Bachelor Thesis in Business Administration

DUSSOULIER Hannah
&
ISHAC Patrick

The impact of Big Data on companies and a lack of skills as the origin of the challenges they are facing

An investigation aimed to understand the origin of the challenges companies are facing with Big Data

Dussoulier Hannah - 950402 - T149
Ishac Patrick - 950629 - T153

Jönköping - May 2018
Abstract

The 21st century saw the rise of internet and with it, the digitalization of our world. Today, many companies rely on technology to run their businesses and Big Data is one of the latest phenomenon that arose from technological evolution. As the amount of data is constantly increasing, ranging from business intelligence to personal information, Big Data has become a major source of competitive advantage for companies who are able to implement it efficiently. However, as with every new technology, challenges and issues arise. What’s more, the learning curve is steep, and companies need to adapt quickly, so as to follow the pace of innovation and develop the skill-set of their employees to remain competitive in their respective industries. This paper investigates how Big Data is impacting companies, the main challenges they are facing within its implementation and looks to determine if these challenges originate from a lack of skills from the current workforce. A qualitative study has been conducted, interviewing nine respondents over eight interviews of 54 minutes on average.

Three main ideas have been outlined through the interviews conducted by the authors. The first is the impact of Big Data in companies with mainly the benefits, challenges, regulations as well as the cohabitation of human beings and technology. The second and third are the optimal profile of a decision-maker and the ideal profile of the employee in companies working with Big Data. The profiles of the decision-maker and employee are composed of characteristics, skills and experience. The decision-maker, in this paper, was defined as a key actor in the success or failure of a company and of great influence on the profile of the employee. His skills, such as strategic, basic, analytical, communication and decision-making were developed, and their correlation was demonstrated. Ultimately, the lack of skills in companies today, often regarded as a challenge by numerous scholars, was shown to be the origin for many of the challenges companies are facing, mainly through bad decision-making and lack of communication.

The authors finally outlined steps for a successful implementation of Big Data in companies and future trends such as regulations and increased technological evolution to carefully and actively pursue for people and businesses alike.
Acknowledgements

We would like to express our gratitude to Rudrajeet Pal, our thesis tutor for providing unwavering guidance and sharing precious time and expertise. We also extend our appreciation to Anders Melander, our course examiner as well as all associate teachers and lecturers: Adele Berndt, Charlotta Mellander, Naveed Akhter, Lucia Naldi, Kristofer Månsson and Mikaela Backman for their seminars and lectures throughout the semester.

A special credit goes to all of our interviewees who kindly agreed to discuss this topic with us and without whom we would never have been able to write this paper: Gilles Yvetot, Frederique Stumm, Mark Eid, Simon Werner Zankl, Gilbert Salhani, Antonio Boulos, Yvette Madi, Mario Montag and Delphine Genin. Thank you again for your time, support and passion.

We would also like to thank Peter Trummel, Marina Bau and Kristin Klinge for their considerable help and support.

Above all, we are immensely grateful and forever thankful to our families who supported us in every possible fashion during our studies and the writing process of this paper.
Table of content

Part 1: Introductory Chapter

1.1 Background ................................................................. 6
1.2 The problem ............................................................. 7
1.3 The purpose ............................................................. 7

Part 2: The state of the literature ..................................................................................... 9

2.1 The definition of Big Data ......................................................................................... 9
2.2 The opportunities of Big Data .................................................................................. 11
2.3 Big Data Challenges ............................................................................................... 14
2.4 The challenge “lack of skills” ................................................................................. 17
2.5 Definitions .............................................................................................................. 20

Part 3: Methodology ........................................................................................................ 22

3.1 Research Design ...................................................................................................... 22
   3.1.1 Philosophical perspective .................................................................................. 22
   3.1.2 Research Approach ......................................................................................... 23
   3.1.3 Research Strategy ........................................................................................... 23
3.2 Methodology ........................................................................................................... 24
3.3 Method .................................................................................................................... 24
   3.3.1 Sampling Method ............................................................................................ 24
   3.3.2 Interview Design ............................................................................................ 27
3.4 Data Collection ........................................................................................................ 29
3.5 Data Analysis .......................................................................................................... 29
3.6 Description of Analysis Strategy ............................................................................ 30
   3.6.1 Analysis Procedure ......................................................................................... 30
3.6.1.3 Identified categories ........................................................................................................... 31
3.6.2 Creation of sub-categories ........................................................................................................ 31

