UNIVERSITETI I EVROPËS JUGLINDORE
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА
SOUTH EAST EUROPEAN UNIVERSITY

FACULTY OF CONTEMPORARY SCIENCES AND TECHNOLOGIES

POSTGRADUATE STUDIES – MASTER STUDIES

STUDY PROGRAM: BUSINESS INFORMATICS

THESIS TITLE
BIG DATA ANALYSIS AND ITS IMPACT ON BUSSINES DECISIONS

STUDENT: **SUPERVISOR:**

Enis Limani ID: 121477 **Assoc. Prof. Dr. Xhemal Zenuni**

.

Abstract

The advancements of technology has proliferated large amounts of data which are easily at disposal to individuals and organizations. If properly analyzed, these large amounts of data can be an added value and can affect the performance of decision makers in business processes. This thesis analyzes the current status how companies use Big Data analytics in business processes from two perspectives. First, it analyses the challenges the companies face in utilization of Big Data analytics to benefit in their business processes. Second, it analyses to what extend is used Big Data analytics to enable revenue growth and to the quality of business processes in companies, thus impacting the decision making and the growth of the company.

Declaration of originality

I hereby confirm that this thesis is my own original work and that I have not sought or used inadmissible help of third parties to produce this work, and that I have clearly referenced all sources used in the work. I have fully referenced and used inverted commas for all text directly or indirectly quoted from a source. This work has not yet been submitted to another examination institution.

Content

List of Figures 5
1. INTRODUCTION 6
1.1. Research Field 9
1.2. Aims of the Research 10
1.3. Research Hypothesis 11
1.4. Research Methodology 12
1.5 Importance of the Thesis 12
2. THEORETICAL BACKGROUND 14
2.1 Big Data 14
2.1.1 Defining Big Data 14
2.1.2 Big Data Today 15
2.1.3 Structure of Big Data 15
2.1.4 Types of Big Data 16
2.1.5 Properties of Big Data 17
2.1.6 Use of Big Data 19
2.2 Big Data Drivers 21
2.1.1 Availability of Data 21
2.2.2 Availability to Aanalyze 21
2.2.3 Availability to Extract Value 22
2.3 Big Data Aanalytics 22
2.3.1 Technology Trends 23
2.3.2 Bussines Intelligence Tools 24
2.3.3 Data Visualization 26
3. PRACTICAL PART 28
3.1 Big Data Trends 28
3.1.1 Storage Capabilities 29
3.1.2 Management of Big Data 31
3.1.3 Analysis of Big Data in Business Settings 32
3.2 Big Data Analysis in Small and Middle-sized Businesses 33
3.2.1 Opportunities 35

3.2.2 Risks 37
3.2.3 Management Challenges 40
3.2.4 Overall Performance 41
3.3 Big Data Analysis in Large Enterprises 42
3.3.1 Technological Requirements 43
3.3.2 Management in Large Enterprises 44
3.3.3 Challenges and Benefits of Big Data 44
4. CONCLUSIONS 46
BIBLIOGRAPHY 47

List of Figures

Figure 1. How Fortune 1000 Executives Report Using Big Data......................................20

Figure 2. Interest in BI Over Time, According to Google Search 2004 – today...............25

Figure 3. Interest in Big Data Over Time, According to Google Search 2004 – today.....26

Figure 4. Capital Spending of Google's Parent Company from 2013-2016......................30

Figure 5 TDWI Survey - Is Big Data Management a Problem or an Opportunity? 31

Figure 6. Big Data Investments by Industry 33

Figure 7. Total Revenue and Growth of Harry & David Inc. 36

Figure 8. EY Global Forensic Data Analytics Survey 38

Figure 9. EY Global Forensic Data Analytics Survey, Top Drivers of FDA Investment .39

Figure 10. 2015 Big Data Survey: Current Implementation Challenges 41

Figure 11. Comparison Between Analytic Capabilities and Performance 45

1. INTRODUCTION

The field of Big Data has experienced great advancement in the last couple of years. With the term becoming more prominent in business settings, as well as the academic community, the research based around its definitions and principles has flourished. The primary purpose of this thesis is to investigate the current state of the art pertaining to Big Data analysis and its impact on bussines decisions. The definition and application in respect to businesses is investigated through the presentation of current available research. To characterize, this thesis firstly include the defining aspects of Big Data analysis and its accompanying technologies. Next, a comparative analysis of research is conducted in order to identify lacking areas and put this master thesis into context. Additionally, the state of the art review available resources and identify the challenges and opportunities Big Data analysis offers in business settings.

The development of new technologies and the proliferation of social networks has enabled gathering of vast amounts of data which can be analysed and structured, thus creating real-time insights in consumer behaviours and potential risks (Mohanty, Jagadeesh and Srivatsa, 2013). This may lead to the conclusion that the only defining characteristic of Big Data is volume. However, this proved to be insufficient to fully define and understand the term. Today, Big Data encompasses four different characteristic which help add new dimensions and clarity in accordance to the advancement in technological fields (Gandomi and Haider, 2015). These characteristic include, the previously mentioned volume, velocity, variety and veracity. Until recently, only the first three, volume, velocity and variety were included in the characterization of Big Data, but due to the emerging new channels which produce data, a fourth characteristic, veracity, was added to the definition (Mediratta, 2015). However, research on the field still often uses the term three V's as the basis for defining Big Data and its properties, explaining only the first three characteristic. Still, the veracity as its fourth characteristic is becoming more and more prominent, thus coining the term four V's.

The characteristic of volume pertains to the most obvious aspect of Big Data, which is the vast amount of generated data, as opposed to other, more traditional sources. Variety characterizes the number of sources and types of data while velocity encompasses the fast-paced gathering and processing of data (Gandomi and Haider, 2015). The term veracity, one of the latest addition to the definition of Big Data, pertains to the quality of gathered data and credibility of its sources (Ernst & Young Global Limited, 2014). This inclusion to the term can be traced back to IBM, the American multinational technology company and their explanation that due to the number of sources which generate data, one should be wary of its accuracy as low data quality ends up creating losses for businesses, which choose to base their strategies around them (Ibmbigdatahub.com, 2017).

Another aspect that is also starting to emerge, but is not yet that prominent in business and academic circles is variability. Along with this term, the aspect of complexity is also added. These aspects originated from SAS Institute, an American multinational developer of analytics software founded in 1976. The institute characterizes the inconsistency of velocity and the complexity of the relationship between sources and acquired data, which need to be addressed to enable better analysing and utilisation of Big Data (SAS Insights, 2017). From the available research, it is evident how complex Big Data is, starting from the definition itself. While the three V's definition has become the status quo in academic circles, the term itself is still being explored as the nature of Big Data evolves with the advancement in technology and development of social networks. Thus, more data-generating sources are established, creating the need for constant, fast and reliable research on the field. However, defining the only the term and its characteristics is not in itself valuable without proper examination of utilization possibilities of Big Data. To understand how Big Data can be applied, the available research seeks to clarify its properties and what Big Data consists of. These examinations can then be applied to examine how Big Data can be managed to further growth and development (Tanner, 2014). Data is generally divided into structured and unstructured data (Marr, 2015). However, only a small percentage of Big Data can be attributed to structured data with 95% being unstructured data (Gandomi and Haider, 2015).

Due to the heterogeneity of Big Data, the classification needed an expansion, thus, semi-structured data is now more and more found in academic papers and books as a category that pertains to the middle ground between structured and unstructured data (Marr, 2015). Structured data is data that has an already assigned field, enabling computerized analysis, while unstructured data consists of information sources which are not categorized and cannot be analysed with traditional tools (Baars and Kemper, 2008). These can include data from social networks, videos, audio files etc.

While unstructured data can be very useful, due to its diverse facets, it can be challenging to analyse and utilise in business strategies, often requiring creative thinking and dedication to investigate. It often pertains to text documents or web pages with XML and JSON, or JavaScript Object Notation files (Mohanty, Jagadeesh and Srivatsa, 2013). Due to their nature, they do contain markers which make them readable to machines, but their lack of categorization prevents them to be fully structured. Precisely the complex nature and heterogeneity of Big Data is a challenge in conducting research on the field. This leads to analysing obstacles of the acquired data. While successful implementation of Big Data analytics does generate business success as it improves business strategies providing valuable insights in consumer needs, extracting valuable data can be a time-consuming process which requires high expertise on the field (Kalyvas and Overly, 2013). Many researches thus emphasize that the key to successful implementation of Big Data does not lie in the data itself but the way it is utilized and structured into meaningful information (Stubbs, 2014). To further characterize, while it is important to successfully analyse the acquired data using precise algorithms and data scientists, the more important aspect is to identify the problem or the question which should be answered by the analysis (Mohanty, Jagadeesh and Srivatsa, 2013). Pertaining to the strategical aspects of Big Data usage, a term that has started to enter business circles and the academic community is Smart Data. While the name Big Data implies that the defining characteristic of such data is only its volume, the term Smart Data emphasizes the strategic benefits Big Data can bring to businesses (Marr, 2015). However, these can be accomplished if the right frameworks, architecture and human capabilities are utilized (Ernst & Young Global Limited, 2014). Additionally, large amounts of data require large frameworks which can accommodate them, as well as proper analysing techniques. The examined research thus found that data analytics and data management have proven to be essential parts of Big Data implementation tactics. While Big Data management pertains to the required processes and technologies that enable its storage and prepare it for analysis, Big Data analytics focus on the techniques which are used to extract information from it (Gandomi and Haider, 2015) The above mentioned is why Big Data is often seen as a costly and time-consuming investment that requires a multidimensional approach across numerous channels. However, more and more research papers are emerging which focus on Big Data implementation on a smaller scale, thus examining principles and strategies smaller companies can use to benefit from Big Data. As such, many services that can be outsourced to enable easier access to Big Data and insights gained from it have emerged. Additionally, data markets are also on the rise, enabling sharing of data on a subscription basis (Mohanty, Jagadeesh and Srivatsa, 2013). All of this creates a stimulative environment for Big Data research, not only on a large scale examining principles that business giants such as Google or Facebook use, but also on a smaller scale, thus providing a wider picture on how Big Data affects all businesses. In addition, as Big Data is often used not only in business settings, but educational, legal and healthcare, the research field can be addressed from multiple standpoints (Marr, 2015).

With it however, some risks are associated. As Big Data is collected from a number of electronic devices and can be more easily accessed than ever, legal issues pertaining to privacy as well as security issues are starting to arise (Ernst & Young Global Limited, 2014). These issues are starting to gain traction and are being more frequently addressed in academic journals and business papers. It should be noted that most available research comes from established companies and individuals with extensive experience in the field of Big Data. The particular interest of this master thesis is however to pertain to business decisions and examine Big Data impact on them. To gain a detailed insight, analysis of both larger enterprises and medium-sized companies is needed. Thus, this thesis provides a detailed analysis while keeping the presented research in mind.

1.1. Research Field

Gathering, analysing and implementing information has always been the driving force of development in almost every field, be it business or personal. Since the dawn of humanity, people used personal, ancestor and peer experience to create better results in their daily lives, improve their chances of success and avoid danger. Thus, information gathering and implementing can be perceived as the focal point of human advancement. Every information starts with small bits of unstructured information or data. Only if the data is analysed and structured can it be called information. The rise of modern technologies has made data acquiring easier than ever, with it being collected through almost every electronic device, from computers, phones or cash registries. Thus, large amounts of data are being generated constantly on personal and business levels enabling an infinite number of analysing possibilities. As such, the term for large data amounts called Big Data has become a more frequently discussed topic among academics as well as businessmen, from CEOs of large enterprises to independent entrepreneurs. The notion of Big Data has started to emerge as one of the focal points in the development of successful business strategies as well as an important growth enabler.

However, from the definition to the possibilities and structuralization, Big Data still lacks overall understanding which would enable better utilization of its possibilities. This is in part due to the complexity and multi-dimensional aspects this research field encompasses. From daily routines, social networks, education, governments or healthcare, Big Data can be found everywhere with staggering amounts of data being generated every second. As such, even though information can be accessed easier than ever, the amount of available data can make it hard to identify which one should be structured and analysed to enable progress for businesses. However, the obstacle does not only lie in the identification of the right data, but also in the right tools and solutions which help create useful information from the provided data. Since the field of Big Data is complex and has multiple aspects ranging from personal to business and legal, this paper will mostly focus on the analyzing aspect and its impact on business decisions of large enterprises and smaller to middle companies.

1.2. Aims of the Research

The chosen topic for this master thesis is Big Data Analysis and its impact on business decisions. As previously mentioned, due to the complexity of the research field, different approaches for the research of Big Data are possible. From government agencies to healthcare and educational institutions, security issues and benefits for businesses, there are many aspects which can be examined to research the impact and development of Big Data and its accompanying technologies. However, this paper focuses primarily on the impact of Big Data analysis on business decisions, examined through a variety of enterprises, from smaller to larger ones. The main goal of the research is to examine the relationship between business decisions and the usage of Big Data analytics. It delves into the issue how the constantly generating data has impacted businesses across the board and analyse its influence on business giants such as Facebook and Google, but also smaller to middle companies. Furthermore, the study examinez the potential risks of Big Data as a decision making tool, pertaining to the volume, velocity and variety of constantly generated data. The relationship between customers and companies in the context of Big Data is also addressed. Due to the constant circulation of communication between consumers and companies, data collected from those consumers in various ways also represents an important aspect of Big Data. Thus, these correlations are also concluded in the research.

The research also delves into the definition of Big Data from the academic aspect, as well as try to clarify some concerns which may arise in an attempt to define the term. Additionally, while some legal and security issues are addressed in the research, however, this thesis keeps its focus on the business decision aspect, aiming to define and determine the relationship between Big Data and business decisions.

1.3. Research Hypothesis

Based on previous conducted research in the respected field, the main hypothesis pertaining to business decisions and Big Data is that:

H1: Effective management of Big Data can add value to the company (quality-process related).

H2: Efficient management of Big Data leads to positive company results (quantitative - cost related).

H3: Due to the costly and time-consuming nature of Big Data, it should only be used within larger companies.

The hypotheses offer a three-way approach with effective and efficient management of Big Data as the foundation of the examination. On one side, the relations between the quality and processes will be examined to conclude whether Big Data can add value to the company. On the other side, the quantitative aspect will also be analysed to determine how management of Big Data impacts the financial results of the company. This will be examined from available online resources and financial reports. Additionally, the cost-to-benefit ratio of Big Data analytics implementation within larger and smaller enterprises will also be researched. The thesis attempts to analyse how complex is it to use Big Data and available business intelligence tools to enable revenue growth and company development. With these three hypotheses as the foundation, this paper will provide insights on how Big Data is acquired within companies and, more importantly, to what extent and how it is utilized to benefit the long-term strategies of a company. Furthermore, the hypotheses will be applied to enterprises of different sizes for the purpose of gaining a wider picture on how they apply in diverse environments. Lastly, there are possibilities of further examination of the selected topic in a different setting pertaining to other aspects of Big Data.

1.4. Research Methodology

The hypotheses will be investigated through quantitative and qualitative research methods. The naturalistic approach of qualitative methods will help provide insights in dynamics created by introduction of Big Data analysis in business environments, as well as investigate the relationship between consumers and companies pertaining to its utilization. Quantitative methods will be applied on a larger scale, through the use of financial reports and case studies on enterprises of smaller to middle size, as well as large multinational companies. This approach will enable insights on how the management of Big Data affects the companies' development and revenue growth. Apart from these research methods, the definition of Big Data itself and its different components will be examined and defined. Thus, the reader will be provided with background knowledge on what Big Data is, how it is collected and lastly, how it can be applied in different settings. However, while aspects such as education and government will be touched upon to provide an overall overview of the topic and help understand it, they will not be investigated in detail. Instead, the research methodologies will focus on business settings. Additionally, the cost of Big Data acquiring and analysing on different scales will be investigated. This will be done to help establish a link between cost and benefit of Big Data. Through these methods the paper will provide both understanding of Big Data, as well as take its multiple variables into account and in turn approve or reject the selected hypotheses. Both quantitative and qualitative research methods were chosen to fully determine how Big Data analysis impacts business decisions and consumer behaviour. In addition, statistical analysis and subsequent interpretation play a key role in the utilization of Big Data. Thus, both methods are necessary to encompass and research the topic of the thesis.

