources is ranked to the second one position. According the gathered answers, 4 business entities respondents (33%) the mentioned subject ranked as the most significant one, 1 respondent (8%) as the second one, while 3 respondents (25%) believed that the lack of human resources is the third one and 4 respondents (33%) – the last one. The final score of the subject is 7.25. The last barriers - Lack of technical tools and Breaches in audit standards and regulations are scored by the same rank and the final score is 6.5. The detailed analysis is provided in the table above.

	1	2	3	4	Total	Score
Lack of technical tools	4	3	2	2	11	7,75
	36%	27%	18%	18%		
Breached in audit standards and regulations	3	4	1	3	11	7,25
	27%	36%	9%	27%		
Data capture	3	2	4	2	11	7,00
	27%	18%	36%	18%		
Lack of human resources	1	2	4	4	11	5,50
	9%	18%	36%	36%		

Table 6. Barriers to successful integration of Big Data into the audit (audit respondents)

According the answers received from audit specialists, lack of technical tools was ranked as the main barrier. The answers were determined as following – 4 audit respondents (36%) ranked it to the first position, 3 audit respondents (27%) – the second one, while 2 audit respondents (18%) – dedicated it to the third and fourth positions. The final score of the mentioned barrier is 7.75. Compared to answers received from the business entities, the mentioned subject was ranked to the third position. Based on collected answers, Breaches in audit standards and regulations was identified as second barrier for successful integration of Big Data in audit market. As it can be seen from the table, 3 audit respondents (27%) ranked it to the first and fourth positions, while 4 audit respondents (36%) – the second position and only one – to the third one. The final score of this barrier is 7.25. Compared to business entities answers, the Breaches in audit standards and regulations was ranked to the fourth position. According the answers from audit specialists, data capture was ranked as the third barrier – 3 audit respondents (27%) ranked it as the most important subject, 2 audit respondents (18%) – as the second one and fourth one, while 4 respondents (36%) dedicated it to the third position. The final score of the mentioned barrier is 7.00. The respondents from the business entities have totally different opinion – ranked Data capture as the most significant barrier to successful adoption of Big Data in audit field. In the end, Lack of human resources was ranked to the last position from the whole barriers. The final score of this subject reached 5.50. It is important to add, that respondents from the business entities ranked this barrier to the second position.

	1	2	3	4	5	6	Total	Score
Sample -based analysis expansion	6	2	3	1	0	0	12	15.25
	50%	17%	25%	8%	0%	0%		
More relevant business insights	2	2	3	2	2	1	12	11.25
	17%	17%	25%	17%	17%	8%		
Significant risk identification of particular account	3	1	3	2	1	2	12	11.25
	25%	8%	25%	17%	8%	17%		
Capability to analyse sub-ledger information	2	3	0	3	2	2	12	10.50
	17%	25%	0%	25%	17%	17%		
Identification of controls efficiency	0	2	2	1	5	2	12	8.25
	0%	17%	17%	8%	42%	17%		
Better understanding the entity and its environment	0	1	1	3	2	5	12	6.75
	0%	8%	8%	25%	17%	42%		

Table 7. Advantages of Big Data Analysis for audit (business entities respondents)

The next question of the questionnaire is oriented to audit field as well. In order to identified the main advantages of Big Data analysis for audit, the respondents were asked to rank items in order of importance with #1 being the most important object to #6 being the least important object. The

following items for ranking are Capability to analyze sub-ledger information, Sample -based analysis expansion, More relevant business insights, Better understanding the entity and its environment, Identification of controls efficiency and Significant risk identification of particular account.

On the strength of received opinions from business entities, Sample - based analysis expansion is identified as the most significant advantage of Big Data analysis. Majority of inquired business entities respondents (50%) ranked the subject as the most important, 2 respondents (17%) identified it as worth the second place, 3 respondents (25%) dedicated it to the third position. The total score of the ranked item is 15.25 which is the highest one from the whole ranked subjects. The second advantage of Big Data analysis for audit is more relevant business insights. The total score of the ranked subject is 11.25. Regarding the results analysis, the mentioned subject was identified as the most important advantage by the 2 respondents (17%), while also 2 respondents (17%) ranked the mentioned item as the second one and 3 respondents (25%) as the third one. Significant risk identification of particular account was ranked as the third subject by importance of advantages of Big Data analysis for audit. Based on received questionnaire answers, 3 experts (25%) ranked it as the main advantage, while one expert (8%) dedicated to it the second position, 3 experts (25%), third position, 2 experts (17%) – fourth position. The total score of the analyzed advantage is 11.25, which is the same as More relevant business insights.