Part 4: Ethics and Trustworthiness .................................................................................................. 33
4.1 Ethics ........................................................................................................................................... 33
4.2 Trustworthiness & quality of research ......................................................................................... 34
  4.2.1 Credibility .............................................................................................................................. 34
  4.2.2 Transferability ......................................................................................................................... 35
  4.2.3 Dependability ......................................................................................................................... 35
  4.2.4 Confirmability ......................................................................................................................... 35

Part 5: Empirical Findings ................................................................................................................ 37
5.1 Impact of Big Data ...................................................................................................................... 37
  5.2 Profile of the decision-maker ...................................................................................................... 39
  5.3 Profile of the employee .............................................................................................................. 41

Part 6: Analysis and Interpretation of Findings ............................................................................. 43
6.1 Impact of Big Data ...................................................................................................................... 43
  6.1.1 Humans and Technology ........................................................................................................ 43
  6.1.2 Data Safety ............................................................................................................................ 45
  6.2 Profile of the decision-maker and employee ............................................................................. 46
    6.2.1 Characteristics and Experience ............................................................................................ 47
  6.2.2 Skills ...................................................................................................................................... 48

Part 7: Discussion ............................................................................................................................. 53
7.1 Findings correlated to the literature .......................................................................................... 53
  7.2 Findings adding new insights to the literature .......................................................................... 55

Part 8: Conclusion ............................................................................................................................ 57

Part 9: Managerial implications ...................................................................................................... 59
  9.1 Further managerial implications related to our research question ........................................... 59
  9.2 Further implications not directly related to our research question ......................................... 61
Part 1: Introductory Chapter

1.1 Background

The 21st century has seen the rise and fast development of new technologies through companies such as Alphabet, Microsoft, Apple and Amazon. Technology has completely blended into our lives and became ubiquitous through our smartphones, laptops and all the Internet of Things (IoT) devices i.e. connected devices available in the market.

In the last decade, we have transitioned to what is widely called today the digital era, where data has become the new raw material (Berners-Lee & Shadbolt, 2011) and digital oil (Yi, Liu, F., Liu, J., & Jin, 2014). As data is generated through every interaction by any individual, be it through an online action or a physical contact, the amount of information generated and collected every second is constantly rising, ever so quickly.

During the last few years, due to the large amount of data available and the rapid increase in data creation, the phenomenon has been generically called Big Data. It mainly emerged from the combination of widespread internet access, growing number of consumer products and services, powerful computing power and large data warehouses. This pool of data is now used in both the public realm and the business world (Cukier & Mayer-Schöneberger, 2014). In the business world in particular, Big Data is regarded as a tool that provides companies with more information, faster than ever before. Big Data is at the core of many firms today as they build their businesses around the phenomenon and is also used as strategy guidance and influences their decision-making activity.

Big Data is however, quite a recent phenomenon as scholars widely started investing it after 2010 (Gandomi & Haider, 2014). The development rate of the technologies Big Data is based on, coupled with the newly developed interest from scholars and companies alike, translates into a broadly unknown field. However, as tech-companies through innovation, and not scholars through researches, are moving the phenomenon forward, the business community has not been able to efficiently adapt. This ultimately resulted in numerous challenges faced by companies when implementing Big Data partly due to a mere lack of time.
As with every innovation and technological progress, there is a learning curve and due to the fast growth of Big Data, companies had little time to develop the skill-set needed to proficiently use this new tool. Now that the business environment has radically shifted to become digital and data-enabled, it requires deep understanding and adaptation from companies and people.

1.2 The problem
Based on the reviewed literature about Big Data, we identified that many companies face various barriers in its usage and understanding. The difficulty for many firms is to know how to select only the valuable data and how to use all this information effectively. Big Data can bring many new opportunities to companies, but this is only possible if the firm has the necessary resources to process all the available data (Marr, 2016).

Indeed, many researches exists about the barriers of Big Data. However, no research has been focused on one particular challenge in order to help companies overcoming it. Existing literature within that field highlights that more investigation on the various issues that Big Data brings to organizations should be done (Chen, & Zhang 2014) and that considerable research efforts are needed to improve the efficiency of data usage and analysis (Chen, Mao, Zhang, & Leun, 2014). Also, some authors ask for deeper investigation on how organizations should restructure themselves in order to embrace Big Data and provide precise recommendations for new strategies implementations (Mauro, Greco, Grimaldi & Ritala, 2017).