1.5 Importance of the Thesis

As Big Data is becoming one of the most discussed topics in the business environment, giving entrepreneurs and managers both new chances and risks, the need for academic analysis becomes more prominent. With the development of modern technologies and the globalisation of markets, insights in consumer behaviours, business principles and correlations become essential for business to stay competitive. However, due to the volume of data being generated through numerous sources, without proper analysis and utilization, the data becomes more of a liability instead of an asset. The circulation of information is more fast-paced than ever, with business looking to implement the acquired information in the most efficient way. To enable this growth, deep understanding of the topic and its functioning is needed. Thus, it is essential for the academic community to examine its principles and determine what aspects are best to be incorporated and how. However, even though the conversation on the topic has taken its place in the business community, with Big Data being perceived as the new growth enabler in almost all fields, the academic community has yet to delve into the details of the topic. Due to the complexity of the topic, there are many aspects that the academic community can chose to investigate to create an overall interpretation of Big Data. This paper will investigate the impact on business decisions, particularly if and how it enables growth not only on a large, but also on a smaller scale. This will provide greater insights on how Big Data works and contribute to the academic discourse on the topic, thus enabling managers and entrepreneurs to better utilize it in their business.

2. THEORETICAL BACKGROUND

2.1 Big Data

Collecting, analyzing and utilizing data has always been the driving force behind most human development. The transformation of data into meaningful insights creates information, which can be applied to further growth and development (Ackoff, 1999). The development of new technologies has not only globalized existing markets and created new ones, but also increased the rate at which new data is generated. To be specific, humans produced the vast majority of data available today in the last couple of years (Axline, 2017). Since more information is being produced, the need for space and storage capabilities keeps growing and with it, the challenge to manage and extract value from the vast amounts of data (Ohlhorst, 2013). However, to be able to utilize and implement Big Data, it is essential to first understand what constitutes Big Data, where it originated and what its main properties are. Since Big Data is a relatively new concept, its definition, as well as it implementation is stillundergoing chabges. This is also due to the fact that Big Data is closely connected to technological advancement. As technology advances, the research revolving around Big Data and its impact on today's society follows these advancements.

2.1.1 Defining Big Data

The term Big Data has undergone many changes in its definition, ranging from business-oriented books and papers, to conferences and guides created by leading companies in the field, such as IBM. Still, in academic circles, there is a certain lack of literature pertaining to Big Data concepts (Gandomi and Haider, 2015). This can be attributed to the fast-paced development of new technologies and the proliferation of social media leading to a rapid surge in amount of data that is generated, leading to the lack of time for proper academic discourse which would shed light on the new technological advancements and enable businesses to incorporate them in their strategies. In most cases, Big Data is described as large volumes of data that cannot be processed with traditional tools, thus making it difficult to extract insights and value from it (Ohlhorst, 2013). As data sets are becoming not only larger, but also more diverse due to the development of networks such as Facebook and advancement in search engines, other characteristic of Big Data have started to emerge (Chen et al., 2014). These characteristics are volume, variety and velocity, coining the three V's definition of Big Data (Laney, 2001). However, as new research is conducted on the field, new characterizations such as veracity, coined by IBM, and velocity, coined by SAS are also being added as defining points of the term (Ibmbigdatahub.com, 2017; SAS Insights, 2017). Nonetheless, the most common definition still relies on the three V's, with other slowly starting to enter the academic discourse. According to the Gartner IT Glossary, an american research and advisory company, the term Big Data is defined as: "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." (Gartner IT Glossary, 2017)

In other word, while volume is one of its defining characteristic, it is not the only one. Speed at which data is generated and its type also play an essential role in the definition of the term.

2.1.2 Big Data Today

The proliferation of social media, development of mobile technologies and other technological advancements have led to a massive popularity and omnipresence of Big Data. Data is collected from almost any electronic device, generating insights from every part of today's society (Ohlhorst, 2013). However, only its volume is no longer the sole purpose of research pertaining to Big Data, as its heterogeneity has initiated research focusing on how to extract value and insights, rather than how to store it (Runciman and Gordon, 2014). As new types of data are being created not only in vast amounts but also real-time, challenges in its analysis arise (Troester, 2012). This is why research today started to shift its focus towards Big Data Analytics, principles, technologies and strategies that are developed to enable easier access to information Big Data can produce (Assunção et al., 2015). The proliferation of Big Data has thus, not only changed the way many industries operate, but also created a new marketplace for companies specializing in Big Data analysis and implementation, helping other companies extract value and include Big Data insights as part of their business strategies. Today, data has become a science, creating new employment opportunities as well as a new career path. Data scientist are specialists in extracting value from large data sets, data visualisation and analytics in general. Their main occupation is to find and structure informations that can be acquired from Big Data (Ernst & Young Global Limited, 2014). As such, Big Data is becoming a concept unavoidable in most business fields and markets.

2.1.3 Structure of Big Data

For Big Data to provide any value, it is first important to understand it complex structure. As large amounts of data are being generated constantly, the distinction between different data types must be made. The issue with its structure lies mainly in the heterogenetic nature of data (Ram, Zhang and Koronios, 2016). It can be gathered from typical sources such as customer login on websites, credit card statements, educational or government statistics, which are public data or enterprise data such as financial reports, number of employees etc. Still, all those are relatively traditional data which can be structured within easily accessible frameworks. Big Data can also be acquired from social networks such as Facebook, Instagram, YouTube and others. Data gathered from such sources is often more challenging, not only to structuralize, but also to analyze (Giannetto, 2015). Due to the nature of social media, every post, photo or action on social networks becomes part of Big Data which can be used to extract valuable customer insights and create business strategies (Kim, 2015). However, since data on social networks is created rapidly, unstructured and real-time, its analysis can be a costly and time-consuming effort. As such, companies often struggle to find data which provide benefits for their own advancement (Murthy and Bowman, 2014). This is due to the fact that there is no one universal structure of Big Data that would fit every businesses' needs and wishes, but rather an endless string of possible variations what types of data can be managed and used in the most effective way. However, the need for structure of data still persist. Its development can be largely attributed to the ever-growing global connectivity through the Internet and the development of new technologies, which enable easier structuralisation of Big Data (Bryant, Katz and Lazowska, 2008). This is where a new industry was created to help business to store, structure and analyze Big Data, reducing the cost of such efforts and opening the door of Big Data utilization to a wide variety of industries and businesses (Mohanty, Jagadeesh and Srivatsa, 2013). The following chapters will further examine the types of Big Data, how they are structured and what its main properties are.

2.1.4 Types of Big Data

As already noted, Big Data is very heterogenetic in its nature, meaning that it consist of various types of data. Each type has its own specifics, analyzing opportunities as well as challenges. Typically, there are three main categories of Big Data. These are structured, unstructured and semi-structured data (Marr, 2015). Previously, data could be divided into structured and unstructured data, but the development of web technologies and programming capabilities has created the semi-structured category . It should however be noted that the vast majority of data belongs in the unstructured group. To be specific, it is estimated that over 90% of Big Data falls within the unstructured category, with only 5% of overall data being completely structured (Gandomi and Haider, 2015). Structured data is data that is stored and processed in a fixed format which can include spreadsheets, databases etc.(Gandomi and Haider, 2015). Another characteristic of structured data is that it can be easily analyzed and extracted by simple search algorithms (Tanner, 2014). On the other side of the spectrum is unstructured data. Unstructured data refers to data which does not have a predetermined organizing model or database and may contain a number of information within just one part (Assunção et al., 2015). Typically, any text, video or audio file can be unstructured data. The proliferation of social media networks has particularly foster the creation of unstructured Big Data, as the vast majority of data produced on social media contains images, text, videos etc, thus posing a challenge in the extraction of valuable information (Gandomi and Haider, 2015). Additionally, the development of smartphones and applications has enabled users to access social networks at any given time, resulting in more volume and a wider variety of Big Data. The third category, semi-structured data combines both structured and unstructured data characteristic. Semi-structured data includes web pages with XML or JSON, thus containing markers which make them machine-readable; however, since they lack categorization, they cannot be fully structured in the traditional sense. This particular aspect leads to their semi-structured characteristic (Mohanty, Jagadeesh and Srivatsa, 2013). As the largest category, unstructured data does come with its own sets of challenges, mostly pertaining its lack of structuralization. However, since social media generates a large portion of unstructured data and provides valuable insights in consumer behavior, needs and wishes, it is one of the topics of Big Data analysis (Mohanty, Jagadeesh and Srivatsa, 2013). To characterize, companies and researchers are constantly developing new ways and tools which can extract value from social media. Big Data from social media comes in many variations, from Facebook posts, tweets, Instagram photos, videos, etc. Over 70% of all Internet users are also on different social networks, which only contributes to the large amounts of unstructured data being generated from it. Additionally, marketers have started to use the potential of social media for advertising in various forms, thus acquiring more precise data on their target audience, which can then be utilized to develop more successful business strategies (Rathinasamy, 2015). While social media and Big Data have a correlation that is large in volume and variety, unstructured data can go beyond the scope of social media. Unstructured data can also be, for instance, customer service calls, which come in audio formats or transcripts and generally include consumers' thoughts on a certain product or service. However, extracting valuable information from such sources can be a daunting and costly process (Bryant, Katz and Lazowska, 2008). It requires large storage capabilities and specialised analysing tools, but also creativity. While unstructured data does hold its potential, often creative and strategic thinking is needed to connect and structuralize raw data, determine which data is necessary to gain insights that can be implemented within business strategies (Troester, 2012).

While the terms structured, unstructured and semi-structured have become the most common when describing Big Data types, it can also be divided into internal and external data (Mohanty, Jagadeesh and Srivatsa, 2013). Internal data does have some similarities with structured data as it operates within given frameworks. It can include everything from transactions, inventory to sales within a company (CHD Expert, 2017). External data, on the other hand, is data generated from an outside source, which can come from various databases, social media etc., It can be implemented for company specific research, specifically if companies do not have the capabilities to generate data internally in a cost-friendly manner, but it should be noted that not all external solutions will cater to the company's needs (Burgess, 2014).

2.1.5 Properties of Big Data

As the name of the term suggests, one of Big Data's main properties is its volume, however it is not the only one, as recent research has started to add other characteristics to the term to further define it. One of the most common explanation for Big Data properties comes as the term three V's, which stand for the main characteristics that are used to describe Big Data - volume, variety and velocity (Gandomi and Haider, 2015). Volume pertains to the amount of data generated from different sources. Data is often considered big only if it is larger than one terabyte, with the threshold for what amount of data is considered Big Data changing as new technologies and storage capabilities are developed (Schroeck et al., 2012). Big Data can be measured in Terabytes, Petabytes, Exabytes, Zettabytes and Yottabytes, with Terabytes being the smallest unit which is made up of 1000 Gigabytes. The largest unit is Yottabytes, which contains 1e12 Terabytes (Mohanty, Jagadeesh and Srivatsa, 2013). As the data grows, the challenge of storage and management of the collected data, due to its increased volume can lead to a costly storage system and management frameworks.

The second “V” in the three V’s is variety. According to Gandomi and Haider: “Variety refers to the structural heterogeneity in a dataset.” (Gandomi and Haider, 2015, p. 138) To characterize, Big Data does not only come in large volumes but from a number of sources and number of types. As explained in the previous chapter, variety of the data pertains to the structured, unstructured and semi-structured type. Additionally, the variety also pertains to the source from which is gathered, not only its type. This is why external and internal data sources also play a role in the variety of Big Data. The third characteristic is velocity. Velocity is defined as: “the rate at which data are generated and the speed at which it should be analyzed and acted upon (Gandomi and Haider, 2015, p. 138). This characteristic has proliferated itself with the development of digital devices and social networks, meaning that data can now be created from a number of sources at any given time, as new technologies and mobile devices have enabled content creation, sharing, commenting and reacting that is not place-bound. Rather, data is streamed real-time from any location users choose to, which only adds to the complexity and heterogeneity of Big Data (Marr, 2015). As research on the field of Big Data is progressing, new characteristics are added to the term to help define it more accurately. One such characteristic is veracity, which pertains to the quality of Big Data and the credibility of its sources (Ernst & Young Global Limited, 2014). This is due to the fact that as technologies develop, so do the sources which generate data. However, not all sources are created equally, with some creating more accurate data, while other should be examined critically to prevent missteps in their analysis which could lead to business losses (Murthy and Bowman, 2014). Such measures are particularly important if the data comes from social media. Due to their nature, social networks are highly diverse and generate data from any message or posts from its users. However, the credibility of these posts can vary, as there are virtually no frameworks within social media frameworks to confirm their credibility (Puget, 2015). The term was first introduced by IBM, coining the term the four V’s (Ibmbigdatahub.com, 2017). Additionally, characteristics such as variability and complexity were also recently become more common when defining the term and its properties (Gandomi and Haider, 2015). According to Gandomi and Haider: “variability refers to the variation in the data flow rates.” (Gandomi and Haider, 2015, p.139). Complexity, on the other hand, deals with the variety of sources data is generated from, and in that sense somewhat correlates with the dimension of veracity. Both terms were introduced by SAS Institute, in an attempt to emphasize the challenges in creating connections and relationships between generated data (SAS Insights, 2017). With this in mind, it is important to highlight that there is a strong correlation between all properties of Big Data. Mainly, if the volume of data changes, there is a strong probability that the variety and velocity also follow those changes (Mohanty, Jagadeesh and Srivatsa, 2013). Thus, each dimension cannot be viewed independently, but rather as a part of the whole Big Data concept. However, businesses, educational institutions, governments or healthcare will use different aspect of Big Data depending on their needs and goals, making the management of Big Data and the identification of the right data sets an essential part of Big Data utilization within given frameworks or strategies.

2.1.6 Use of Big Data

The data can be gathered from a number of sources and comes in many forms. As such, there are a number of ways Big Data can be implemented, not only in business environments but in education, governments and healthcare as well. The main question does not lie in whether Big Data can be used, but rather, how it can be most proficiently used (Ernst & Young Global Limited, 2014). One of the trends for Big Data usage is in business process optimisation. Large retailers are using Big Data to optimize their offers, stock or delivery routes by analysing data gathered from various sources (Virmani, 2017). By analysing the data, they can identify where improvement is needed and considerably cut costs or provide better value to their consumers. Another way Big Data can improve performance is by providing insights in customers' wishes, needs and behaviours (Bryant, Katz and Lazowska, 2008). By implementing Big Data analytics, it is possible to predict consumer behaviours and create strategies according to those predictions. While data for this purpose does come from many different sources and can enable finding correlations between, for instance, consumer behaviours and external factors, it also comes from social networks (Kim, 2015). As mentioned, data from social networks is often unstructured and comes at high velocity and high variability. While it can provide valuable insights, it can also pose a challenge to store and analyse (Murthy and Bowman, 2014).

Big Data today is being also used in sports, science, law enforcement and traffic among others (Schroeck et al., 2012). For instance, professional sports clubs can use Big Data analytics to improve their performance by tracking individual players' scores and comparing them. They can also be used to analyse games and develop better strategies according to the opposing team's strengths and weaknesses (Marr, 2017). Traffic is another field that can benefit from Big Data. By analysing traffic events in large cities and creating a correlation between other external factors, traffic jams can be prevented. Additionally, it can be used to improve existing routes and identify parts of the transport system which need improvement (Marr, 2017). However, Big Data is not only for large enterprises, governments or scientific establishments. On a personal level, Big Data can also be used, creating a relationship between individual users and companies providing a certain service. For instance, smart watches and bracelets can help maintain health by measuring heart rate, blood pressure etc. This provides the user with valuable real-time insights, while generating data that the company can use for research or future development (Andreescu, 2013). Use of Big Data can be found in almost any field.

The main appeal of Big Data for businesses lies the ability to make strategic decisions based on insights provided from large datasets (McAfee and Brynjolfsson, 2012). Purchases can be tracked to analyse what customers want, or delivery routes can be optimized to provide a more efficient service. However, how businesses implement Big Data within their frameworks depends heavily on the industry itself and the company's needs at a given time (Mohanty, Jagadeesh and Srivatsa, 2013). For instance, retailers will use gathered data to predict customer behaviours, offer products that are more likely to appeal to their target audience or improve the brand image. Financial services, banks etc., are able to predict economic trends with more accuracy and adapt to them, while marketers can use insights gathered from social media to predict which content users are more likely to interact with and craft their marketing strategies according to the results. However, to achieve positive results from Big Data, it is necessary to use adequate storage capabilities, analysing tools and manage data in a cost efficient way (Bean, 2017). Since the amount of gathered data for a particular field can be enormous, is storage and management can be challenging, particularly for smaller companies. Figure 1. presents a survey by NewVantage Partners, conducted on Fortune 1000 companies, the 1000 most successful companies chosen by Fortune magazine and ranked by revenue. The survey is based around the usage of Big Data in business environments, what particular part businesses chose to improve with Big Data and how it pertains to the acquired value for the business. The survey shows mostly positive results, with the greatest success being in decreasing expenses. This is mostly done by identifying challenging areas and adopting principles to reduce production, delivery costs etc. (New Vantage Partners, 2017).