Capability to analyse sub-ledger information is the advantage of Big Data analysis for audit process which was identified in the fourth position regarding the research. The total score of the ranked item is 10.50. As it can be seen from the table below, 2 respondents (17%) ranked it as the first one advantage, 3 respondents (25%) – second one, 2 respondents (24%) – fourth one, while 2 respondents (17%) – dedicated it to fifth position. According the collected opinion, identification of controls efficiency is in the fifth position regarding the importance of advantages of Big Data analysis for audit process. The final score of the advantage is 8.25. None of respondents have an opinion, that the mentioned advantage should be ranked as the most important, 2 respondents (17%) – ranked it as the second one and third one, while the majority ranked it in the last two positions regarding the importance. At least, in the last position of ranking is better understanding the entity and its environment. No respondents ranked it in the first position, 1 respondents (8%) ranked in the second and third position, while the majority of respondents (42%) dedicated it to the last position.

	1	2	3	4	5	6	Total	Score
Capability to analyse sub-ledger information	4	4	1	0	0	2	11	12.50
	36%	36%	9%	0%	0%	18%		
Sample -based analysis expansion	2	3	1	2	3	0	11	10.75
	18%	27%	9%	18%	27%	0%		
Significant risk identification of particular account	2	0	2	4	2	1	11	9.25
	18%	0%	18%	36%	18%	9%		
More relevant business insights	2	0	3	2	1	3	11	8.75
	18%	0%	27%	18%	9%	27%		
Better understanding the entity and its environment	0	2	3	2	2	2	11	8.50
	0%	18%	27%	18%	18%	18%		
Identification of controls efficiency	1	2	1	1	3	3	11	8.00
	9%	18%	9%	9%	27%	27%		

Table 8. Advantages of Big Data Analysis for audit (audit respondents)

As it could be seen from the table above, audit respondents have different opinion compared to respondents from business entities. According the audit specialists, the main advantage of Big Data Analysis is Capability to analyse sub-ledger information. The total score of the ranked item is 12.50 which is the highest one from the whole ranked subjects. According the collected ranking answers, the majority of audit specialists (36%) have ranked the mentioned subject to the first and second positions, while only one specialists (9%) ranked in to the third one and none of the respondents – to fourth and fifth positions. It is important to add, that business entities respondents have a different opinion and the mentioned subject was ranked to the fourth position. Sample -based analysis expansion was ranked as the second one advantage for Big Data Analysis. The total score of the mentioned advantage is 10.75. Based on received questionnaire answers, 2 audit respondents (18%) identified this subject to the first position, 3 audit respondents (27%) – second one and only one (9%) – third position. Business entities respondents classified the mentioned subject as the most important advantage. Significant risk identification of particular account was ranked to the third place. To be more detailed, 2 audit respondents (18%) ranked the mentioned advantage to the first, third and fifth places, while the majority of the respondents (36%) dedicated it to the forth position. The final score of this subject is 9.25. In this case, the business entities and audit respondents opinions coincide as the both groups ranked significant risk identification of particular account as the third advantage of Big Data Analysis for audit. Based on the answers received, More relevant business insights advantage was ranked to the forth position regarding the importance. It can be seen from the table below, that 2 audit respondents (18% ranked this subject to the first and fourth positions, while 3 respondents (27%) – third and sixth positions and only one (9%) – to the fifth one. The total score of the subject is 8.75. To compared with answers of business entities respondents, the mentioned advantage was ranked to the second place. Better understanding the entity and its environment was ranked to the fifth place. The total ranked score is 8.50. The answers of the respondents were allocated by the following way: no one of the audit respondents ranked it to the first position, while 2 specialists (18%) ranked it to second, fourth, fifth and sixth positions, while 3 audit specialists (27%) dedicated it to the third place. The answer by the business entities respondents was quite similar – Better understanding the entity and its environment was identified as the sixth advantage. At least, Identification of controls efficiency was ranked as the last advantage of Big Data Analysis for audit. Majority of inquired audit respondents (27%) ranked the subject to the last two positions, while 1 respondent (9%) dedicated it to the first, third and fourth positions and finally 2 respondents (18%) have an opinion that this subject is the second one regarding the importance. The answer by the business entities respondents was also quite similar – the mentioned subject was ranked to the fifth place.

The last open question was for general respondents opinions about the future of Big Data adoption into the business and its audit process. The majority of opinions were quite similar, as the respondents have an insights regarding the advantages of Big Data and Big data analytics usage. The following opinions were revived:

— Big Data will be developed and become the main thing in business running.
— Struggle to align one solution to wide variety of companies will remain an important issue. However, Big Data solutions should be implemented and developed in order to increase efficiency of audit process.
— As data sets constantly continue to grow, audit process starts getting more complex and requires more advanced tools to work with Big Data.