The problem in this research paper is studied from a management’s perspective where Big Data is used as a strategic tool influencing decision making.

1.3 The purpose
Big Data is a relatively new phenomenon which impacts almost every company nowadays (Slessor, 2017) and even if literature exists within that field, many questions are still not answered and Big Data remains a vague concept for many people.

With this research paper, we aim to study Big Data and its impact on businesses as well as discover the challenges these companies face, how they arose and where they originate from. We have chosen to investigate one specific issue which is the lack of skills within companies as the potential origin for the challenges.

In this regard, our research should be relevant for companies who seek to understand why and how a potential lack of skills will affect their business performance and how to address. This
research should also be of interest to business students who will most probably work in a data-enabled company in the future.

Existing literature defines lack of skills as the challenges companies face, related to the lack of knowledge and understanding, that translates into an inefficient use of Big Data (Marr, 2016). In this sense, lack of skills represents the challenges that employees face by not understanding Big Data because of their limited knowledge and due to potential lack of training in that field. The term lack of skills is also related to management and leadership not able to align knowledge transmission between the different departments of the company (McAfee & Brynjolfsson, 2012). The lack of technical abilities results in a bad understanding of data that directly impacts a company’s performance. The impact is on the one hand, errors in decision making and on the other hand, not using Big Data as an opportunity for growth.

In this sense, our research questions are:

*How is the implementation of Big Data impacting companies today?*

*What challenges arise from the implementation of Big Data and do they originate from a lack of skills?*
Part 2: The state of the literature

This part of the paper presents an analysis of the stream of research on Big Data, the opportunities and main barriers that companies face, based on 55 articles published on internet as well as in various books, between 2001 and 2018. The used articles come exclusively from scientific and academic entities.

2.1 The definition of Big Data

During the last decade, the term “Big Data” has started to emerge rapidly within the academic literature (Mauro, Greco & Grimaldi, 2016). However, scholars are struggling to agree on a formal definition of Big Data, emphasizing the fragmented discourse on what constitutes it and what metrics define it (Sivarajah, kamal, Irani & Weerakkody 2017). The main reason behind this academic condition is the fast adoption of Big Data by public and private sectors which left little to no time for academics to develop and settle on a formal definition (Gandomi & Haider, 2014). This is partly due to Moore’s Law that states that the number of transistors doubles approximately every 2 years, implying that the computing power doubles as well, thus attesting that technology is constantly moving forward. As of 2018, the majority of scholars still argue on an “official” description of this phenomenon.

However, this same majority shares a common ground through Laney’s definition in 2001. Laney (2001) has famously outlined Big Data’s 3Vs: Volume, Variety and Velocity. Volume stands for the size of the data, Variety refers to its structure and Velocity is the rate at which it is produced. This 3Vs framework is shared by Chen, Chiang and Storey as well as Kwon, Lee and Shin (2012) but also Gandomi and Haider (2014) to only cite a few.

First, Volume fundamentally defines the physical space required or the hard drive capacity necessary to store the data. The Volume of data is a very interesting characteristic because what is considered a large amount of data today, is most probably going to be of insignificant size in the future considering the speed at which data is created. Regarding Velocity, scholars are unanimous, data production is going to increase exponentially in the future. According to a study by Gantz and Reinsel (2012), between 2005 to 2020, the
digital space will have grown by a factor of 300 reaching 40 Zettabytes. In 2010, we were at 1,3 Zettabytes. In 2017, we were at approximately 2,5 Zettabytes.

Some of the reasons for this massive increase are the proliferation of Internet of Things (IoT) devices, the increased availability of internet in third-world countries and the ever-increasing data generated by each individual due to more devices and more reliance on internet-based services (Gantz & Reinsel, 2012; “Data never sleeps 5.0”, 2017; “Extracting business value from the 4Vs of Big Data”, 2016).

Finally, Variety is also becoming a more challenging characteristic. Research has outlined three types of data: structured, semi-structured and unstructured. Structured, which represents 5% of all existing data (Cukier, 2010), is the data available in spreadsheets for example. Texts, images, videos and audio are some examples of unstructured data, the latter represents more than 90% of the data today. Semi-structured data is a cross between these two as it may have some structure that can be used for analysis but lacks the strict arrangement found in database or spreadsheets (Marr, 2016).