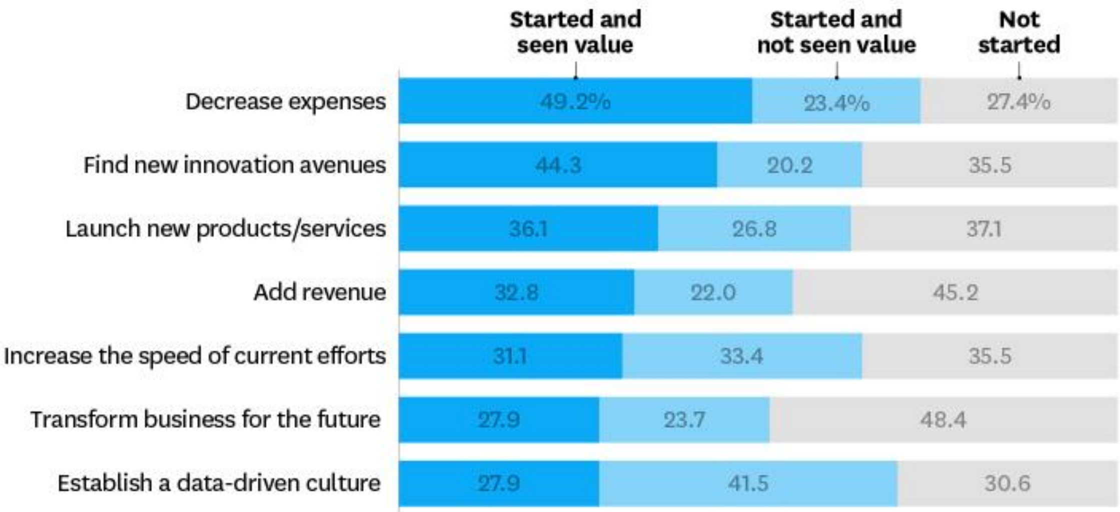

Figure 1. How Fortune 1000 Executives Report Using Big Data

However, when it comes to business setting, there are some risks to be aware of. These mostly pertain to organizational initiatives of data within companies and a clear management strategy (Platt et al., 2014). As large volumes of data are created constantly, companies are faced with the challenge, not if they should utilize Big Data, but rather how (Marr, 2015). Thus, one of the roles of management is to identify which problems can be solved with Big Data. The principles on how Big Data is managed and how its management impacts business decisions will be analyzed in the practical part of this paper.

2.2 Big Data Drivers

The first step in Big Data implementation within business strategies is to understand which drivers enable Big Data usage. Since the Big Data market is constantly expanding and can be costly to participate in, companies first strive to identify on which fields and how Big Data can be implemented (Connolly, 2012). This is why, according to Bernard Marr, the aim should not be to gather vast amounts of data with no goal in sight, but rather gather smart that which caters to goals the business, brand or company have set (Marr, 2015). To do so, three Big Data drivers are necessary. Those are availability of data, availability to analyse and availability to extract value. To further characterize, availability of data refers to sources from which data is collected, how it is stored and how it can be most efficiently accessed (Ernst & Young Global Limited, 2014). This particularly pertains to the heterogeneity of data, as its nature can pose many challenges. Availability to analyse refers to analytical tools that are necessary to process vast amounts of data, while availability to extract value deals with how valuable insights can be gained through its analysis (Connolly, 2012). Without these three drives, Big Data becomes just amounts of data that serve no real purpose.

2.1.1 Availability of Data

It is important to note that availability of data does not pertain to its performance, rather it deals with whether enough data is available to analyze it without creating biases or false conclusions (Lyon, 2015). The main challenge lies in the multitude of sources from which Big Data can be generated and identifying which data sets are relevant to the business or a company (SAS Insights, 2017). Additionally, data should be delivered in a timely manner, which can also pose a challenge, as its complexity can lead to slower storage within analysing platforms. In conjunction with this challenge, once data is processed, it can be stored in a number of ways, which can also lead to problems as it may not be accessed whenever it is needed. DB2 tables, or IBM's relational database management system tables are often used to approach this challenge, however, due to a number of concerns, such as hot spots or old data, the tables may not always perform properly, requiring constant readjustments (Lyon, 2015).

2.2.2 Availability to Aanalyze

Big Data is only valuable if it can be analysed and, additionally, if value can be extracted from it. In the traditional sense, large data amounts in itself are not that challenging, if provided storage is available, however, depending on the goal of analysis and type of data, its availability to analyse can drastically degrade with its growing size (Murthy and Bowman, 2014). Additionally, analysing large heterogeneous data sets can lead to noise accumulation, spurious correlations or incidental endogeneity among others (Gandomi and Haider, 2015). Spurious correlation refers to the perceived causal relationship between variables that does not exist. To further explain, variables are not causally related, but it may seem that they are due to simple coincidence (Simon, 1954). This phenomenon is notable with larger data sets. Incidental endogeneity, also commonly found in Big Data sets, is the correlation between variables and the error term, which in itself brings the statistical methods for regression analysis into question (Fan and Liao, 2014). However, despite the challenges, there are available tools for Big Data analysis which constantly evolve to keep track with new types of data, as well as its growing amount.

2.2.3 Availability to Extract Value

To extract valuable insights from Big Data, it is necessary to firstly analyze what questions or problems should information acquired from Big Data tackle. These can range from providing business value, being a key enabler for return on investment or provide insights in consumer behavior and needs. However, to be able to extract data, simple analysis is not enough - rather performance management, decisions science, social measurement and data exploration are all necessary to obtain value (Parise, Iyer and Vesset, 2012). One of the key components is for the leadership to understand not only what Big Data is, but also how to use it as a strategic asset (Ismail, 2017). Thus, it is essential to incorporate technical knowledge and strategic thinking to be able to implement Big Data within business strategies (Gupta, 2014). However, as it will be discussed further, to extract real value from Big Data can initially a costly investment.

2.3 Big Data Aanalytics

For Big Data to become valuable, it is first necessary to analyze the gathered data, thus creating information that can be incorporated within business strategies and generate value. To do so, proper analytics tools are necessary. IBM defines Big Data analytics as "the use of advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/batch, and different sizes from terabytes to zettabytes. (IBM Analytics, 2017). Big Data analytics typically can go in one of four directions: prescriptive, predictive, diagnostic and descriptive analysis (Declues, 2017). Prescriptive analysis pertains to the best possible solution for a problem. This type of analysis takes multiple factors into account to determine the best course of action. However, it is heavily under-utilized due to the many factors that are needed to provide the best solution (Maydon, 2017). Predictive analysis refers to a likely event, thus forecasting possible outcomes. Through the use of a variety of data, this type of analysis forms a correlation between components that enables future predictions. Diagnostic analysis deals with the root of the problem, or rather, why something is happening (Declues, 2017). Lastly, descriptive analysis is the most common type of analysis found in businesses, mainly due to the straightforward approach that often does not require analysing a number of complex factors. Descriptive analysis deals with what is happening, rather than why or to what future outcomes (Huddleston, 2016)). Some examples of descriptive analysis are for instance revenue reports or demographic information.

To conduct any type of analysis, first data must be implemented within analytical tools. There is a wide variety of techniques to gain information. These include Online Analytical Processing or OLAP; ever processing, data mining etc. (Mohanty, Jagadeesh and Srivatsa, 2013). Traditionally, to process of implementing data within analytical frameworks was referred to as ETL, which is short for Extract, Transform and Load (Rouse, Martinek and Stedman, 2017). However, due to the high volume, variety and velocity of Big Data, new technological solutions are necessary to process and analyse the data. As such, technology trend pertaining to Big Data analysis are constantly evolving, thus enabling easier and a more cost-efficient access to information.

2.3.1 Technology Trends

Since a large percentage of Big Data is unstructured or semi-structured data that does not fit within traditional data initiatives, which often pertain only to structured data, new technological trends have arisen to accommodate the growing heterogeneity of data. The first technological advancement came with the development of Hadoop in 2005, an open-source framework that can store and process large amounts of data (Lo, 2017). Additionally, as NoSQL, a framework of databases which processes data on a large scale, was developed, new solutions for Big Data processing surfaced. Today many businesses turn to tools such as YARN, MapReduce, Spark, HBase and others to collect and analyse data (Rouse, Martinek and Stedman, 2017). One particular advancement that has started to emerge is the usage of Hadoop data lake, a concept which allows for storing and processing raw data directly within Hadoop clusters (Search Technologies, 2017). However, these types of approaches still require organizing and data management.

2.3.1.1 Large Scale Search

One of the first implementations of Big Data was through search queries which initially matched keywords, thus providing a list of results. However, large-scale search has since developed at a fast pace, with search becoming more refined, enabling inclusion or exclusion of certain clusters to provide better results. Additionally, with the development of NLP, or natural language processing technologies, even unstructured Big Data can be classified and categorized (Mohanty, Jagadeesh and Srivatsa, 2013). One such example are search engines, which now are able to identify semantical principles within languages, thus providing more accurate results.

2.3.1.2 Multimedia Content

Unstructured data is a common pitfall in Big Data analysis and a large part of it is multimedia content, which can range from photos, videos or audio files. In their nature, they are mostly user-produced, at high variety and velocity, making them particularly challenging to analyse (Gandomi and Haider, 2015). However, in the marketing industry, this is especially valuable content as it can provide insights in consumer behaviours in a very naturalistic manner. Today, new speech and video analytics are being constantly developed to automatize the extraction of such data. This is possible due to the advancement of technologies such as speech-to-text transcription and CBIR, or Content-Based Image Retrieval (Mohanty, Jagadeesh and Srivatsa, 2013).

2.3.1.3 Sentiment Analysis

Sentiment analysis refers to the summarization of context pertaining to unstructured data. It is often referred to as opinion mining as it determines whether the speaker’s opinion was positive, neutral or negative (Lexalytics, 2017). Most commonly it is used by brands to determine how users perceive the brand in general or a particular product. To conduct a sentiment analysis, technologies such as NLP or natural language processing, computational linguistics and cross-referencing multiple sources are often implemented (Mohanty, Jagadeesh and Srivatsa, 2013) . Sentiment analysis is also often conducted on social media, analysing posts, reviews etc., thus providing insights that can be incorporated into marketing strategies.

2.3.1.4 Data Contextualization

As vast amounts of data are being constantly produced in a number of sources, noise within those datasets is only to be expected. However, it is necessary for businesses to identify which data is relevant to them in a process called data contextualization (Lorentz, 2017). Data contextualization is mostly implemented by cross-referencing data from various sources, such as social media and online chats. With the development of Facebook and Google ID's, contextualizing data has become more advanced and approachable (Mohanty, Jagadeesh and Srivatsa, 2013). Nonetheless, it is a field that requires more research.

2.3.1.5 Exploratory Analytics

Exploratory analytics or exploratory data analysis analyses data for the purpose of discovering a new insight or information (Seltman, 2015). While it does implement a variety of techniques, it presents more an approach, rather than a set of rules that need to be implemented (Information Technology Laboratory, 2017). Exploratory analytics often requires cross-referencing seemingly unrelated data to find a common denominator between variables. To characterize, this approach aims to connect data outside of the traditional OLAP or Online Analytical Processing (Mohanty, Jagadeesh and Srivatsa, 2013).

2.3.1.6 Operational Aanalytics

Operational analytics, as opposed to exploratory analytics, focuses on existing operations and how to improve them (Techopedia.com, 2017). Again, operational analytics use a variety of methods such as data mining and aggregation, but its main aim is to develop more effective decision making tactics. This type of analytics is particularly prominent in the healthcare industry, as data gathered from medical equipment is often analysed with operational analytics (Mohanty, Jagadeesh and Srivatsa, 2013).

2.3.2 Bussines Intelligence Tools

Business intelligence tools, or BI tools go hand in hand with Big Data as they pertain to the transformation of available data into information that can be used to benefit businesses in a number of ways, from internal processes to consumer relations (Ram, Zhang and Koronios, 2016). While Big Data Analytics deal with storage and data analysis, BI tools are used to extract information from that data pool (Pepalis, 2015). To further characterize, businesses use BI tools to not only cut costs and optimize existing processes, but also to identify new opportunities that enable development and growth. Another important aspect of Business Intelligence is its focus on data-driven decisions. Rather than relying on their own intuition and experience, company leaders and management are able to make decisions based on processed data (Joly, 2016). What Business Intelligence is and how it is applied within companies will be explained in the following chapters.

2.3.2.1 Defining Bussines Intelligence

According to Gartner IT Glossary, Business Intelligence is defined as "an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve decisions and performance." (Gartner IT Glossary, 2017). While the term was first mentioned in 1865 by Richard Millers in his Cyclopaedia of Commercial and Business Anecdotes, it was not until the mid-20th century that the term became more prominent in business settings. This was due to new technological advancements that started to emerge in the 20th century, as technology became an essential part of modern business intelligence. The development of modern business intelligence began with IBM's development of the first hard disk and other storage technologies which enabled storage of data that could later be processed by computers (Joly, 2016). The new technology was still in its infancy, but with the establishment of personal computers in almost every setting, personal and business, the interest for the term and its capabilities started to arise.

2.3.2.2 Types of Bussines Intelligence Tools

While Big Data consist of both structured and unstructured data, traditional BI tools still mostly pertain to the analysis of structured data (Baars and Kemper, 2008). There are generally three categories in which Business Intelligence tools operate: guided analysis and reporting, self-service Business Intelligence and advanced analytics (Sherman, 2017). Guided analysis and reporting deals with traditional BI tools and is used for creating reports, spreadsheets and dashboards (Sherman, 2017). Self-service BI usually enables users to define specific metrics which enable revelation of patterns and insights. One particular aspect of self-service BI is that it does not require an IT-specialist, but can rather be accessed by individual businesspeople providing a more general and user-friendly approach but lacking custom-made aspects that a business may need (Rouse, 2017). OLAP or online analytical processing, data visualisation and data discovery are often used in self-service BI. The last category is advanced analytics which are often deployed by data scientist and are used for predictive and prescriptive analysis (Ernst & Young Global Limited, 2014). Since creativity and analytical knowledge are necessary to conduct those types of analysis, advanced analytics operate on a scope between Business Intelligence and Big Data analytics often employing booth to achieve the best results (Marvin, 2016). Nowadays, a large number of businesses opt for tools such as SAS Business Intelligence, Oracle BI or Tableau to conduct data and information analysis.

2.3.2.3 Big Data and Bussines Intelligence

Big Data and Business Intelligence are terms that are often used interchangeably, indicating a misunderstanding on the definition of the terms. The most notable difference is that Big Data, while useful, is not necessary to implement Business Intelligence within companies (Pepalis, 2015). The proliferation of smart devices has led to a surge in data volumes, creating Big Data as it is perceived today. On the other hand, BI and its principles have been a subject of interest before the rise of Big Data. This is also indicated in the search volumes of the terms. The term Business Intelligence, while still highly researched, is slowly experiencing a decline in search volume.

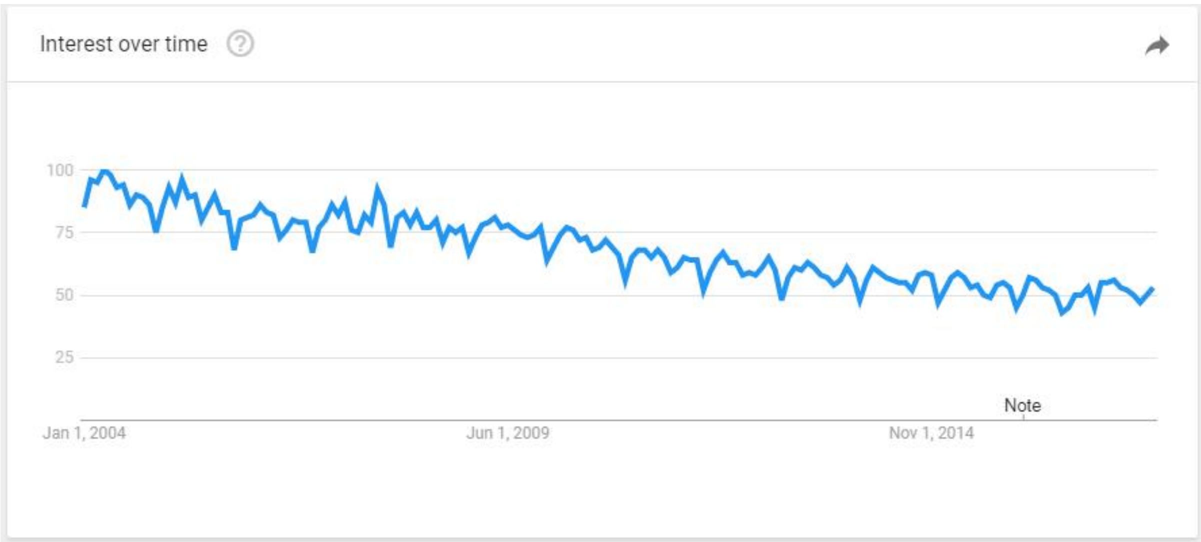

Figure 2. Interest in BI Over Time, According to Google Search 2004 – today.