— Implementation of Big Data into business processes is just the matter of time and resources as advantages of Big Data Analysis is being used by different sector. Biggest rival of Big Data in audit sector in my opinion is security.
— Using Big Data, companies get a much better understanding of customers' needs. Then, this knowledge will allow companies to tailor its products or services. And, finally, companies are able to improve their marketing strategy. All these factors lead to an increasing number of clients, better retention rates and higher profits.

4.4. Insights of the research

The last section of the research part of the Master Thesis consists of the insights regarding the Big Data and Big Data Analytics adoption in Lithuanian audit market. The questionnaire answers by auditors and business entities specialists were quite different. The contrast between the answers demonstrate the different overview of the topic by diverse specialists. Auditors and Business entities might have dissimilar opinions about Big Data and Big Data Analytics implementation in Lithuanian audit market as each group of respondents evaluate the topic through their own perspective.

Sectors oriented into Big Data implementation		
	Auditors	**Business Specialists**
1	Finance and Insurance	Manufacturing
2	Information Services and Data Processing	Construction
3	Wholesale and Retail	Information Services and Data Processing
4	Telecommunication	Wholesale and Retail
5	Government and Public administration	Transportation and Warehousing

Table 9. Comparison of opinions regarding the sectors oriented into Big Data implementation

For instance, the opinions regarding the sectors oriented into Big Data implementation were totally dissimilar. The opinions coincided for Information Services and Data Processing and Wholesale and Retail industries as they were selected by both groups of the respondents. On the other hand, auditors opinions for other industries diverse from business entities outlook. One of the reason of diverse opinions could be that each respondent have an overview though its own perspective. For instance, Auditors understand that Big Data implementation have an impact to auditors, accounting and tax services providers, for this reason they selected Finance and Insurance sector. On the other hand, auditors have practical experience with various clients, their controls and value of data. Business entities have opinion based on their own experience and perspective as well. Their understanding of intense to use Big Data is more individually and dispose to particular business entity – their workplace.

Reasons to use Big Data		
	Auditors	**Business Specialists**
1	High amount of data	Wide data analysis
2	Decision making	High amount of data
3	Wide data analysis	Automatically processes
4	Greater data accuracy	Greater data accuracy
5	Automatically processes	Decision making

Table 10. Comparison of opinions about reasons to adopt Big Data

Opinions of both groups about the main reasons to use Big Data in business entities were quite similar. Respondents have an opinion that High amount of data and Wide Data analysis are the main reasons why companies adopt Big Data to their systems. The opinions were unsimilar for Decision making reason, as business entities respondents selected it for the last position, while auditors – the second one. The reason of different opinions might be the same as it was mentioned above, both groups of the respondents answered based on their own intentions. For instance, auditors believed that Big Data provide more opportunities for auditors to make decisions regarding the companies activities, while business entities preferred other reasons as the most important.

Areas where companies seek to adopt Big Data applications		
	Auditors	**Business Specialists**
1	eCommerce, eBusiness, Online Operations	Direct and online marketing
2	Information Technology	Customer service
3	Customer and market analysis	Risk management
4	Risk management	Customer and market analysis
5	Finance and administration	eCommerce, eBusiness, Online Operations

Table 11. Comparison of opinions about areas where companies seek to adopt Big Data applications

As it could be seen from the table above, the respondents opinions about areas where companies seek to adopt Big Data applications were diverse as well. The responses quite similar regarding eCommerce, eBusiness, Online Operations, Customer and market analysis and Risk management fields, as both groups of respondents identified the mentioned areas as the significant ones. On the other hand, answers regarding the other areas were different.

Challenges for the business companies in Big Data implementation		
	Auditors	**Business Specialists**
1	Data security issues	Storage
2	Data management issues	Data management issues
3	Lack of technical tools	Data security issues
4	Financial budget	Financial budget
5	Storage	Lack of human resources
6	Lack of human resources	Lack of technical tools

Table 12. Comparison of opinions about challenges for the business companies in Big Data implementation

Responses about challenges for the business companies in Big Data implementation were quire similar. The following subjects - Data security issues, Data management issues, Financial budget and Lack of human resources were ranked in the similar positions, while opinions about Storage were unsimilar. Business entities respondents identified this subject to the firs position, when auditors – the fifth one. Moreover, auditors ranked Lack of technical tools to the third position, when business entities respondents dedicated it to the last position.