Building upon these 3Vs, Dijcks (2013) adds a fourth V: Value, he states data can be used to create valuable insights that may lead to an increased competitive advantage for companies. Schroeck, Shockley, Smart, Romero-Morales and Tufano. (2012) believe Veracity is another characteristic as some sources within the Big Data Supply Chain are inherently unreliable. Tech-companies such as IBM, Oracle and SAS also tend to share these two new characteristics (“SAS - What is Big Data”, 2018; “Extracting Business Value from the 4Vs of Big Data”, 2016).

Besides, Suthaharan (2014) has his own framework with the 3Cs: Cardinality, Complexity and Continuity. SAS does share the Complexity point of view but ads Value as well to its own definition (“SAS - What is Big Data”, 2018). IBM developed their own framework labeled the 5Vs that include Laney’s 3Vs in addition to Veracity and Value (“Extracting Business Value from the 4Vs of Big Data”, 2016). Finally, Sivarajah et al. (2017) argue there is a need for a 6th V: Visualization, as they claim owning the data and analyzing it properly is challenging but once this has been done, visualizing it in a timely and effective manner is one of its most important aspects. Today, most companies thrive to visualize data in real-time, thus making visualization a critical characteristic and a very though-after skill from recruiters (Sivarajah et al., 2017).
Other scholars and companies have focused their research and definitions of Big Data based purely on its technological aspect. Microsoft Research (2013) defined Big Data as “the process of applying serious computing power, the latest in machine learning and artificial intelligence, to seriously massive and often highly complex sets of information.” Chen, Chiang and Storey (2012) defined it as “the data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis and visualization technologies” (p. 1166). Dumbill (2013) described it as “data that exceeds the processing capacity of conventional database systems” (p. 1). We can thus see in this very large array of definitions that scholars still have not agreed upon a common definition and that their views diverge quite substantially.

Two researches bring another dimension: social impact. Boyd and Crawford (2012) state that Big Data is a “cultural, technological and scholarly phenomenon that rests on the interplay of technology, analysis and mythology” (p. 662) while Mayer-Schönberger and Cukier (2013) define it as a “phenomenon that brings three key shifts in the way we analyze information that transform how we understand and organize the society: more data, messier data and correlation overtakes causality”.

Ultimately, Mauro et al. (2016) summarized in their paper all the definitions of Big Data and grouped them into four categories: Information, Technology, Methods and Impacts. They then based their own definition of Big Data on these four categories and described it as “the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value” (p. 131). We believe the latter to be the most complete and relevant definition of Big Data to-date and the most in-line with our paper.

When understood and used efficiently, Big Data can be regarded as a major source of competitive advantage for companies as it brings many opportunities.

### 2.2 The opportunities of Big Data

Scholars are unanimous, Big Data is regarded as today’s new *Digital Oil* (Yi, Liu, F., Liu, J., & Jin, 2014) and *New Raw Material* (Berners-Lee & Shadbolt, 2011). In addition to being compared to gold and oil, Big Data is seen as one of the most powerful sources of competitive advantage for companies of the 21st century. It is repeatedly associated with enormous growth
opportunities as large tech-companies, such as IBM or Oracle, believe high-value information could be extracted from Big Data given the right tools, processes and skills. Scholars also agree with that assessment (Gandomi & Haider, 2014) and even state that it “changes the way people think about knowledge and research altogether” (Boyd & Crawford, 2012)(p. 665). Lazer, Pentland, Adamic, Aral, Barabási, Brewer, … and Van Alstyne (2009) affirmed that it offers “the capacity to collect and analyze data with an unprecedented breadth and scale” (p.722). Today, Big Data is enabling researchers to analyze, sort and visualize large amounts of data faster and better than ever before. This translates into less tedious tasks for scholars and gives them more time to analyze the data at hand or enables a faster publication of articles for example.

There is a common discourse amongst academics that Big Data will affect every industry or sector in the business world today. From computer scientists to economists, mathematicians, sociologists, physicians to doctors, all will have potential benefits in embracing Big Data, granted that they have the resources to use it. The benefits are endless. Janssen and Kuk (2016) wrote that it has “the potential to fundamentally transform the way organizations make their decisions” (p. 9).