However, it is interesting to note, that while the term BI is being less researched, the term Big Data is experiencing a surge in the number of searches. This is due to the relative novelty and unfamiliarity with the term and what it represents, but also due to the advancement in technologies which put more focus on Big Data and insights that can be gained from it.

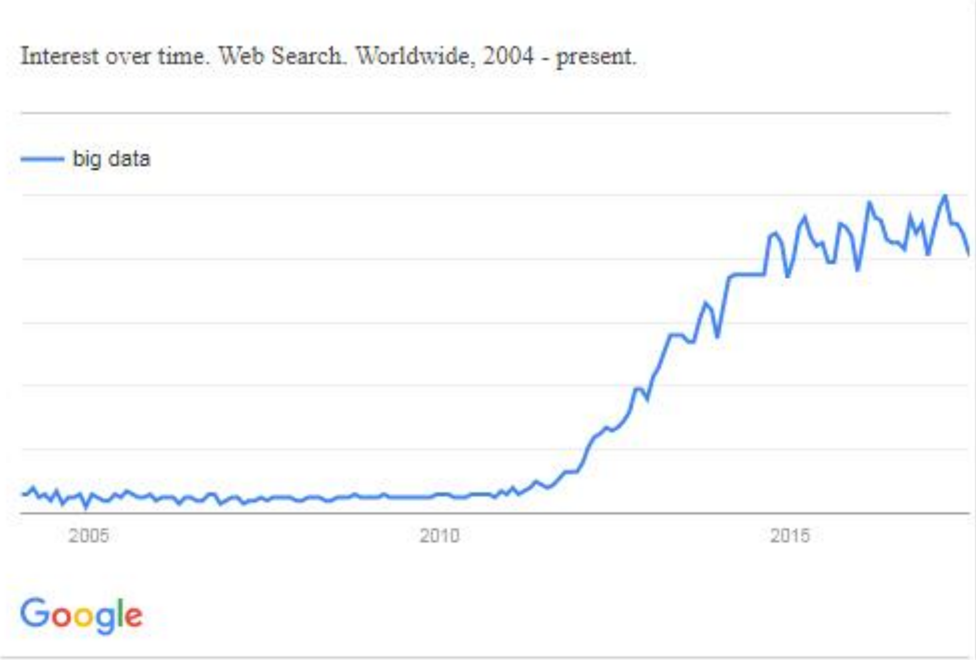

Figure 3. Interest in Big Data Over Time 2004 -present, According to Google Search

Still, Business Intelligence and Big Data can nowadays be seen as two term that operate within same frameworks, but with different purposes. While Big Data focuses on gathering and analysing vast amounts of data, BI mostly pertains to the decision-making process (Ram, Zhang and Koronios, 2016).

2.3.3 Data Visualization

Analyzing data can provide valuable insights, but if it is not presented in an understandable way, it will not garner success. This is where Data Visualization becomes essential for business strategies and development. Data Visualization refers to the presentation of data in an easy to process and graphically interesting way (SAS Institute Inc, 2017). Its main aim is to enable understanding of concepts, patterns and processes with a visual approach. Additionally, successful Data Visualization provides an easy overview of gathered information, which in turn can be implemented within business frameworks in a timely manner. Today, Data Visualization is the norm for any BI tool as well as Big Data Analytics.

2.3.3.1 Data Visualization Technologies

Data Visualization generally falls within two main categories: visual reporting and visual analysis (Eckerson and Hammond, 2011). Charts, graphics, dashboards etc., are the main examples of visual reporting as they present data in a simple overview enabling users to extract a certain metric and use it accordingly (Rouse, 2017). On the other hand, visual analysis consists of more complex data that can help identify patterns, trends or anomalies (Eckerson and Hammond, 2011). Visual analysis are thus more exploratory in their nature, while visual reporting enables data presentation against predefined metrics. Line charts, pie and donut charts as well as scatter plots are all common techniques in data visualization. However, as was the case with other fields, the sheer volume, velocity and variety of Big Data does present a challenge to traditional Data Visualization techniques. While new tools such as SAS Visual Analytics are being constantly developed to support the data, depending on the amount, some compromises may be necessary. Those mostly pertain to the identification of which data is important to highlight and visualize, again leading up to the management to ask which questions should be answered for the company to be able to advance.

2.3.3.2 Data Visualization Principles

For Data Visualization to provide value and insights, some principles are necessary. One of the main principles of successful Data Visualization is clarity in its presentation as well as detailed information on what is being presented (Burns, 2016). Graphical integrity in units and numerical quantities is one of the most important aspects that can lead to distortion of data if not handled properly. According to Cleveland, Data Visualization should always provide clear vision, clear understanding, scales and a general strategy (Cleveland, 1994). While these principles can be further developed with additional guidelines, they provide a general outline of Data Visualization.

3. PRACTICAL PART

3.1 Big Data Trends

With the growing amounts of Big Data and more mentions of the term in both academic and business settings, it is only natural that new trend emerge with the development of new technologies and practical implementation, both successful and less successful examples. The emerging trends of Big Data today deal mostly with the development of more accurate analytics, security and veracity of Big Data, as well as dealing with its functionality and how it pertains to both large and small businesses (Harvey, 2017). Additionally, Big Data was in the past seen as a costly investment that was reserved for large enterprises but with the development of general solutions like subscription-based data providers and storage options, Big Data became available to smaller and middle-sized companies as well. However, emerging Big Data trends do not cater to only one industry, but rather, multiple industries and markets have started to implement Big Data as a part of their business strategies. From traditional brick-and-mortar to online businesses, Big Data is being adapted to improve performance (Marr, 2015). Additionally, its adaption is not limited to businesses only, but fields such as education and healthcare have also started to incorporate Big Data.

However, with the development of new technologies, advancement of Big Data is also to be expected. Today, the focus is slowly shifting from just large data sets to intelligent, or rather, smart data (O'Dowd, 2017). This is why intelligent apps, predictive analysis, machine learning and IoT, or Internet of Things are cited as the new emerging trends in Big Data. Machine learning in particular is starting to gain momentum as artificial intelligence becomes a more prominent topic. Machine learning is a method that enables computers to learn without direct commands or programming (Hall et al., 2014). It is also often referred to as deep learning. To simplify, a learning algorithm is implemented within computers that enables them to differentiate one thing from another by answering questions, thus being able to learn what is correct and what false. One of the most prominent examples of machine learning is found in smart cars. Connected through different applications, smart cars would be able to adapt to their owner, adjust to current traffic conditions and drive itself. The search engine giant, Google is one of the leading companies on that field. Starting their research in 2009, but recently developing new models within its parent company Alphabet Inc. in 2012, Google is using incorporating their driving equipment within multiple types of cars, ranging from Toyota and Lexus to their own custom-built cars (Marr, 2016). But Google is not the only company integrating new Big Data trends within vehicles. Tesla Inc. an American automaker specializing in electric cars, also came out with their own version of a smart car in 2015. Uber, the American transportation and delivery focused company is developing smart cars as well. While there has been great progress on, misread data and software reliability are still one of the main challenges in this field (Zhou et al., 2016).

However, Big Data trends do not happen only on a large scale and within large companies looking to develop new technologies. On a smaller scale, companies are rapidly starting to adapt parts of Big Data analytics such as predictive and prescriptive analysis which in the past could mostly be implemented by larger enterprises due to the substantial investments that they required. This is possible due to the development of new tools that offer more general, open-source solutions on a subscription basis. While not custom-made, they enable Big Data implementation on a much smaller scale and in a cost-efficient way (Murthy, Sawyer and Bowman, 2014). This approach can already be utilized in a number of ways, particularly in the digital marketing segment, as one of the main forces in predictive analysis are social networks such as Facebook, that collect data from every action their users make (Monnappa, 2017). Based on consumers gender, age, interests and action, marketers are able to, not only identify, but also predict which customers are more likely to purchase their products and services. Big Data and it social media aspect has thus changed a number of industries, particularly those marketing to the end-consumers. Small businesses but also large ones are able to gain a so-called 360 degree view of their customers, meaning they not only know how they consumers act and look like, but what their occupation is, how they communicate and what they anticipate from the brand (Woodie, 2016). This is particularly important for the retail industry. Retail giants such as Amazon or Ebay depend heavily on Big Data. From a smaller scale such as offering consumers new products based on their purchase history, to trend forecasting and price optimization based on their competition. This forecasting trend has also developed a new kind of a retailer, one that provides the end-consumer with options based on their preferences. Stitch-fix is an online subscription service founded in 2011. The concept of the personal styling service developed rapidly through the use of social media and by implementing algorithms that help determine which outfits and clothes their consumers would prefer. While the first impression comes from the questionnaire the customer solves before the first purchase, with every additional purchase, the service improves, as it receives feedback based on the clothes the customer chooses to keep or return, thus creating data specifically for this customer (Murray, 2017). With digital marketing tactics and aggressive social media efforts such as PR boxes, the service has developed a following and a $730 million revenue (Stitch Fix Newsroom, 2017). This concept has however not only been confined to the fashion industry. The food industry has also started to see Big Data in conjunction with consumer-behaviour predictions. Services such as Blue Apron and Freshly bring the same concept to their customers. By analysing their needs and current trends, these services provide fresh produce and ingredients, as well as recipes with which to use the ingredients.

However, to provide successful service on almost any level, storage capabilities, analysis and management are essential. The following chapters will examine how businesses adapt to these Big Data requirements.

3.1.1 Storage Capabilities

Storage is one of the challenges of Big Data, particularly for smaller businesses. Constant flow of data coming from various sources first needs to be stored in an accessible manner for companies to extract value from them. The previous chapters have already discussed some storage possibilities that businesses can choose to use. From custom solutions to open-source frameworks and cloud computing, the development of new technologies is constantly bringing advancement in storing vast data amounts. However, the type of storage businesses choose will not only depend on their needs but also their ability to invest in them, as data warehouses and other storage capabilities can require large funding. Large companies such as Facebook, Google, Apple and other often have whole buildings devoted to Big Data storage. Those are called hyperscale computing environments and they include commodity servers with direct-attached storage, or DAS (Adshead, 2013). While there is no shared storage in this type of environment, these systems also often run on Hadoop and NoSQL to process their data through clusters (DeZyre, 2016). However, these large environments are incredibly costly to maintain. For instance, Google's capital investments increased by 10% in 2016, to $10.9 billion dollars reflecting the heavy investments in data facilities and their infrastructure (Sverdlik, 2017). Figure 4 shows capital spending of its parent company Alphabet Inc. The recent spike in these investments suggest an upward trend in the future.

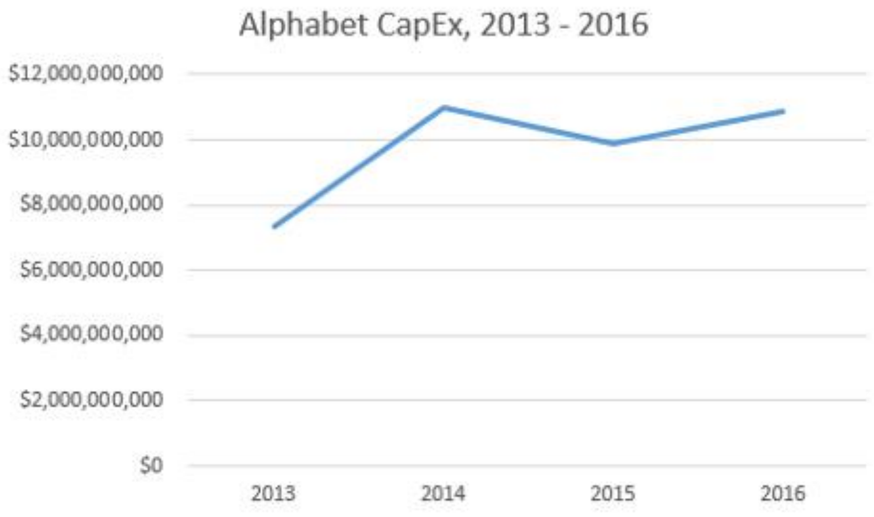

Figure 4. Capital Spending of Google's Parent Company from 2013-2016

Still, the need persist for smaller companies to implement Big Data initiatives, thus being inclined to find storage capabilities that venture outside of the hyperscale large enterprises are able to use. The already mentioned DAS storage systems are a fast but expensive option that may not suit the needs of smaller companies that do not need to process large data sets as fast as, for instance Facebook or Google, which is why alternatives such as NAS or network attached storage are gaining more and more popularity. However, NAS systems are easier to set up and do not require specialized knowledge in storage options, but if multiple devices are incorporated simultaneously, it does not necessarily mean that they are scalable and can be managed efficiently (ComputerWeekly.com, 2017). Rather, every device acts as its own sphere of data, making the management of data challenging. This is why companies are starting to turn to clustered NAS, which enable cohesion between individual elements, thus creating a more scalable and approachable data environment (Adshead, 2013).

The rising need for storage capabilities has also lead to the creation of a new market place, namely, storage providers that are used by smaller to middle sized businesses. Cloud service and content direct service or CDN are becoming the prominent options storage providers offer. Depending on the company, they may choose one or the other, or incorporate both (Ferkoun, 2014). This is the particular case with Netflix, the American entertainment company founded in 1997. Netflix deploys both cloud computing and CND, with their own internal CND and cloud computing by Amazon's AWS Cloud. Ready-made storage services from providers such as Stackpath or Akamai can start as low as $400 to $8000 per month, depending on data volumes and the needs of the business (Pierson, 2017).

3.1.2 Management of Big Data

The recurring thread throughout the themes of Big Data is often data management. Its size and heterogeneity often leads to management challenges such as low scalability due to a multitude of systems and databases (Gandomi and Haider, 2015). This is why, management is a topic often discussed in business circles. As already mentioned, to implement Big Data, companies first need to know, which business questions data should answer. One such example is the grocery chain from Venezuela, who had issues with its management. The company already collected Big Data, but being dispersed over a multitude of databases, it did not provide valuable insights. Additionally, without knowing what issues to address, data tracking was challenging and inefficient. Their particular problem was with pricing and distribution, posing a problem for the company to examine where their assets were. To solve the issue, the company has invested in information integration across the board, creating a unified system that could be used to analyse market conditions and adjust to them (Schroeck et al., 2012). While initial costs of such integration may be substantial, most companies choose to adjust their management tactics nonetheless. According to the Industrial Insights Report for 2015 in collaboration with General Electric, the American conglomerate corporation, 84% of correspondents were of the opinion, Big Data will change the competitive landscape within industries. Another 89% have stated that lack in Big Data adoption will lead to risks of losing market share (Bloom, 2015).

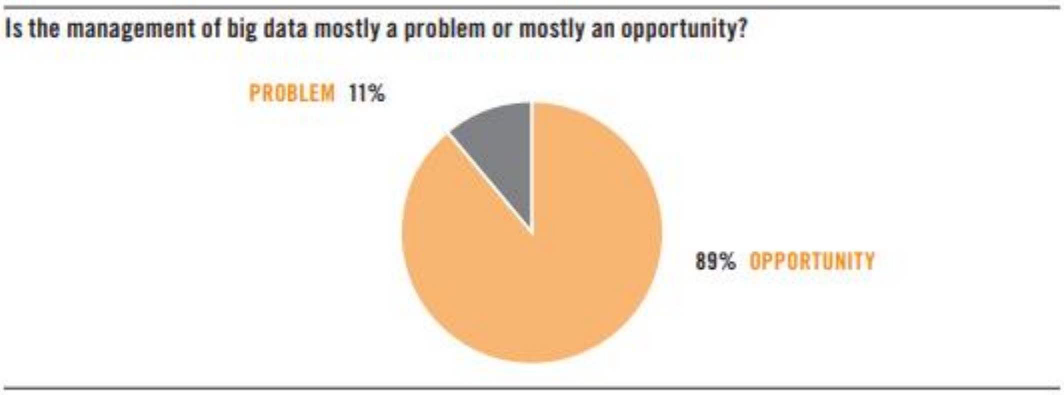

Figure 5. TDWI Survey - Is Big Data Management a Problem or an Opportunity?