Barriers to successful integration of Big Data into the audit		
	Auditors	**Business Specialists**
1	Lack of technical tools	Data capture
2	Breaches in audit standards and regulations	Lack of technical tools
3	Data capture	Breaches in audit standards and regulations
4	Lack of human resources	Lack of human resources

Table 13. Comparison of opinions about barriers to successful integration of Big Data into the audit

Moreover, answers about barriers to successful integration of Big Data into the audit were also similar. Auditors and business entities specialists identified Lack of technical tools, Breaches in audit standards and regulations and Lack of human resources to the similar ranking positions. The only one difference was specified for Data capture as auditors ranked it to the third position, when business entities respondents – to the first one. In this case, auditors answers are based on experience and practical knowledge, while business entities are able to identify the risks and barriers only through general understanding. The majority of business entities are audited, but their understanding is narrow and based on their sector audit.

Advantages of Big Data Analysis for audit		
	Auditors	**Business Specialists**
1	Capability to analyse sub-ledger information	Sample -based analysis expansion
2	Sample -based analysis expansion	More relevant business insights
3	Significant risk identification of particular account	Significant risk identification of particular account
4	More relevant business insights	Capability to analyse sub-ledger information
5	Better understanding the entity and its environment	Identification of controls efficiency
6	Identification of controls efficiency	Better understanding the entity and its environment

Table 14. Comparison of opinions about advantages of Big Data Analysis for audit

The final question about Big Data Analytics in Audit field was regarding the advantages of Big Data Analysis for auditors. Opinions about the majority of advantages were the same. Both business entities respondents and audit specialists confirmed that the main advantages are sample -based analysis expansion and significant risk identification of particular accounts. Moreover, both groups of respondents also ranked to the lowest ranges subjects as better understanding the entity and its environment and identification of controls efficiency. The respondents agreed that the mentioned advantages are less important compared others. On the other hand, audit specialists dedicated capability to analyse sub-ledger information to the first position from all listed advantages, when business entities specialists ranked it to the fourth position. Also, auditors have an opinion that More relevant business insights is in the middle in the ranking, while business entities respondents ranked it to the second one position. The opinions about the advantages of Big Data Analytics in Audit field were similar as expectations from the business entities and auditors understanding of audit processes improvement are rather related.

According the answers received, it could be concluded that Big Data and Big Data Analytics value added and opportunities provided are evaluated in different perspectives by respondents by business entities and auditors. The respondents are highly experienced and qualified to provide appropriate opinions which are based on their own review. In my opinion, auditors need to improve their development in technologies according the clients, their needs and systems used. As the Big Data implementation is just a matter of time, the availability to analyse it could be a huge advantage in competitive audit market.

Conclusions

1. The theoretical analysis of Big Data disclosure that Volume, Variety, and Velocity are the main dimension of Big Data definition. The word Big in Big Data perspective determine the volume of data. It is predicted that according the rapidly growing of data, in nearby future, it supposed to be collected in zettabytes instead of petabytes. In Big Data perspective, velocity is an intention that deals with the data speed which is coming from various data sources. For this reason, traditional systems are not available to perform the analytics on the data that is continually in movement. Variability of data is related to inconsistencies of the data flow. The data variety consists of six characteristics: formats, structure, unstructured data, natural language, media, complex systems.
2. The theoretical review of Big Data processing, identified two the most appropriate Big Data techniques - MapReduce and Apache Spark. After evaluation of Processing Method, Time Efficiency, Error Recovery and Memory requirements, the conclusion was made that Apache Spark could be suggested as the most suitable alternative for the Big Data application and processing.
3. Based on the theoretical review of the literature, Big Data could be divided into the main four phases: data generation, data acquisition, data storage and Big Data Analytics. Data generation is called as the first step of Big Data implementation and is related to data collection from various types of resources. Moreover, data acquisition process includes data collection, data transmission and data pre-processing. Data storage is oriented to the data keeping, device is used to store, manage and analyse data. Data analysis is the last and the most significant part in the value phases of Big Data, which the main aim is extract useful values and provide suggestions or decisions for organization activities. Data analysis consist of two types – Traditional and Big Data analysis. Traditional data analysis is more statistical and mathematical, while Big Data analysis is more oriented to high amount of data which is used in the organizations. Moreover, the mentioned analysis is more complex process which cover the large volume of data to make some trends, forecasts and some assumptions to make decisions.
4. During the theoretical evaluation, the main Big Data issues were identified: data management, privacy and security, data access and sharing of information, storage and tools and talent issues. Big Data requires investments on IT systems, its storage, processing process, human resources and various other tools. Big Data is formed by multiple heterogeneous sources with totally different nature and representation. For this reason, data management process requires innovative, modern, high performance and multi-dimensional management tools, otherwise, the results could be unappropriated. One of the most important issue with Big Data is privacy and security that includes conceptual, technical as well as legal significance. Personal information is sensitive data of a person and could be inference of additional facts about the person when data is combine with external large data sets. Data sharing about Company‘s clients and internal procedures is harmful for secrecy and competitiveness of the business. Therefore, Big Data implementation involve the whole organization and personnel, for this reason, employees need to be trained and prepared.
5. According the theoretical review, the purpose of audit is to perform particular procedures and provide auditor’s opinion about financial statements. Lithuanian Financial Audit is reglamented by the the audit requirements specified in Law on the Audit of Financial Statements. .During the planning stage, the auditor’s targets are to pay attention to the most important areas, recognize potential problematic areas and solve it on a timely basis and be responsible for