In an Accenture study published in 2014, 89% of respondents rated Big Data as “very important” or “extremely important” to their businesses’ digital transformation, and 82% stated Big Data is a major value-creating source for their companies. According to the research, 94% of executives said Big Data is providing additional ways to identify new sources of revenue, 90% stated it allows their company to acquire more customers (“Companies Are Satisfied with Business Outcomes from Big Data and Recognize Big Data as Very Important to Their Digital Transformation”, 2014). A more recent study by Accenture in 2017 has outlined that 89% of users across various industries believe that a successful Big Data implementation will revolutionize the way companies operate. What’s more, they trust the impact on a company’s operations and processes will be as profound as the one internet had almost three decades ago (Slessor, 2017).

As data is generated throughout every internet-based activity today, every sensor and IoT device (Columbus, 2016), the amount of data produced will soon surpass the computing power we currently have to analyze it in a timely and cost-efficient manner. Thus, efficiency is a key characteristic when determining if a company is investing in a value-creating endeavor. This
Data Deluge (Sivarajah et al., 2017), if handled properly, will enable us to expose new knowledge, make faster and better decisions and give innovative and actionable insights to businesses (Jukić N., Sharma, Nestorov & Jukić B., 2015).

A UPS Case Study, done in 2013, is one of the best examples of an innovative and value-adding endeavor through the use of Big Data. UPS equipped more than 46 000 delivery vehicles with telematics sensors (Samuels, 2017). In 2013, the company launched its new software called ORION (On-Road Integrated Optimization and Navigation) aimed at redesigning the entire UPS’s drivers route structure and navigation systems. By combining the data from the sensors onboard the vehicle (such as speed, drivetrain performance and braking) with external data (such as traffic management, online mapping and weather conditions), ORION optimized automatically and in real-time every driver’s delivery route. UPS claims the project has led to savings of more than 31 million liters of fuel in 2011 by cutting more than 136 million kilometers from daily routes. The company also estimates that by reducing each drivers’ daily route by only 1.6 kilometers, it can save approximately 25 million euros on a daily basis (Davenport & Dyché, 2013).

According to Chen and Zhang (2014), Big Data will provide rich business intelligence for better-informed business decisions and Brown Chui and Manyika (2011) believe Big Data will create opportunities through value creation. Kumar, Niu and Re (2013) state it will support in enhancing visibility and flexibility of supply chain and resource allocation.

On the one hand, the Dublin City Council has implemented in 2013 a big data analytics solution to better manage the traffic flow of its public transportation services. The solution was based upon the deployment of GPS-systems on all the buses coupled with live cameras providing the ability to gather information on the 1000+ buses in real-time. Controllers can display each bus on a virtual city map, locate areas of congestions and instantly activate the live camera feeds to assess the cause of the delay. The system also automatically generates live arrival times and adjust transit times accordingly. The implementation of this system has allowed the City Council to manage more efficiently its bus fleet through real-time tracking, faster identification of issues and thus faster response-time and reduced emissions. What’s more, the City Council can offer peace-of-mind to the riders as it can provide real-time and accurate arrival times and/or explanations as to what caused a delay ("Big Data Helps Dublin Improve Public Bus Transport Network", 2013).
On the other hand, Big Data provides innovative ways to leverage business assets for companies. Some of the real-world applications comprise investment firms embracing social media data. In 2012, Cha indicated that more than 50% of Wall Street investment firms incorporated social media data such as tweets or product reviews on Amazon in their trading activities. It only contributed to less than 2% in 2007.

Another example is through face recognition technology and sensors, stores can track in-store movement patterns and cross-match them to purchasing behavior providing very valuable insights into consumer buying behavior (Kudyba, 2014). This information is leveraged within business decisions such as product placement, promotions, advertising and staffing (Gandomi & Haider, 2014).

To conclude, companies need to include Big Data in their core strategy in order to benefit from it. However, implementing Big Data also brings along many challenges that companies need to face and overcome to ensure success.

2.3 Big Data Challenges

Today, most companies sit on a wealth of data; the difficulty is to know how to effectively analyze it. The latter is referred to as “Big Data Analytics”.

According to Rajaraman (2016), data analytics can be defined as the extraction of knowledge from Big Data and consist of a correlation between different variables. Four types of data analytics can be highlighted.