However, the two main problems with Big Data management lie in the approach companies often implement, either to cut costs or due to the lack of expertise in how to incorporate Big Data. While companies across the board are starting to put considerable efforts into collecting data, creating large databases and organizing them into spreadsheets, they still lack the effective usage of that data (Carter, 2014). This is mostly due to the fact that databases are often created without a specific goal or problem in mind, thus creating databases of raw data with little to no value. Another issue lies on the other side of the spectrum. Namely, the large volumes of data often require some cleansing and centralizing. However, if these processes take the focus of Big Data projects, the acquired databases may not provide an accurate picture due to leaving out important aspects to create an organized database (Butler, 2015). Nonetheless, despite the issues, Big Data management is perceived as an opportunity, rather than a challenge. Figure 5 demonstrates this through a survey conducted by TDWI among 461 respondents, professionals in the data management field (TDWI Research, 2013).

3.1.3 Analysis of Big Data in Business Settings

Once data is gathered from various sources, it needs to be analyzed to provide value and insights. With the development of new technologies and tools, such as Hadoop or NoSQL based systems, businesses have a variety of ways that large data sets can be analyzed. However, the approach often depends on the size of the company, the issues that need to be addressed and the financial efforts that are necessary (Mohanty, Jagadeesh and Srivatsa, 2013). Companies thus, can choose to implement ready-made open-source Big Data solutions or they can develop their own frameworks.

Analysis of Big Data in business setting often pertains to specific problems within the company's operational or strategic efforts. These can include broad business strategies, definition of the business scope, marketing initiatives, but also logistic problems. For instance, a retail business may conclude that their supply chain is inefficient. Using Big Data and analysing their data on all parts of the supply chain, the business can identify weak spots, thus reducing costs that arise due to the inefficiency. A particular example that many retailers use is by implementing visual delivery routes. While traditional routing software has become the norm for logistics problems in delivery, the need for more accurate and timely data is becoming more and more prominent (Woodie, 2016).

By implementing Big Data from cloud-based systems which provide real-time data and geoanalytical mapping techniques, creating an accurate map that adapts to the current situations in traffic, businesses are able to save 15-20% on supply costs (Brocca, Kotlik and Greiser, 2014). Other examples include predicting consumer behaviour, personalizing in-store experience and increasing conversion rates by analysing consumer information and interactions (Hitchcock, 2017). These techniques are utilized by online retailers such as Amazon, but also brick-and-mortar based businesses such as Walmart or Nordstrom. However, Big Data analytics are not only implemented in the retail industry, but also telecommunications, banking, manufacturing and others. This is characterized by Figure 6 , showing which industries have the largest investments in Big Data. Most notably, telecommunication, followed by travel and airlines demonstrate the highest investment in Big Data analytics. Energy and resources show the lowest, however still notable, investments.

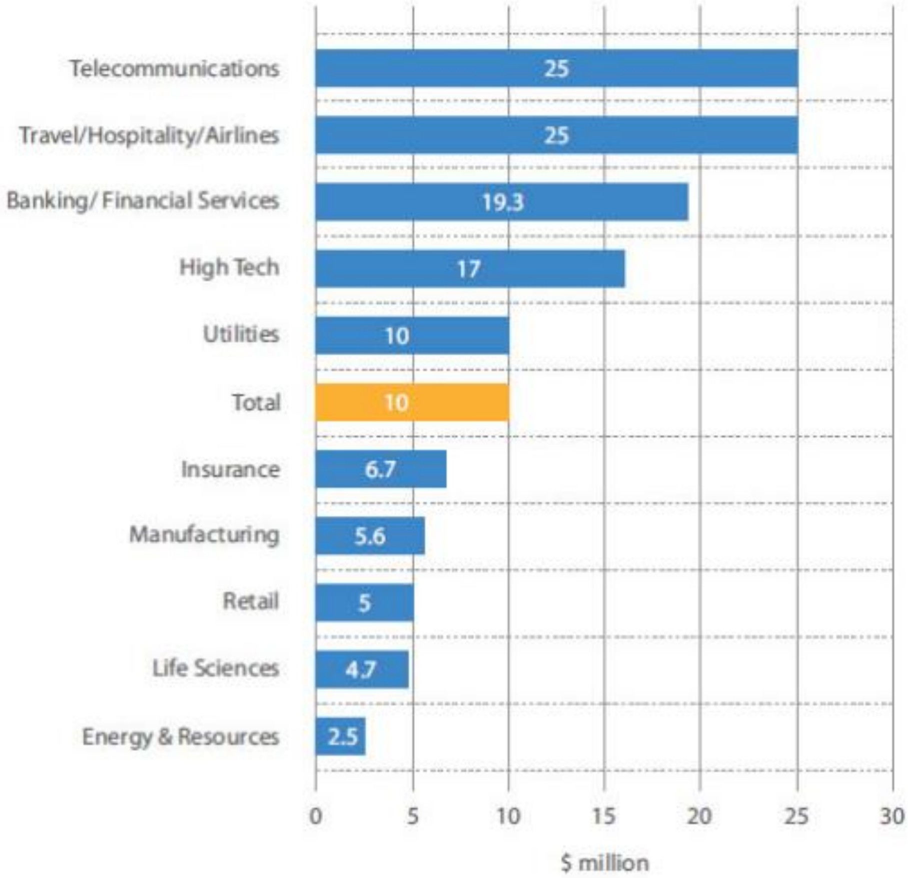

Figure 6. Big Data Investments by Industry

3.2 Big Data Analysis in Small and Middle-sized Businesses

Big Data is often mentioned in terms of high volume, velocity and variety requiring data warehouses, specialized analysing tools and data scientists to properly store, analyze it and lastly, utilize within business strategies (Gandomi and Haider, 2015). Due to the complexity of its characteristics, Big Data is perceived as a costly investment that does not generate enough return on investment, or ROI for smaller businesses to justify the significant cost it may require (Jones, 2017). However, the advancement in open-source data analysing tools and the development of affordable storage solutions has enabled Big Data utilization even for smaller and middle-sized businesses. These can range from free tools such as Google Analytics and Facebook which provide insights, for instance, about the demographics and interests of the target audience, to subscription tools which can target more specific problems in a visually coherent way (Lebied, 2017). The subscription-based tools are often referred to as self-service analytics, as they enable the user to conduct the analysis through intuitive interfaces without the need to invest in data scientists (Wicks, 2016). Additionally, a completely new market has also developed revolving around Big Data solutions on a smaller scale. Companies such as Canopy Labs, Kissmetrics, Qualtrics and others cater specifically to small businesses. However, even larger companies such as IBM or SAS also created tools that can be implemented, not only by large enterprises, but also smaller companies.

When it comes to storage solutions on a smaller scale, the development of cloud technology has enabled an alternative to large data warehouses (Angeles, 2016). Google Drive, Google Cloud Platform and Amazon Web Services or AWS are just some of the tools available for cloud storage. The key points of cloud storage lie in its simplicity, easy access from any smart device and low investments that it requires. However, some privacy and security issues, as well as potential increase in costs as the company developes are some of the risks associated with cloud storage. Still, tools such as Apache Hadoop, Hive, PolyBase and others also offer storage solutions that can be implemented even in small business settings. Apache Hadoop is a particular example which is often integrated within, or used as a base for other tools such as Google Cloud Storage (Patterson, 2015). Hadoop is developed for processing large clusters of data across distributed servers. One of its main advantages lie in its capability to extract data from unstructured data sources. Apache's Hadoop project has become the main tool for Big Data storage and processing. However, it does cater mostly to larger enterprises due to the technical skill it requires. A study by Murthy and Bowman published in the Journal Big Data & Society dealt with this particular issue. The researchers attempted to use Hadoop to store and analyse tweets from Twitter and examine what opportunities and challenges lie in the usage of Hadoop for smaller data sets. The authors created two separate storage and analysis setups, one a traditional relational database management system, or RDBMS, the other a Hive-Hadoop setup. Hive is also an Apache project built on top of Hadoop, enabling data summarization and analysis (Hive.apache.org, 2017). Their experiment found that the Hive-Hadoop setup processed data faster and more efficient that the traditional RDBMS approach at a lower price point. However, the Hive approach, while efficient in data processing, proved to be lacking when it came to extracting specific data. Additionally, in smaller systems, the risk of exhausting data storage as the business progresses does persist and needs to be kept in mind when exploring Big Data solutions on a smaller scale (Murthy and Bowman, 2014) Still, as more and more businesses enter the realm of data-based decision making pertaining to Big Data analytics, many analytics tools and Big Data storage providers adapt their products to Hadoop, creating an easy alternative for business users that operates on a basis that is well-established in Big Data efforts.

What tools small and middle-sized businesses may adapt within their Big Data efforts will depend on their goals and needs. As larger enterprises have the funds and capabilities to explore Big Data possibilities and employ data scientists for predictive or prescriptive types of analysis, smaller companies often have to identify their goal beforehand and focus on the particular issue to justify the investment Big Data can require and gain insights that can be used to solve a particular issue, rather than data that does not prove to be useful (Brown, 2016).

3.2.1 Opportunities

Generally, Big Data's main advantage is its ability to provide insights through analysis of large data volumes being gathered from various internal and external sources, ranging from social media to transactions or supply chain information (Kim, 2015). These opportunities do not change with size. As large enterprises are able to implement Big Data to improve and develop their business strategies, small to middle-sized companies also have the same opportunity. However, due to the cost of Big Data storage and analysis, smaller companies turn to ready-made solutions both in terms of storage and analysis. As already mentioned, for small businesses in particular, Big Data initiatives start with a goal or a problem that needs to be addressed, rather than relying on large data sets without a specific purpose in mind (Reisenwitz, 2017). The goals can range from providing a 360 degree view of potential customers to identifying challenges within company frameworks to reduce costs and improve efficiency. However, Big Data almost always requires a strategic approach regardless of the size of the company (MIT Sloan Management Review, 2016).

An example of successful Big Data usage on a smaller scale is the company Audience Audit. Audience Audit provides custom segmentation solutions that enables custom insights for marketing agencies and their clients (Baier, 2017). Founded by Susan Baier, a marketing strategist, the company has decided to focus on providing valuable research for their clients catering specifically to smaller businesses. However, quality research can be time-consuming which is why Audience Audit decided to implement Big Data solutions to cater to a wider array of clients without sacrificing their quality. According to the founder, Susan Baier, the company develops surveys and provides analysis with a base in Excel. Their clients vary greatly in their demands and needs, which is why the company implemented Tableau as part of their Big Data efforts (Tableau Software, 2017). Tableau is a data visualization tools that enables rapid data processing from various data sources providing a clear visualization in form of dashboards (Tableau Software, 2017). By implementing the tool, Audience Audit has managed to lower their fees and reduce project timelines by 50% as it enabled a more streamlined approach to data analysis and visualisation (Fields, 2014).

Another example how Big Data offers opportunities is the company Twiddy & Company Realtors. The company was founded in 1978 by Doug Twiddy offering vacation home rentals (Twiddy & Company, 2017). To identify problem areas and cut maintenance costs, the company used Big Data to examine maintenance charges and cost effectiveness. By implementing a ready-made solution in form of SAS Analytics, the company has leveraged Big Data to its advantage (SAS, 2017) The process has enabled the business to eliminate processing errors lowering their costs by 15%. Additionally, the end-result included increase in bookings and a 10% increase in company inventory (Kelleher, 2014).

SAS Analytics is one of the more popular data solutions for smaller to middle-sized businesses. With a list of clients ranging from universities, hospitals to retailers, fashion brands and others, the management software provides a tool that enables data-based decision making. Harry & David Holdings Inc., an American premium food and gift producer operating from retail stores and ecommerce is also an example of Big Data utilization, but with a slightly different issue that had to be addressed. Founded in 1910, the company had to adapt to technological developments to enable growth and remain relevant in their industry. This was particularly the case in 2011 when the company filed for bankruptcy due to the economic crisis which started in 2008 (Ovide, 2011). Thus, the company needed to adapt and they did so by shifting their focus to personalization and segmentation. Their main problem was in the outdated infrastructure and scarce investments in new technologies. However, by implementing SAS Analytics, the company started to segment their customer base to identify which customers bring high-value.

Additionally, through the use of outside data and its analysis, the company was able to predict consumer behavior and sales, enabling careful strategic planning pertaining to their marketing efforts. To characterize, instead of marketing the product itself, Harry & David Holdings Inc. used Big Data to identify consumers' needs and wishes, crafting their marketing strategies according to the analysed data (SAS, 2017). This has led to an increase in profits of 20%, as well as an increase in customer retention by 14% (SAS, 2017). The surge in revenue is notable in 2013, when the company started to implement Big Data within their frameworks. Since then, while total revenue growth has been declining, Harry & David Holdings Inc. maintain a steady growth of 1.7%. Reporting a yearly revenue of 554.6 million for the fiscal year 2016 (eMarketer Retail, 2017). Their growth is presented in Figure 7 in the timespan of 2013, shortly after the filed bankruptcy. The figure presents total revenues, as well as total revenue growth.

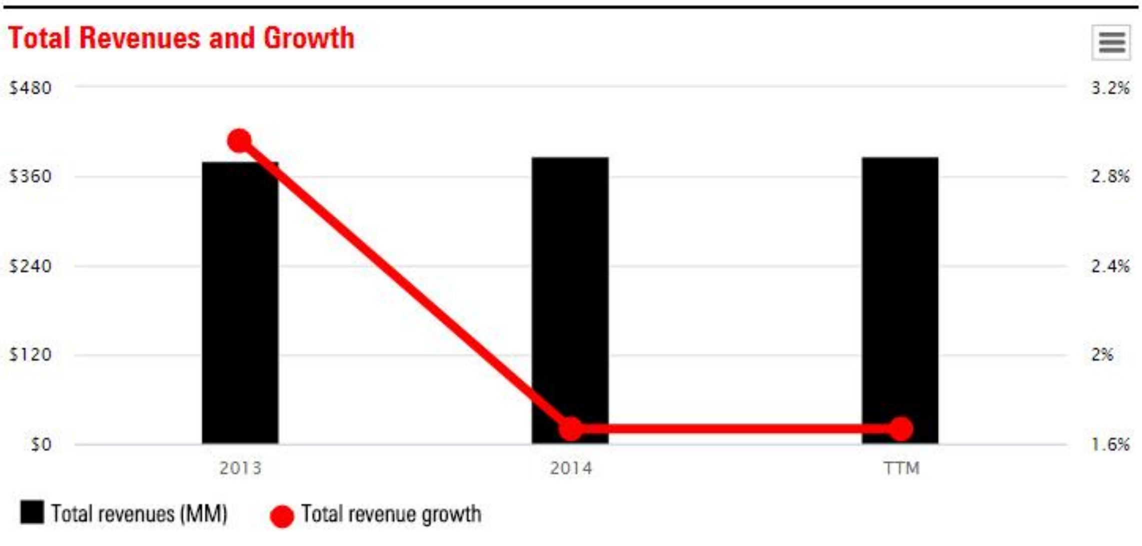

Figure 7. Total Revenue and Growth of Harry & David Inc.

To further characterize, small to middle-sized businesses do have a wide array of opportunities when it comes to Big Data utilization. Streamlining data management, creating customer-specific marketing messages, improving operational efficiency and reducing operational expenses are just some of the opportunities Big Data offers. However, selecting the right analyzing tool, as well as a keen focus on which issues need to be addressed are the focal points of successful Big Data implementation within business strategies and company efforts. Additionally, challenges and risks pertaining to Big Data usage on a smaller scale also need to be addressed.

3.2.2 Risks

As data is being gathered from a number of sources, external, as well as internal, opportunities for data-driven decisions are being constantly created. However, due to the gathering of data from almost any device, as well as its volume bring forth some issues and risks. These risks mostly pertain to privacy issues and security of data. Additionally, as large data amounts are being gathered, often once the project is finished and the issue resolved, the companies keep the data nonetheless for fear of losses making them more vulnerable to attacks and privacy concerns in turn. In addition, growing data amounts lead to an increase in storage needs creating new costs for storage capabilities.

The main issue with Big Data lies in the possibility of data breaches that can strike large enterprises, but also small to middle-sized businesses. According to Brian McGinley, the Identity Theft 911 Senior Vice Presiden, smaller businesses are particularly vulnerable to cyber attacks and data breaches as they often lack in security protocols which would reduce the risk of attacks or curb the financial hit if such attack would happen (McGinley, 2011). Brian McGinley further characterizes that any business that handles private customer and employee data is at risks but cites legal practice and insurance agencies as particularly vulnerable due to the detailed customer data that they gather from their clients. Such data often includes social security number and credit card records putting a number of consumers at risk if their data is not well protected.