effective and efficient organization of the engagement. The analysis of financial statements is performed by the audit procedures which are based on accounts. The last stage of the audit process is to provide auditor's opinion on Financial Statements.

6. According the research, auditors have an opinion that Finance and Insurance, Information Services and Data Processing, Wholesale and Retail, Telecommunication and Government and Public administration sectors are the most oriented to Big Data adoption. The respondents of business entities have different opinions, their most oriented sectors are Manufacturing, Construction, Information Services and Data Processing, Wholesale and Retail and Transportation and Warehousing. Moreover, both groups of respondents agreed that High amount of data, Wide data analysis and Greater data accuracy are the most important reasons to implement Big Data in business systems. In terms of areas, where companies seek to adopt Big Data applications, auditors and respondents from business entities agreed that these fields are eCommerce, eBusiness, Online Operations, Customer and market analysis and Risk management. Additionally, respondents of both groups believe that the most challenging issues for Big Data implementation is companies are Data security issues and Data management issues.
7. After qualitative research, it was identified that Lack of technical tools is ranked to the most significant barrier to successful integration of Big Data into the audit, while Breaches in audit standards and regulations was ranked to the last positions as less important subject. Moreover, both respondents group have an opinion that the most important Advantages of Big Data Analysis for auditors are Sample -based analysis expansion and Significant risk identification of particular account. On the other hand, the opinions differed regarding the advantage of capability to analyse sub-ledger information, auditors selected this advantage as the most important, while business entities - identified this subject as less significant. Opinions about Big Data and Big Data Analytics implementation in audit market differed as auditors and business entities respondents reviewed to the topic from different perspectives.

Recommendations. Based on the research results, several recommendations for companies were prepared:

1. Big Data and Big Data analysis might be an advantage for better understanding of clients, suppliers, customer needs and their expectations. It is important to add that the whole company needs to be prepared for this implementation in the system. Management of the companies should provide Big Data trainings for personnel and introduce the value, opportunities and main problems by using Big Data.
2. Big Data implementation between business entities is growing in particular sectors, for this reason Big Data Analytics usage could be a competitive advantage against the other audit companies.
3. Expectations for audit companies and auditors is quite high and required to be in step with new technologies development. Audit companies should provide trainings for auditors regarding Big Data Analytics usage. Moreover, audit companies need to pay attention to software which could be mandatory for Big Data engagements.