Equifax Inc., a consumer credit reporting agency collecting information on over 800 million users experienced this particular type of security data breach, endangering information from 143 million users (King, 2017). The attack launched an investigation to determine the hackers and security issue which enabled the compromisation of data. It was concluded that the attack occurred due to a security flaw in the Apache Struts software (Cybersecurity Incident & Important Consumer Information, 2017). While one of the larger examples, Equifax security breach is not an isolated incident. According to a report from Gemalto, stolen or lost data experienced a 164% rise since 2016 putting numerous consumers and business individuals at risk of identity theft and damaged personal credit (Gemalto, 2017). Additionally, according to Ernst & Young's global forensic data analytics survey, 665 companies have identified cyber breach and fraud risks as the main issues pertaining to large data sets (EY, 2017). This is also partly due to the growing amounts of unstructured data which can include personal emails, social media interactions and voice messages. The heterogeneity of such unstructured data is thus particularly vulnerable to security breaches as its nature poses a challenge in security protocols and frameworks developed to maintain the privacy of the users. With the rising risk, anti-fraud and anti-corruption initiatives have started to emerge on the field of Big Data. Figure 8 demonstrates the concerns companies identify as their main problem when it comes to Big Data.

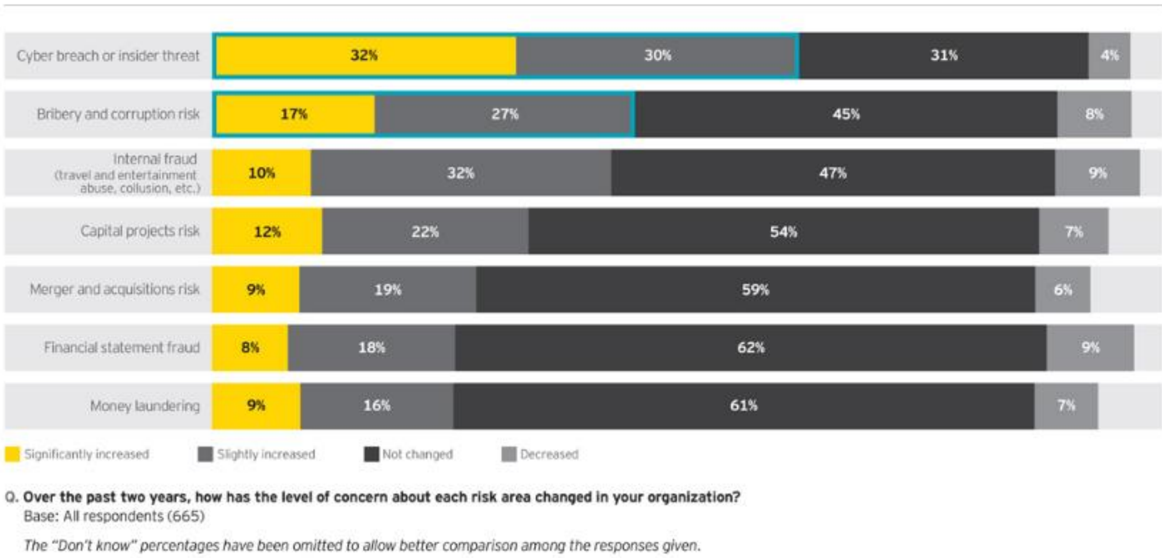

Figure 8. EY Global Forensic Data Analytics Survey

Cyber breach and insider threat are the leading problem with 32% of correspondents identifying a slight increase. In accordance to the security question, one particular term that has started to emerge is Forensic Data Analytics, which are used to determine and prevent suspicious behavior, transactions and patterns (EY, 2017). Recently, companies have started to invest additional resources towards FDA initiatives. Figure 9 presents the top drivers for such decisions among the 665 surveyed companies. It should however be noted, that from the 665 surveyed companies, 405 of them reported future plans for Forensic Data Analytics investments. With 53%, growing cybercrime risks are cited as the main reason with increased regulatory scrutiny following with 43%. Increasing merger and acquisition activities, as well as whistle-blower calls and reports show the lowest, however still notable percentage, with 13% and 9% respectively. Of the surveyed companies, 4% reported additional FDA investments, however cited other reasons as the focal point of such decision.

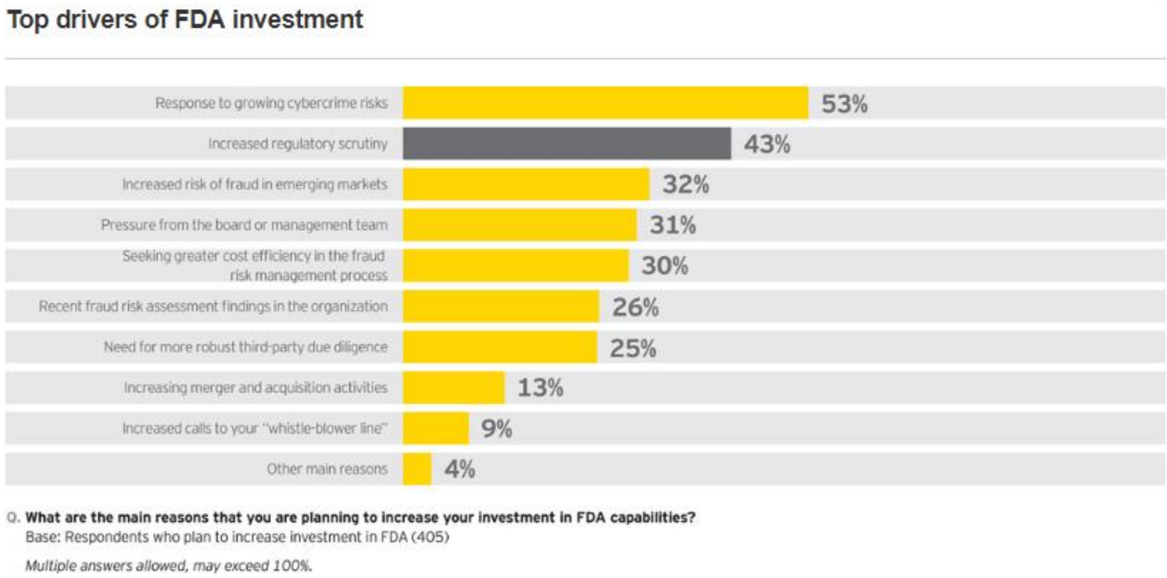

Figure 9. EY Global Forensic Data Analytics Survey, Top Drivers of FDA Investment

As mentioned, there are many flaws hackers can use to conduct an attack and gain valuable consumer information. One such example comes from the company VTech, founded in 1976. Operating from Hong Kong, the company develops learning toys catering to infants, preschoolers and preteens. Due to the nature of their business, the company relies on sign up forms for their products, thus gathering sensitive data such as consumers' gender, age, name and date of birth. Specifically, the hacker targeted the Learning Lodge, Kid Connect, PlanetVTech and V.Smile Link data, which include data from the children's' profiles to parent information etc. However, since the databases do not contain any financial data, no credit card information was stolen (FAQ about Cyber Attack on VTech Learning Lodge, 2016). The attack was conducted through an SQL injection, a code technique which places a malicious code in SQL statements which enables the attacker to retrieve data by going around authorization processes (Acunetix, 2017). This is one of the most common types of hacker attacks. The attack resulted in significant losses for VTech, not only in revenue, but also brand equity as such security breaches created a sense of mistrust among its customers. According to the company's report, 4 863 209 parent accounts and 6 368 509 children's profiles were affected, resulting in significant revenue loss, as the databases did not function over the course of two months (FAQ about Cyber Attack on VTech Learning Lodge, 2016). As such, VTech's weak encryption and general lack of security resulted in an incident which affected the company negatively. However, these particular incidents do show an increasing need for stronger network security across the board. Data security, privacy, as well as bad analytics and bad or irrelevant data can create enormous losses for businesses (Marr, 2015). Data security and data privacy need to be maintained to protect not only consumer data, but also internal company data. On the other hand, bad analytics and false or irrelevant data may lead to the wrong conclusion, thus generating losses instead of providing applicable insights.

3.2.3 Management Challenges

While management challenges of large volumes of high-velocity data are found on every level, as well as in businesses of all sizes, due to the limitations in capabilities and funding, smaller to middle-sized companies often struggle with Big Data. The main problem occurs if business owners do not start with an overarching data strategies that covers the Big Data basics, such as where and how will data be stored, as well as how the stored data will be analysed. Without such strategy, collected data is often processed without a purpose in mind often leading to miscalculations or missteps in business strategies (Murthy and Bowman, 2014).

One of the most common ways management of data is handled within smaller companies is by addressing operational issues. According to IBM Institute for Business, 40% of companies incorporate Big Data to solve problems pertaining to operational challenges (Schroeck et. al, 2012). However, companies are also using Big Data to determine their marketing strategies, which in turn often leads to analysis of unstructured data. The main challenge in analysing such data is that it can often be incomplete, outdated or of poor quality. This can pertain to the simplest mistakes, such as spelling errors in names or variations in information which, amplified, can lead to faulty conclusions (Petty, 2016). One of the reasons for bad data which in turn leads to unsuccesful management is lack of a coherent project strategy and investments in checking the data quality (Marr, 2015). Additionally, without a coherent strategy, relevant data can get lost in the vast amounts of data that may not be necessary for the project at hand. According to Bernard Marr, a strategic performance consultant and Big Data guru, this is one of the main problems in Big Data management, as reports are often overstuff with irrelevant information. Thus, his approach centers around clear data presentation with easy-to-understand graphics addressing only the relevant problem (Marr, 2015).

Security and privacy are also challenges that need to be addressed by the company's management. Since small and middle-sized companies often outsource their Big Data initiatives, it is essential to choose reputable storage alternatives and analysing tools. One such example is cited by Jen Cohen Crompton, a writer, digital strategist and a small business owner. She founded FUEL Cycle Fitness in 2015, a small fitness studio, also implementing Big Data initiatives. To address the issues of privacy and security, the company choose to store their data through an outsourced site, with the ability to export and store content within their own frameworks. However, they have limited the access of the outsourced service to consumer-sensitive data such as credit card numbers, thus creating a security barrier that protects the consumers, as well as the company itself (Cohen Crompton, 2017)

Another notable point in Big Data management is the human factor. While intricate analysing tools do provide valuable insights, it is ultimately up to the management to incorporate the gained information within business strategies. To do so, conflicts and human biases need to be addressed in a way that provides an accurate picture, rather than focusing on one aspect (Petty, 2016).

3.2.4 Overall Performance

With new technologies and the development of companies providing Big Data solutions, small and middle-sized companies are able to utilize Big Data within their frameworks, without investing notable resources that may not generate enough Return on Investment, or ROI to justify the Big Data initiatives. The concerns for Big Data initiatives within companies lies in the heterogeneity of Big Data that requires specialized infrastructure to accommodate it. A general agreement is that Big Data is still perceived as challenging but also necessary for further growth. A study conducted by Knowledgent, a data analytics and consulting company shows the main concerns with Big Data implementation, citing infrastructure as the main challenge, with qualified resources showing following behind. It is also notable, that the highest percentage of correspondents cited qualified resources as extremely challenging (Knowledgent, 2015).

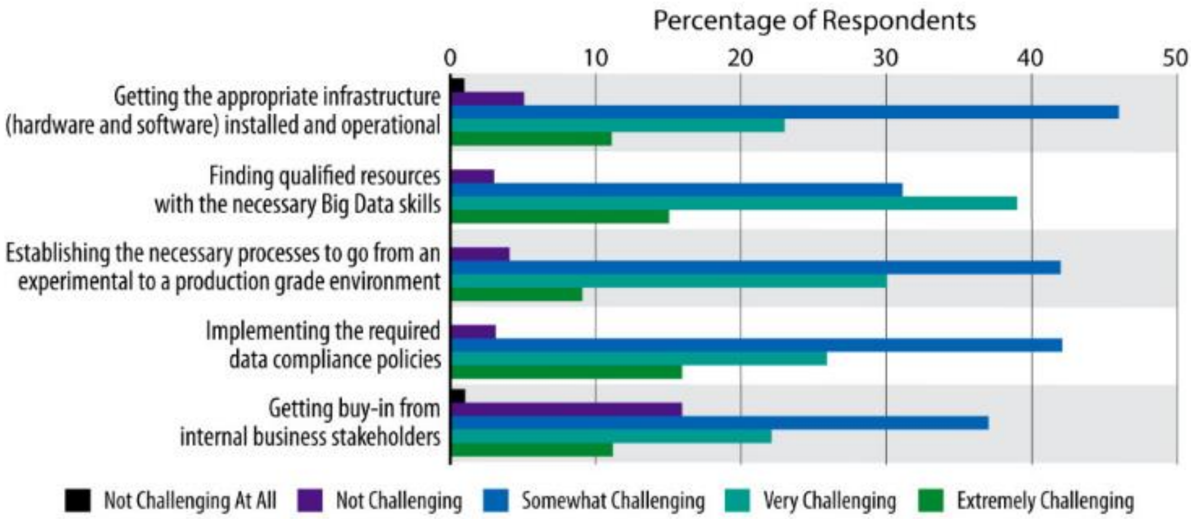

Figure 10. 2015 Big Data Survey: Current Implementation Challenges

However, despite the challenges, Big Data initiatives are experiencing a rise in implementation with companies either already implementing Big Data or developing plans to do so.

Thus, the overall performance of Big Data initiatives shows positive results due to the development of new technologies which enable data analysis for small businesses as well. Nonetheless, an overarching strategy is one of the key components for successful implementation, as it prioritizes which problems and issues need to be addressed (Quinn, 2017). Still, it should be noted that the current research of Big Data initiatives mostly caters to larger business enterprises, as the data on small to middle-sized companies often lacks proper analysis which would include the cost of Big Data initiatives in accordance to the success rate.

3.3 Big Data Analysis in Large Enterprises

Due to the available funds and human resources, large enterprises are first to utilize Big Data initiatives, as well as develop new tools that can be outsourced to smaller companies. Depending on the type of industry and company goals, Big Data trends and initiatives vary from company to company. Business giants such as Google, Facebook, IBM and others are constantly improving their Big Data initiatives often devoting considerable investments to data warehouses and analyzing solutions. According to surveys from Accenture, a global management consulting firm, IBM as well as GE or General Electric, 92% of company executives experience positive Big Data results, with 82% citing Big Data as very or extremely important (Burger, 2014). Additionally, with the development of new technologies, the maintenance cost of Big Data has also experienced a cost decrease. Namely, storage and bandwidth costs have dropped on an annual basis by 38% and 27% respectively (Bloom, 2015).

As opposed to Big Data solutions on a smaller scale, large companies often incorporate their own data warehouses as well as a team of data scientist. To characterize, larger companies incorporate more complex Big Data initiatives which operate on a much larger scale. One of the most notable example of Big Data initiatives comes Wal-Mart Inc, an American retailing corporation founded in 1962. Walmart operates 20 000 stores in 28 countries making it the largest retailer worldwide. However, Walmart has successfully integrated new technologies within their company, namely machine learning and the IoT or Internet of Things. As such, Walmart generates vast amounts of data, that need to be stored and processed. As such, Walmart is currently building a large private cloud with processing capabilities of 2.5 petabytes per hour (Marr, 2017). Their Big Data usage varies from relatively simple product stocking problems to initiatives with machine learning and facial recognition. In the aforementioned stocking example, Walmart's analysts noticed a particular cookie which experienced high sales in most stores, except for two. By noticing the anomaly, the analysts were able to conclude that a stocking mistake caused the cookies to not be even placed on the shelves, thus explaining sales declines (Marr, 2017). Such automated alerts when certain metric change enable the company to quickly adapt to changes and issues, thus curbing potential revenue losses due to mistakes. Another example of their Big Data efforts is in the managing their supply chain by leveraging Big Data finds to determine the most efficient routes for delivery which in turn cuts their delivery costs as it helps schedule their deliveries and decreases possible problems along the route (Amato-Mccoy, 2017). Personalizing the shoppers preferences is another popular Big Data initiative that is not only incorporated by Walmart, but by retailers such as Amazon or Ebay as well. These initiatives revolve around analysing the past and current purchases of the customer, thus extracting information which predict the customer's future purchases. Companies are then able to create a more personalized shopping experience and craft deals accordingly.

However, the retail industry is not the only industry that incorporated Big Data on a larger scale. Business giants such as Facebook and Google do not only create Big Data initiatives within their own frameworks but also outsource their tools to smaller and middle-sized businesses, thus enabling a circle of Big Data usage. For instance, Facebook, the largest and most popular social media network has changed the way online marketing works. This is particularly notable in marketing efforts of smaller companies and brands, with social media providing a platform which can target their specific audience. As such, Facebook offers customers insights and advertising space by collecting consumers' data, analysing it and segmenting to be further used for marketing initiatives (Marr, 2015). The role of social media marketing has thus spawned its own industry focusing solely on Facebook advertising and how it can be implemented within marketing efforts. Additionally, the Big Data success of Facebook can also be contributed to the personal nature of the platform, where users can post and interact with the content generating a consumer-centric approach (Chakraborty, 2017).

Similarly to Facebook, Google has also changed the way data is perceived and used. As the most popular search engine, Google focuses heavily on providing the best possible search experience while also developing solutions for Big Data usage. Their success can be contributed to intricate search algorithms which enabled them to generate search results based on keywords, additional ranking the pages in its index (Marr, 2015). Since, Google refined its algorithms and built its index by sending specific robots, called crawlers, across the web to analyze web pages and rank them (Google.com - The Fundamentals of Search, 2017). The robots scroll through the content of web pages ad copy them onto Google's data centres. Once the information are stored, the search engine is able to quickly match the search query to possible results.

3.3.1 Technological Requirements

Large data amounts require specific technological capabilities which can be a costly investment. This is why custom solutions and storage options are often the choice for larger companies as they leverage large data amounts across multiple projects. For instance, Google has its own data centers which accommodate and analyse the vast amounts of data it gathers. Additionally, to maintain the data warehouses, particularly for large companies, specialists in data analysis and management are also essential (Google.com - Data Centers, 2017). Facebook, for example, operates on custom servers and power saving technologies. However, as is the case with many large enterprises, it also operates on an Hadoop based system. Particularly for larger companies, Hadoop systems are the most common choice but rarely the only one. LinkedIn, a business- and employment-oriented service incorporates tools such as Oracle, Hive, Kafka and MySQL in addition to investments in Hadoop (Bloom, 2015). Other enterprises may also include third-party solutions within their strategies. Such is the case with the already mentioned Netflix, which incorporates Amazon Web Services or AWS, but also some traditional BI tools such as MicroStrategy (Marr, 2015). One particular characteristic of Big Data storage on a larger scale is the lack of cloud-based storage. Users such as Google operate from their own hyperscale data centers with network-attached storage, or NAS. This is due to the fact that, while cloud storage provides an efficient alternative to smaller and middle-sized companies, it is also challenging in terms of constant data transporting on a large scale, which is why, large enterprises still find NAS storage more cost-efficient (Lee, 2017).

However, as is the case with Big Data tools and software solutions for smaller companies, the choice of storage, analyzing tools and investments will depend on the goals and requirements of the company. Thus, while storage and processing capabilities are essential for large companies to extract value from Big Data, how they choose to organize the data will be different from company to company.

3.3.2 Management in Large Enterprises

Similarly to management of data in smaller businesses, large enterprises also need to maintain an overarching data strategy to successfully manage large data amounts. However, unlike smaller companies, the management of data will often include teams of analysts, data scientists and custom-made tools to accommodate the large data volumes being gathered at a high-velocity (Mohanty, Jagadeesh and Srivatsa, 2013). One particular challenge that is associated with Big Data in general is the quality of gathered data. With data being constantly generated from a number of sources, it is not uncommon for companies to have outdated data (Harvey, 2017). Additionally, Big Data within large companies can be stored in separate databases, that my not be connected to each other, losses can occur. One particular example is if a retailer has multiple customer information databases, but spread across the company's frameworks. If one of the informations in those databases is found to be lacking or faulty, it can lead to inaccuracies in the business. Such examples demonstrate the lack in strategic implementation of Big Data. However, companies are starting to identify and address those problems. By creating written strategies and data management policies, implementation of Big Data initiatives is considerably simplified across the organisation. Additionally, investing in employee training as well as hiring a chief data officer can also simplify Big Data management as it streamlines the ongoing data management processes. One example of successful Big Data management comes from the logistics company UPS or United Parcel Service. According to the chief information and engineering officer, Juan Perez one of their most notable current initiatives is the development of Network Planning Tools, meant to optimise UPS's logistics network by using artificial intelligence and real-time data to enable smarter usage of their assets (Samuel, 2017). The new initiative is scheduled to come out in 2018. However, the main point Perez cites as a downfall in management within large companies is focus on only one dataset (Samuel, 2017). One dataset may provide some answers but it fails to provide an overall picture that could impact the company as a whole.

3.3.3 Challenges and Benefits of Big Data

For larger enterprises in particular, the main benefits of Big Data initiatives pertain to improved decision making as subjective assumptions are being replaced with data-based decisions (Marr, 2015). Additionally, operational issues, customer retention and innovative marketing solutions can all be contributed to Big Data projects. However, as was the case with smaller companies, challenges pertaining to data credibility, human errors, storage and analysing issues are present in larger companies as well. A notable example of faulty Big Data analysis comes from the Big Data giant, Google. In 2008 Google's data scientists attempted to predict the flu based on user searches. While the algorithm first showed positive results at first, it proved to be extremely faulty as it correlated terms that were not actually connected, thus falsely predicting flu prevalence in 2013 by 140% (Lazer and Kennedy, 2015). However, Google's miscalculations have pointed issues with Big Data analysis and have shifted the focus to the creation of more robust analysing tools.

Nonetheless, despite the challenges, according to the Bain Big Data Diagnostic Survey presented in Figure 10, larger investments in Big Data analytics show more positive results and affect companies success.

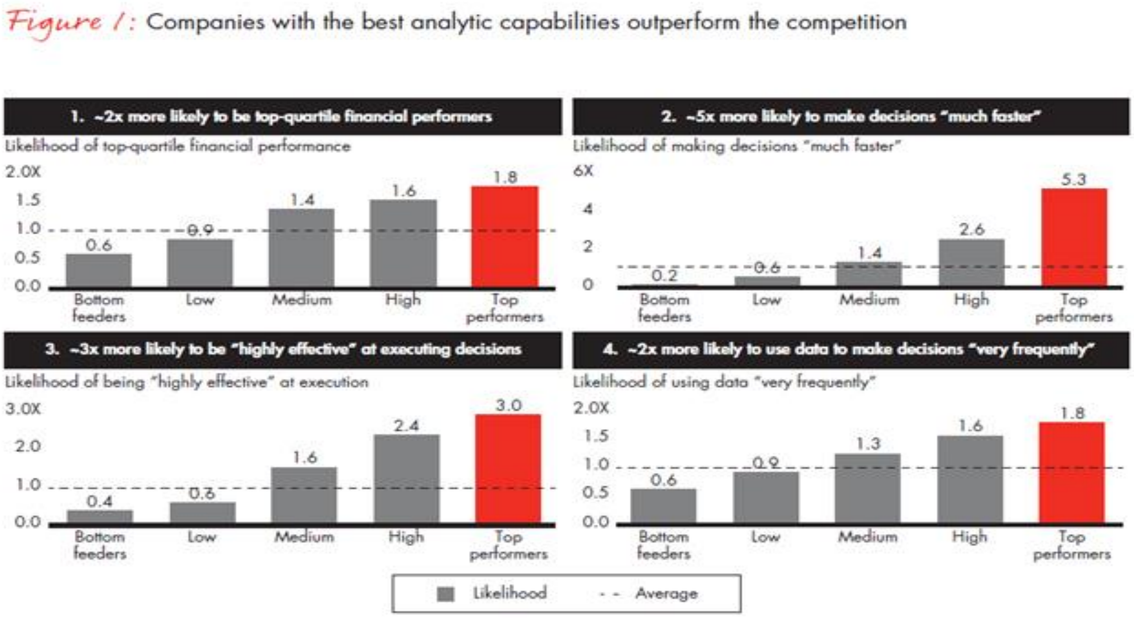

Figure 11. Comparison Between Analytic Capabilities and Performance

4. CONCLUSIONS

Through the research, the paper has presented an overview of Big Data analytics, as well as specific examples of Big Data utilization in small to middle-sized companies and large enterprises. While strategies pertaining to Big Data initiatives vary from industry to industry, some common points can be concluded. Big Data can offer valuable insights and applicable information if storage capabilities, proper analyzing tools and data from trusted sources is present. However, if there is no overarching strategy and specific problem Big Data needs to answer, regardless of the size of the company, it can prove to be a lacking investment as it will not provide sufficient information. This is due to the fact that without a long-term goal in mind, data becomes obsolete as it does not answer the companies problems. Thus it can be concluded that the first hypothesis is proven to be correct. The case studies in the practical part have shown that effective management of data does indeed tackle operational issues and streamlines processes within company frameworks. In conjunction, the second hypothesis is also correct, as the case studies have shown cost reductions if Big Data initiatives are managed properly. When it comes to smaller companies, Big Data solutions can be costly and time-consuming, but the development of outsourced, ready-made solutions has enabled Big Data usage even for small business owners without analytical specialty knowledge. Thus, the third hypothesis is only partially correct, as Big Data initiatives can be costly, but the development of open-source solutions has enabled usage across the board.

BIBLIOGRAPHY

Ackoff, R. (1999). Ackoff's best. 1st ed. New York: John Wiley & Sons, Inc.

Acunetix. (2017). What is SQL Injection (SQLi) and How to Fix It. [online] Available at: https://www.acunetix.com/websitesecurity/sql-injection/

Adamson, D. (2017). Big Data in Healthcare Made Simple. [online] Health Catalyst. Available at: https://www.healthcatalyst.com/big-data-in-healthcare-made-simple

Adshead, A. (2013). Big data storage: Defining big data and the type of storage it needs. [online] ComputerWeekly.com. Available at: http://www.computerweekly.com/podcast/Big-data-storage-Defining-big-data-and-the-type-of-storage-it-needs

Amato-Mccoy, D. (2017). Five ways Walmart uses big data |Chain Store Age. [online] Chain Store Age. Available at: https://www.chainstoreage.com/article/five-ways-walmart-uses-big-data/

Andreescu, A. (2013). Big Data for Personal Use Is More Popular Than Ever - SmartData Collective. [online] SmartData Collective. Available at: https://www.smartdatacollective.com/big-data-personal-use-more-popular-ever/

Angeles, S. (2016). 8 Big Data Solutions for Small Businesses. [online] Business News Daily. Available at: http://www.businessnewsdaily.com/6358-big-data-solutions.html

Assunção, M., Calheiros, R., Bianchi, S., Netto, M. and Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing, 79-80, pp.3-15.

Axline, A. (2017). How long can we keep gathering and storing data about everything? - CertMag. [online] Certification Magazine. Available at: http://certmag.com/long-can-keep-gathering-storing-data-everything/

Baars, H. and Kemper, H. (2008). Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework. Information Systems Management, 25(2), pp.132-148.

Baier, S. (2017). Why Audience Audit? - Audience Audit. [online] Audienceaudit.com. Available at: https://audienceaudit.com/why-us/

Bean, R. (2017). How Companies Say They’re Using Big Data. [online] Harvard Business Review. Available at: https://hbr.org/2017/04/how-companies-say-theyre-using-big-data

Bloom, A. (2015). 20 Examples Of ROI And Results With Big Data. [online] Content.pivotal.io. Available at: https://content.pivotal.io/blog/20-examples-of-roi-and-results-with-big-data

Brocca, M., Kotlik, L. and Greiser, C. (2014). 3 ways big data can improve your supply chain. [online] World Economic Forum. Available at: https://www.weforum.org/agenda/2015/05/3-ways-big-data-can-improve-your-supply-chain/

Brown, J. (2016). Big Data: 3 Insights That Will Make You Actually Love Big Data. [online] Topbrainiac. Available at: https://www.topbrainiac.com/big-data-3-insights-that-will-make-you-actually-love-big-data

Bryant R. E., Katz R. H., & Lazowska E. D. (2008). Big-Data Computing: Creating

revolutionary breakthroughs in commerce, science, and society: A white paper prepared for the

Computing Community Consortium committee of the Computing Research Association.

Bryson, S., Kenwright, D., Cox, M., Ellsworth, D. and Haimes, R. (1999). Visually exploring gigabyte data sets in real time. Communications of the ACM, 42(8), pp.82-90.

Burger, A. (2014). Accenture Survey Finds High Level of Big Data Satisfaction. Telecompetitor.

Burgess, T. (2014). Internal vs. External Big Data: How Your Servers and Network Are Affected. [online] Samanage Blog. Available at: https://blog.samanage.com/insights/internal-vs-external-big-data-how-your-servers-and-network-are-affected/

Burns, E. (2016). Data visualization techniques, tools at core of advanced analytics. [online] SearchBusinessAnalytics. Available at: http://searchbusinessanalytics.techtarget.com/feature/Data-visualization-techniques-tools-are-core-of-advanced-analytics

Butler, B. (2015). 5 Problems with big data. [online] Network World. Available at: https://www.networkworld.com/article/2973963/big-data-business-intelligence/5-problems-with-big-data.html

Carter, K. (2014). Managing big data: The two biggest mistakes companies make. [online] SearchBusinessAnalytics. Available at: http://searchbusinessanalytics.techtarget.com/feature/Managing-big-data-the-two-biggest-mistakes-companies-make

Chakraborty, S. (2017). How is Facebook Deploying Big Data? - DZone Big Data. [online] dzone.com. Available at: https://dzone.com/articles/how-is-facebook-deploying-big-data

CHD Expert. (2017). Database Fundamentals: Internal and External Data Sources. [online] Available at: https://chdexpert.wordpress.com/2012/03/30/database-fundamentals-internal-and-external-data-sources/

Chen, M., Mao, S., Zhang, Y. and Leung, V. (2014). Big Data: Related Technologies, Challenges and Future Prospects. 1st ed. Springer.

Cleveland, W. (1994). The elements of graphing data. Murray Hill, New Jersey: AT & T Bell Laboratories.

ComputerWeekly.com. (2017). NAS management comes of age with clustered NAS. [online] Available at: http://www.computerweekly.com/podcast/NAS-management-comes-of-age-with-clustered-NAS

Cohen Crompton, J. (2017). 5 Data Management Challenges Facing Small Business Owners. [online] Entrepreneur. Available at: https://www.entrepreneur.com/article/281058

Connolly, S. (2012). 7 Key Drivers for the Big Data Market - Hortonworks. [online] Hortonworks. Available at: https://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/

Cybersecurity Incident & Important Consumer Information. (2017). Cybersecurity Incident & Important Consumer Information | Equifax. [online] Available at: https://www.equifaxsecurity2017.com/

Declues, J. (2017). Four Types of Big Data Analytics and Examples of Their Use. [online] Ingrammicroadvisor.com. Available at: http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-their-use

Devens, R. (2013). Cyclopaedia of Commercial and Business Anecdotes. New York: HardPress Publishing (Original work published 1865).

DeZyre. (2016). Top 10 Industries using Big Data and 121 companies who hire Hadoop Developers. [online] Available at: https://www.dezyre.com/article/top-10-industries-using-big-data-and-121-companies-who-hire-hadoop-developers/69

Eckerson, W. and Hammond, M. (2011). Data Visualization Technology | Transforming Data with Intelligence. [online] Transforming Data with Intelligence. Available at: https://tdwi.org/Articles/2011/11/09/RESEARCH-EXCERPT-Data-Visualization-Technology.aspx?Page=1

eMarketer Retail. (2017). Harry & David | Revenues & Sales | eMarketer Retail. [online] Available at: https://retail-index.emarketer.com/company/data/5411e319ddb53b4e043d24d9/5411e31fddb53d765417c8ec/lfy/false/harry-david-revenues-sales

Ernst & Young Global Limited (2014). Big Data: Changing the Way Businesses Compete and Operate. 1st ed. EYGM Limited.

EY. (2017). Big risks, big data – and big decisions for the board. [online] Available at: http://www.ey.com/gl/en/issues/governance-and-reporting/ey-big-risks-big-data-and-big-decisions-for-the-board

Fan, J. and Liao, Y. (2014). Incidental Endogeneity in High Dimensions. Princeton University and University of Maryland.

FAQ about Cyber Attack on VTech Learning Lodge. (2016). .

Fields, E. (2014). The World of Big Data Is Now Open to Small Companies. [online] Entrepreneur. Available at: https://www.entrepreneur.com/article/238118

Ferkoun, M. (2014). Cloud computing and big data: An ideal combination - Cloud computing news. [online] Cloud computing news. Available at: https://www.ibm.com/blogs/cloud-computing/2014/02/cloud-computing-and-big-data-an-ideal-combination/

Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), pp.137-144.