List of references

1. Cambridge University. *Cambridge Dictionary*. Retrieved from: https://dictionary.cambridge.org/dictionary/english/big-data
2. ACCA. (2016). *Professional Accountants – the Future: Drivers of Change and Future Skills*. Retrieved from ACCA Global: https://www.accaglobal.com/an/en/technical-activities/technical-resources-search/2016/june/professional-accountants-the-future-report.html
3. ACCA. (2019). *Machine Learning: More Science than Fiction*. Retrieved from ACCA Global: https://www.accaglobal.com/content/dam/ACCA_Global/professional-insights/machine-learning/pi-machine-learning-report.pdf.
4. ACCA. (2019). *Audit and Technology*. Retrieved from ACCA Global: https://www.accaglobal.com/us/en/professional-insights/technology/audit-and-tech.html.
5. Agresti A. (2018). *An Introduction To Categorical Data Analysis*. Wiley.
6. Apache Storm. (2015). Retrieved from Apache Storm: https://storm.apache.org/
7. Ater B., Gimbar C., Jenkins G., Saucedo G. and Wright N. (2018). Audit roles and the review process: workpaper preparers' and reviewers' differing perspectives. Retrieved from Emerald Insight: https://www.emerald.com/insight/content/doi/10.1108/MAJ-05-2018-1896/full/html
8. Bajaj R. H. and Ramteke P. L. (2014). *Big Data - The New Era of Data*. Retrieved from International Journal of Computer Science and Information Technologies: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.639.5363&rep=rep1&type=pdf.
9. Bank of Lithuania. (2019). *Lithuanian Economic review*. Retrieved from Bank of Lithuania: https://www.lb.lt/en/publications/lithuanian-economic-review-september-2019
10. Barta G. (2018). *The increasing role of IT auditors in financial audit: risks and intelligent asnwers*. Retrieved from Business Management and Education Vol 16: https://journals.vgtu.lt/index.php/BME/article/view/2142
11. Beyer M. and Laney D. (2012). *The importance of Big Data : a definition*. Gartner. Retrieved from: https://www.gartner.com/doc/2057415.
12. Chen G., Cai Q. and Zhan Y. (2016). *Approaches on personal data privacy preserving in cloud: a survey*. Proceedings of The Third International Conference on Data Mining, Internet Computing, and Big Data. Retrieved from: https://www.scribd.com/document/319294948/Approaches-on-Personal-Data-Privacy-Preserving-in-Cloud-A-Survey
13. Corbett P., Schroeck C. and Schroeck R. (2013). *Analytics: The real-world use of big data*. IBM Global Business Services Business Analytics and Optimization. Retrieved from: https://www.ibm.com/downloads/cas/LKMQWLPY
14. CSA Big Data working Group. (2013). *Expanded top ten Big Data Security and Privacy Challenges*. CSA Cloud Security Alliance. Retrieved from Big Data Working Group: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf
15. Dagilienė L. and Klovienė L. (2019). Motivation to use Big Data and Big Data Analytics in external auditing.

16. Davenport T. and Harris H. (2017). *Competing on Analytics: The New Science of Winning, rev. ed*. Boston, MA, USA: Harvard Business Review Press.
17. Deloitte. (2018). *For internal audit, big data represents a big opportunity* . Retrieved from: https://deloitte.wsj.com/cio/2018/02/06/for-internal-audit-big-data-represents-a-big-opportunity/
18. Early C. E. (2015). *Data Analytics in Auditing: Opportunities and Challenges*. Business Horizons Journal, Iss5, 493-500.
19. EY. (2015). *How Big Data and Analytics are transforming the audit.* Retrieved from EY Global: https://www.ey.com/en_gl/assurance/how-big-data-and-analytics-are-transforming-the-audit
20. EY. (2017). *Big Data and Analytics in audit process.* Retrieved from EY Global: https://www.ey.com/Publication/%20vwLUAssets/ey-big-data-and-analytics-in-the-audit-process/$FILE/ey-big-data-and-analytics-in-the-audit-process.pdf
21. EY. (2019). *Unlocking the power of data to improve health outcomes: five trends to watch*. Retrieved from EY Global: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/life-sciences/life-sciences-pdfs/ey-unlocking-the-power-of-data-to-improve-health-outcomes.pdf
22. Eriksonas L. (2019). *Lithuania's capital city Vilnius becomes a top destination for foreign direct investment in Fintech.* China CEE Institute. Retrieved from: https://china-cee.eu/wp-content/uploads/2019/11/2019e1086-%EF%BC%8810%EF%BC%89-Lithuania.pdf
23. Espinosa J., Kaisler S., Armour F. and Money V. (2019). *Big Data Redux: New Issues and Challenges Moving Forward.* Proceedings of the 52nd Hawaii International Conference on System Sciences. Retrieved from: https://scholarspace.manoa.hawaii.edu/bitstream/10125/59546/0106.pdf
24. European Court of auditors. (2020). *Big Data and Digital Audit.* Retrieved from: https://www.eca.europa.eu/en/Pages/NewsItem.aspx?nid=13397
25. Fatt Q. K. and Ramadas A. (2018). The Usefulness and Challenges of Big Data in Healthcare. *Journal of Healthcare Communications*.
26. FDi Intellligence. (2019). Retrieved from Tech Start-up FDI Attraction Index 2019: https://www.fdiintelligence.com/article/75880
27. Fujitsu. (2012). *Big Data: The definitive guide to the revolution in business analytics.* Fujitsu Services Ltd. 32-40.
28. Gandomi A. and Haiden M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, Iss. 2, 137-144.
29. Griffin P. A. and Wright A. M. (2015). *Commentaries on Big Data's importance for accounting and auditing*. Accounting Horizons: Vol. 29, No. 2, 377-379.
30. Gu L. and Li H.. (2013). *Memory or time: Performance evaluation for iterative operation on hadoop and spark*. Retrieved from: https://ieeexplore.ieee.org/abstract/document/6831988
31. Hassani H., Huang X. and Silva E. (2017). Digitalisation and Big Data Mining in Banking. *Big Data and Cognitive Computing Journal, Volume 2, Iss. 3*.
32. International Financial Reporting Standards. Conceptual Framework for Financial Reporting. Retrieved from: https://www.ifrs.org/issued-standards/list-of-standards/conceptual-framework