Gartner IT Glossary. (2017). What Is Big Data? - Gartner IT Glossary - Big Data. [online] Available at: https://research.gartner.com/definition-whatis-big-data?resId=3002918&srcId=1-8163325102

Gartner IT Glossary. (2017). Business Intelligence - BI - Gartner IT Glossary. [online] Available at: http://www.gartner.com/it-glossary/business-intelligence-bi/

Gemalto. (2017). Gemalto releases findings of 2016 Breach Level Index. [online] Available at: http://www.gemalto.com/press/Pages/Gemalto-releases-findings-of-2016-Breach-Level-Index.aspx

Giannetto, D. (2015). Understanding The Structure of Big Data. [online] Adotas.com. Available at: http://www.adotas.com/2015/01/understanding-the-structure-of-big-data/

Google.com. - The Fundamentals of Search (2017). How Google Search Works | Crawling & Indexing. [online] Available at: https://www.google.com/search/howsearchworks/crawling-indexing/ [Accessed 9 Oct. 2017].

Google.com - Data Centers. (2017). Data Centers – Google. [online] Available at: https://www.google.com/about/datacenters/

Gupta, A. (2014). Why Extracting Value From Big Data Is Difficult. [online] Cioinsight.com. Available at: http://www.cioinsight.com/it-management/expert-voices/why-extracting-value-from-big-data-is-difficult.html

Hall, P., Dean, J., Kabul, I. and Silva, J. (2014). An Overview of Machine Learning with SAS Enterprise Miner. 1st ed. SAS Institute Inc.

Harvey, C. (2017). What is Big Data Management - Datamation. [online] Datamation.com. Available at: https://www.datamation.com/big-data/big-data-management.html

Harvey, C. (2017). Top Big Data Trends in 2017 - Datamation. [online] Datamation.com. Available at: http://www.datamation.com/big-data/big-data-trends.html

Hive.apache.org. (2017). Apache Hive TM. [online] Available at: https://hive.apache.org/

Huddleston, J. (2016). 5 Types of Big Data Analytics and How They Help Customer Success. [online] business2community. Available at: http://www.business2community.com/big-data/5-types-big-data-analytics-help-customer-success-01519563#qQRqgMl5sJ8GwPV7.97

IBM Analytics (2017). What is big data analytics? – IBM Analytics. [online] Ibm.com. Available at: https://www.ibm.com/analytics/us/en/technology/hadoop/big-data-analytics/

Ibmbigdatahub.com. (2017). IBM Big Data & Analytics Hub | Understanding big data for the enterprise. [online] Available at: http://www.ibmbigdatahub.com/

Information Technology Laboratory. (2017). What is EDA?. [online] Available at: http://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm

Ismail, N. (2017). How can a business extract value from big data? - Information Age. [online] Information Age. Available at: http://www.information-age.com/wheres-value-big-data-123467197/

Johnson, L., Adams, S., Cummins, M., Estrada, V., Freeman,

A., & Ludgate, H. (2013). The NMC Horizon Report: 2013 Higher Education Edition. Austin, TX: The New Media Consortium. Retrieved from

http://www.nmc.org/pdf/2013-horizon-report-HE.pdf

Joly, S. (2016). A Brief History of Business Intelligence - Blog | Pyramid Analytics. [online] Pyramid Analytics. Available at: http://www.pyramidanalytics.com/pages/blog/2016/02/brief-history-business-intelligence.aspx

Jones, M. (2017). It's Time for Small Business to Embrace Big Data. [online] Adage.com. Available at: http://adage.com/article/digitalnext/time-small-business-embrace-big-data/308191/

Kalyvas, J. and Overly, M. (2013). Big data: A Business and Legal Guide. Auerbach Publications.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. Application Delivery Strategies, META Group.

Kelleher, K. (2014). What 3 Small Businesses Learned From Big Data. [online] Inc.com. Available at: https://www.inc.com/magazine/201407/kevin-kelleher/how-small-businesses-can-mine-big-data.html

Kim, H. (2015). Big Data: The Structure & Value of Big Data Analytics. In: Twenty-first Americas Conference on Information Systems.

King, M. (2017). Can Your Small Business Survive a Big Data Breach? | Lendio. [online] Lendio. Available at: https://www.lendio.com/blog/can-small-business-survive-big-data-breach/

Knowledgent. (2015). 2015 Big Data Survey: Current Implementation Challenges. [online] Available at: https://knowledgent.com/whitepaper/2015-big-data-survey-current-implementation-challenges/

Lazer, D. and Kennedy, R. (2015). What We Can Learn From the Epic Failure of Google Flu Trends. [online] WIRED. Available at: https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/

ebied, M. (2017). Strategy Guide For Big Data & Business Intelligence for Small Business. [online] BI Blog | Data Visualization & Analytics Blog | datapine. Available at: http://www.datapine.com/blog/business-intelligence-for-small-business/

Lexalytics. (2017). lexalytics. [online] Available at: https://www.lexalytics.com/technology/sentiment

Lo, F. (2017). What is Hadoop and NoSQL?. [online] Datajobs.com. Available at: https://datajobs.com/what-is-hadoop-and-nosql

Lorentz, A. (2017). With Big Data, Context is a Big Issue. [online] WIRED. Available at: https://www.wired.com/insights/2013/04/with-big-data-context-is-a-big-issue/

Lyon, L. (2015). Ensuring Data Availability in Critical Big Data Applications — DatabaseJournal.com. [online] Databasejournal.com. Available at: http://www.databasejournal.com/features/db2/ensuring-data-availability-in-critical-big-data-applications.html

Manzoni, J. (2017). Big data in government: the challenges and opportunities. [online] Gov.uk. Available at: https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities

Marr, B. (2015). Big Data: Using SMART Big Data; Analytics and Metrics To Make Better Decisions and Improve Performance. 1st ed. John Wiley & Sons Ltd.

Marr, B. (2017). Forbes Welcome. [online] Forbes.com. Available at: https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/#7582ec666c10

Marr, B. (2017). How is Big Data Used in Practice? 10 Use Cases Everyone Must Read. [online] Bernard Marr. Available at: https://www.bernardmarr.com/default.asp?contentID=1076

Marr, B. (2016). Top 10 AI And Machine Learning Use Cases Everyone Should Know About. [online] Forbes.com. Available at: https://www.forbes.com/sites/bernardmarr/2016/09/30/what-are-the-top-10-use-cases-for-machine-learning-and-ai/#253feb8394c9

Marvin, R. (2016). Predictive Analytics, Big Data, and How to Make Them Work for You. [online] PCMAG. Available at: https://www.pcmag.com/article/345858/predictive-analytics-big-data-and-how-to-make-them-work-fo

Maydon, T. (2017). The 4 Types of Data Analytics. [online] Kdnuggets.com. Available at: http://www.kdnuggets.com/2017/07/4-types-data-analytics.html

McAfee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution. [online] Harvard Business Review. Available at: https://hbr.org/2012/10/big-data-the-management-revolution

McDonald, C. (2017). 5 Big Data Trends in Healthcare for 2017 | MapR. [online] Mapr.com. Available at: https://mapr.com/blog/5-big-data-trends-healthcare-2017/

McGinley, B. (2017). Big Data Risks for Small Businesses. 1st ed. Country Financial.

Mediratta, A. (2015). Big data: Terms, Definitions and Applications. EMC.

MIT Sloan Management Review. (2016). Lessons from Becoming a Data-Driven Organization. [online] Available at: http://sloanreview.mit.edu/case-study/lessons-from-becoming-a-data-driven-organization/

Mohanty, S., Jagadeesh, M. and Srivatsa, H. (2013). Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics. Berkeley, CA: Apress.

Monnappa, A. (2017). How Facebook is Using Big Data - The Good, the Bad, and the Ugly. [online] Simplilearn.com. Available at: https://www.simplilearn.com/how-facebook-is-using-big-data-article

Murray, S. (2016). Data analytics is on trend with fashion houses. [online] Ft.com. Available at: https://www.ft.com/content/621d20c0-7033-11e6-a0c9-1365ce54b926?mhq5j=e5

Murthy, D. and Bowman, S. (2014). Big Data solutions on a small scale: Evaluating accessible high-performance computing for social research. Big Data & Society, 1(2)

NewVantage Partners (2017). Big Data Executive Survey 2017: Executive Summary of Findings. NewVantage Partners LLC.

O'Brien, A. (2012). The Impact of Big Data on Government. Iron Mountain.

O'Dowd, S. (2017). Top 10 Big Data Trends in 2017 for Financial Services | MapR. [online] Mapr.com. Available at: https://mapr.com/blog/top-10-big-data-trends-2017-financial-services/

Ohlhorst, F. (2013). Big data analytics. Hoboken, New jersey: John Wiley & Sons, Inc.

Ovide, S. (2011). Harry & David Bankruptcy: Everything You Need to Know. [online] WSJ. Available at: https://blogs.wsj.com/deals/2011/03/28/harry-david-bankruptcy-everything-you-need-to-know/

Parise, S., Iyer, B. and Vesset, D. (2012). Four Strategies to Capture and Create Value from Big Data. Ivey Business Journal, (July/August 2012).

Patterson, L. (2015). How Hadoop Keeps Even Small Businesses In the Loop for Big Data Analytics. [online] Business2Community. Available at: http://www.business2community.com/big-data/hadoop-keeps-even-small-businesses-loop-big-data-analytics-01257559#8PyCJ45KD38jIiDs.97

Pepalis, B. (2015). What’s the Difference Between Big Data & Business Intelligence?. [online] Izenda.com. Available at: https://www.izenda.com/blog/whats-different-between-big-data-business-intelligence/

Petty, A. (2016). 6 Big Challenges Managers and Organizations Face With Data. [online] The Balance. Available at: https://www.thebalance.com/challenges-managers-and-organizations-face-with-data-4100645

Platt, J., Souza, R., Checa, E. and Chabaldas, R. (2014). Seven Ways to Profit from Big Data as a Business. [online] https://www.bcg.com. Available at: https://www.bcg.com/publications/2014/technology-digital-seven-ways-profit-big-data-business.aspx

Pierson, F. (2017). Top 3 server storage companies to keep your data safe at a bargain price. [online] Big Data Made Simple - One source. Many perspectives. Available at: http://bigdata-madesimple.com/top-3-server-storage-companies-to-keep-your-data-at-a-bargain-price/

Press, G. (2013). Forbes Welcome. [online] Forbes.com. Available at: https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/

Ram, J., Zhang, C. and Koronios, A. (2016). The Implications of Big Data Analytics on Business Intelligence: A Qualitative Study in China. Procedia Computer Science, 87, pp.221-226.

Puget, J. (2015). Optimization Is Ready For Big Data: Part 4, Veracity (IT Best Kept Secret Is Optimization). [online] Ibm.com. Available at: https://www.ibm.com/developerworks/community/blogs/jfp/entry/optimization_is_ready_for_big_data_part_4_veracity?lang=en

Quinn, B. (2017). Big Data for Small Businesses? Not so Fast | CustomerThink. [online] Customerthink.com. Available at: http://customerthink.com/big-data-for-small-businesses-not-so-fast/

Ram, J., Zhang, C. and Koronios, A. (2016). The Implications of Big Data Analytics on Business Intelligence: A Qualitative Study in China. Procedia Computer Science, 87, pp.221-226.

Rathinasamy, S. (2015). Unstructured Data: Growth and Challenge. EMC.

Reisenwitz, C. (2017). How Small Businesses Can Increase Customer Loyalty Using Data Science. [online] Blog.capterra.com. Available at: https://blog.capterra.com/how-small-businesses-can-increase-customer-loyalty-using-data-science/

Reyes, J. (2015). The skinny on big data in education: Learning analytics simplified. TechTrends, 59(2), pp.75-80.

Rouse, M. (2017). What is data visualization? - Definition from WhatIs.com. [online] SearchBusinessAnalytics. Available at: http://searchbusinessanalytics.techtarget.com/definition/data-visualization

Rouse, M. (2017). What is self-service business intelligence (BI)? - Definition from WhatIs.com. [online] SearchBusinessAnalytics. Available at: http://searchbusinessanalytics.techtarget.com/definition/self-service-business-intelligence-BI

Rouse, M., Martinek, L. and Stedman, C. (2017). What is big data analytics? - Definition from WhatIs.com. [online] SearchBusinessAnalytics. Available at: http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics

Ruppel, D. (2017). Four Ways Governments Gain With Big Data Analytics. [online] Microstrategy.com. Available at: https://www.microstrategy.com/us/blog/stories/four-ways-governments-gain-with-big-data-analytics

Samuels, M. (2017). Big data case study: How UPS is using analytics to improve performance | ZDNet. [online] ZDNet. Available at: http://www.zdnet.com/article/big-data-case-study-how-ups-is-using-analytics-to-improve-performance/

SAS. (2017). Analytics bears fruit: Harry & David. [online] Available at: https://www.sas.com/en_us/customers/harry-and-david.html .

SAS Institute Inc. (2017). Data Visualization Techniques: From Basics to Big Data With SAS Visual Analytics. 1st ed. SAS Institute Inc.

Schroeck, M., Shockley, R., Smart, J. and Romero-Morales, D. (2012). Analytics: The real-world use of Big Data. How innovative enterprises extract value from uncertain data. IBM Institute for Business Value.

Search Technologies. (2017). A Data Lake Architecture with Hadoop and Open Source Search Engines. [online] Available at: https://www.searchtechnologies.com/blog/search-data-lake-with-big-data

Seltman, H. (2015). Experimental Design and Analysis.

Sherman, R. (2017). Understanding BI analytics tools and their benefits. [online] SearchBusinessAnalytics. Available at: http://searchbusinessanalytics.techtarget.com/feature/Understanding-BI-analytics-tools-and-their-benefits

Siemens, G. (2011). Learning and Academic Analytics | Learning and Knowledge Analytics. [online] Learninganalytics.net. Available at: http://www.learninganalytics.net/?p=131 [Accessed 18 Sep. 2017].

Simon, H. (1954). Spurious Correlation: A Causal Interpretation*. Journal of the American Statistical Association, 49(267), pp.467-479.

Stitch Fix Newsroom. (2017). Stitch Fix Enjoys Growth & Client Loyalty - Stitch Fix Newsroom. [online] Available at: https://newsroom.stitchfix.com/press-releases/stitch-fix-enjoys-growth-client-loyalty/

Stubbs, E. (2014). Big data, Big Innovation: Enabling Competitive Differentiation through Business Analytics. 1st ed. Wiley and SAS Business Series.

Sverdlik, Y. (2017). Google Ramped Up Data Center Spend in 2016. [online] Data Center Knowledge. Available at: http://www.datacenterknowledge.com/archives/2017/02/01/google-ramped-data-center-spend-2016

Tableau Software. (2017). Audience Audit completes more work in half the time. [online] Available at: https://www.tableau.com/solutions/customer/audience-audit-completes-double-work-half-time

Tanner, J. (2014). Analytics and Dynamic Customer Strategy: Big Profits from Big Data. New York, NY: John Wiley & Sons.

TDWI Research (2013). Managing Big Data. TDWI Best Practices Reports Series. Renton.

Troester, M. (2012). Big Data Meets Big Data Analytics. SAS Institute.

Twiddy & Company. (2017). About Twiddy & Company Outer Banks Vacation Rental Specialists. [online] Available at: https://www.twiddy.com/about/

Virmani, A. (2017). How Big Data is Transforming Retail Industry. [online] How Big Data is Transforming Retail Industry. Available at: https://www.simplilearn.com/big-data-transforming-retail-industry-article

Wicks, D. (2016). A Quick Guide to Big Data Tools for Small Businesses. [online] Aurora Beacon-News. Available at: http://www.chicagotribune.com/suburbs/aurora-beacon-news/community/chi-ugc-article-a-quick-guide-to-big-data-tools-for-small-bus-2016-11-07-story.html

Woodie, A. (2016). 9 Ways Retailers Are Using Big Data and Hadoop. [online] Datanami. Available at: https://www.datanami.com/2016/07/20/9-ways-retailers-using-big-data-hadoop/

Woodie, A. (2016). How Retailers Use Big Data to Gobble Up Sales. [online] Datanami. Available at: https://www.datanami.com/2016/11/23/retailers-use-big-data-gobble-sales/

Zhou, L., Pan, S., Wang, J. and Vasilakos, A. (2016). Machine learning on big data: Opportunities and challenges. Neurocomputing, 237, pp.350-361