33. International Standard on auditing 700. *Froming an opinion and reporting on financial statements 700 (Revised).* Retrieved from: https://www.ifac.org/knowledge-gateway
34. International Standards on Auditing 200. *International Standards on Auditing.* Retrieved from: https://www.ifac.org/knowledge-gateway
35. International Standards on Auditing *315 (Revised).* Retrieved from: https://www.ifac.org/knowledge-gateway
36. International Standart on Auditing, 300. *Planning and Audit of Financial Statements.* Retrieved from: https://www.ifac.org/knowledge-gateway
37. International Standart on Auditing, 320. *Audit Materiality.* Retrieved from: https://www.ifac.org/knowledge-gateway
38. Invest Lithuania. (2020). *The Fintech Landscape in Lithuania Report 2019-2020*. Vilnius: Invest Lithuania. Retrieved from: https://investlithuania.com/fintech-report-2019/
39. Kaisler S., Armour F., Espinosa A, Kaisler V.M. and Armour S. (2013). *Big data: Issues and challenges moving forward*. 2013 46th Hawaii International Conference on System Sciences.
40. Katal A., Wazid M. and Goudar R. H.. (2013). *Big Data: Issues, Challanges, Tools and Good Practices*. 2013 Sixth International Conference on Contemporary Computing (IC3).
41. KPMG. (2017). *Audit 2025, the future is now.* Retrieved from Forbes insights: https://assets.kpmg/content/dam/kpmg/us/pdf/2017/03/us-audit-2025-final-report.pdf
42. Laney D.. (2001). *Application Delivery Strategies*. Retrieved from META Group: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
43. Seimas of the Republic of Lithuania. (1999). *Law on the Audit of Financial Statements* (15 June 1999 No VIII-1227). Retrieved from: https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/40997e20566b11e78869ae36ddd5784f?jfwid=7cihrgzru
44. Lyko K., Nitzschke M. and Ngomo A. C. N. (2016). *Big Data Acquisition*. New Horizons for a Data-Driven Economy. 39-61
45. Lithuanian Chamber of Auditors. (2019). *Annual Report.* Vilnius: Lithuanian Chamber of Auditors . Retrieved from: https://lar.lt/www/new/page.php?25
46. Lukoianova T. and Rubin V. L. (2013). *Veracity Roadmap: Is Big Data Objective, Truthful and Credible?* Retrieved from: https://journals.lib.washington.edu/index.php/acro/article/view/14671/12311
47. McKinney E., Yoos C. J. and Snead K. (2017). The need for 'skeptical' accountants in the era of Big Data. *Journal of Accounting Education, Vol. 38,* 63-80.
48. McNulty E. (2014). *Understanding Big Data: The seven V's.* Retrieved from: https://dataconomy.com/2014/05/seven-vs-big-data/
49. Moura J. and Serrao. (2015). *Security and Privacy Issues of Big Data*. Retrieved from: https://arxiv.org/ftp/arxiv/papers/1601/1601.06206.pdf
50. Mukherjee S. and ShawR. (2016). Big Data – Concepts, Applications, Challenges and Future Scope. *International Journal of Advanced Research in Computer and Communication Engineering*. Vol. 5, Issue 2.
51. National Institute of Standards and Technology. (2015). *NIST Big Data Interoperability Framework: Volume 1, Definitions*. Retrieved from:

https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf

52. Ohlhorst F J. (2013). *Big Data Analytics.* Hoboken, New Jersey: John Wiley & Sons, Inc. 55-62.
53. Padgavankar M. H. and Gupta S. R. (2014). *Big Data Storage and Challenges.* International Journal of Computer Science and Information Technologies, Vol. 5. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.661.538&rep=rep1&type=pdf
54. Savitz E. (2012). *10 critical tech trends for the next five years*. Retrieved from: http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years/
55. Savitz E. (2012). *Top 10 strategic technology trends for 2013*. Retrieved from: http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technology-rends-for-2013/
56. Snijders Ch., Matzat U. and Repis U. D.. (2012). *Big Data: Big Gaps of Knowledge in the Field of Internet Science.* International Journal of Internet Science. Volume. 1-5.
57. *Statista*. (2019). *Combined revenue of the Big Four accounting / audit firms worldwide from 2009 to 2018*. Retrieved from: https://www.statista.com/statistics/473959/big-four-accounting-firms-global-combined-revenue/
58. *Statista*. (2019). *Nuskaityta iš Forecast revenue big data market worldwide 2011-2027*: https://www.statista.com/statistics/254266/global-big-data-market-forecast/
59. *Statista*. (2019). *Revenue of the Big Four accounting / audit firms worldwide in 2019, by geographical region*. Retrieved from: https://www.statista.com/statistics/250944/big-four-accounting-firms-geographical-breakdown-of-revenues/
60. Tang J. and Karim K. (2018). *Financial fraud detection and big data analytics – implications on auditors' use of fraud brainstorming session.* Managerial Auditing Journal. Volume 34.
61. *The Apache Software Foundation Blog*. (2014). Retrieved from: https://blogs.apache.org/foundation/?date=20140820&cat=general
62. The Institute of Internal Auditors. (2017). *Global perspective and insights: Internal Audit and External Audit.* Retrieved from: https://businessdegrees.uab.edu/blog/internal-auditing-vs-external-auditing/
63. Wani M. A. and S. Jabin. (2018). *Big Data: Issues, Challenges, and Techniques in Business Intelligence.* Big Data Analytics. 613-628
64. Wu X., Zhu X. Wu G. Q. and Ding W. (2014). *Data mining with big data.* IEEE Transactions on Knowledge and Data Engineering. Volume: 26, Issue 1.
65. Zhang, J., Yang, X. and Appelbaum, D. (2015). *Toward Effective Big Data Analysis in Continuous Auditing*. Accounting Horizons, Vol. 29 No. 2, 469-476.

Appendix 1

Questionnaire:

1. **Which of the following categories best describes the industry you work in?**
   - — Transportation and Warehousing
   - — Telecommunications
   - — Information Services and Data Processing
   - — Finance and Audit
   - — Real Estate, Rental and Leasing
   - — Legal Services
   - — Construction
   - — Other

2. **Year of work experience:**
   - — Under 1
   - — 1 to 3
   - — 3 to 7
   - — 7 to 10
   - — Over 10
3. **Which of the following positions best describes your role(s)?**
   - — Manager (Business unit or department)
   - — Senior Specialist/ Senior Team member
   - — Specialist/Team member
   - — Other (please specify)
4. **Which sectors could be the most oriented into Big Data implementation in Lithuania?**
   - — Construction
   - — Manufacturing
   - — Wholesale and Retail
   - — Transportation and Warehousing
   - — Telecommunications
   - — Information Services and Data Processing
   - — Finance and Insurance
   - — Real Estate, Rental and Leasing
   - — Education
   - — Health Care and Social Assistance
   - — Hotel and Food Services
   - — Government and Public Administration
   - — Other industry (please, specify)
5. **What are the main reasons for business companies to use Big Data?**
   - — High amount of data
   - — Automatically processes
   - — Greater data accuracy
   - — Decision making
   - — Wide data analysis
   - — Other

6. **In your opinion, in which areas do organizations seek to implement Big Data applications?**
    - — eCommerce, eBusiness, Online Operations
    - — Direct and online marketing
    - — Fraud detection / management
    - — Customer and market analysis
    - — Customer service
    - — Supply Chain management and logistics
    - — Information Technology
    - — Finance and administration
    - — HR and people development
    - — Risk management
    - — Other (please specify)
7. **In your opinion, what are the main challenges for the business companies in Big Data implementation process? Please rank the following items in order of importance with #1 being the most important object to #6 being the least important object.**
    - — Data management issues
    - — Data security issues
    - — Storage
    - — Lack of human resources
    - — Lack of technical tools
    - — Financial budget
8. **What are the main barriers to successful integration of Big Data into the audit? Please rank the following items in order of importance with #1 being the most important object to #4 being the least important object.**
    - — Data capture
    - — Lack of human resources
    - — Lack of technical tools
    - — Breached in audit standards and regulations
9. **What are the advantages of Big Data analysis for audit? Please rank the following items in order of importance with #1 being the most important object to #6 being the least important object.**
    - — Capability to analyse sub-ledger information
    - — Sample -based analysis expansion
    - — More relevant business insights
    - — Better understanding the entity and its environment
    - — Identification of controls efficiency
    - — Significant risk identification of particular account
10. **What is your opinion regarding the future of Big Data implementation into the business and its audit process?**