The first one is descriptive analytics which explains the current state of a business situation to make developments, patterns and exceptions evident, by producing standard reports for instance (Joseph & Johnson, 2013). Secondly, predictive analytics is used to analyze available data in order to explain what is likely to happen in the future. Thirdly, discovery analytics consists of collecting unexpected relationships among variables, as well as collecting data from various sources and analyzing them to provide new information. Finally, optimizing solutions to different problems by using gathered data, is called prescriptive analysis. This analysis can help a company to achieve a certain goal by improving its decision-making process and trying to reduce business expenses. Consequently, data itself is a low-value asset: the only way to extract value from Big Data is to have an efficient data analytics process in order to turn it into
meaningful insights and enable faster and/or better decision-making (Gandomi & Haider, 2014).

Kaisler, Armour, Espinosa and Money (2013), explain in their article that the volume of Big Data has vastly increased in the past years since data is created by everyone at any time. This makes the data management process more difficult for structured and unstructured data. Thus, companies need to look for and master new techniques to cope with that increase by developing new skills and nurture employees’ knowledge.

Sivarajah et al. (2017) tackle the subject of challenges for Big Companies to store, analyze, visualize the data in a cost-efficient manner and emphasize the fact that costs are going to be a substantial problem for companies in the future because of the sheer size of the data produced, collected and the need for its analysis. Scholars group Big Data challenges into three main categories.

The first category is related to Data Challenges, which are the basic characteristics of the data itself. They agree with Kaisler et al. (2013), that the high volume of data represents a great challenge for companies. Its variety in forms and sources, such as messages, transactional data (business transactions for instance) and web data (images for example), makes it difficult to comprehend and manage. Moreover, its Veracity, as some sources are inherently unreliable, Velocity, as more and more data is created every second, and Variety as the amount of unstructured data ever-increasing, represent further challenges for companies. This is also discussed by Gandomi and Haider (2014), who explain that Veracity of data imposes companies to connect, match and clean the data from different sources.

Sivarajah et al. (2017), also present another “V”, which is Visualization, stating that analyzing the data is great, but showing quick and easy ways to visualize it so that the information that needs to be transmitted is easy to understand, is another big challenge that leads to wrong decision making. Many factors can influence this challenge such as communication skills, in which the difficulty is to transmit your understanding of Big Data to someone else in the company.

The second category represents Process Challenges, which correspond to capturing, integrating, transforming and selecting the right analysis model and results (Kaisler et al.,
However, in order to perform these tasks effectively, one needs to have a broad, though intricate, knowledge of the processes. In other words, acquiring data from diverse sources and storing it for value generation purposes, represents an issue for many companies since warehousing data implies increasing costs. These costs are expected to boom within the next two years as the amount of data produced will grow exponentially. It seems that many companies face inadequate infrastructure and inappropriate data warehouse architecture. This is also outlined by Alharthi et al. (2017), who explain that many companies do not have a clear strategy to address the various sources and formats of Big Data and some firms are even deleting old, but potentially very important, data because of warehousing infrastructure overload.

In this sense, Alharthi et al. (2017) and Russom (2011) add that with this new phenomenon, companies need to innovate their processes and change their working procedures. Building effective Big Data Analytics systems and database software able to process all the information quickly, seems to be a great challenge for many companies. Thus, a deep understanding of Big Data Analytics and of the company’s core business model is required to be able to implement the necessary changes and to make the best use of the technology at hand. In other words, organizations encounter a lack of infrastructure readiness, where investments in software and hardware are too low, and where most of the utilized technologies are not designed to process Big Data effectively. Organizations also encounter difficulties when extracting and cleaning the data collected from a large scale of unstructured data and need to integrate the structured data within the business processes.

In this regard, infrastructure readiness can be linked to technological challenges that companies face. New platforms must be built to store high volumes of data and companies need to adapt their technology tools to embrace the volume, velocity and variety of data (McAfee & Brynjolfsson, 2012). Chen, Mao, Zhang and Leung (2014) develop this aspect by outlining the “Data Life Cycle Management issues” which arise from this Data Deluge. Sivarajah et al. (2017) add to this idea, that analyzing and modelling the data once it has been captured, stored, mined, cleansed and integrated will imply new costs for companies since firms need more computing power and more warehousing capabilities. Gandomi & Haider (2014) in their paper, also outline the lack of current computing power and that companies must react accordingly to the further increase of Big Data production within the next few years.

Finally, the third identified category by Sivarajah et al. (2017) is related to Management Challenges that cover security, ethical and privacy issues for instance. The people creating the
data are not the ones in control of it. Significant challenges in terms of truth, control and power appear (Boyd & Crawford, 2012). These challenges are also highlighted by Alharthi et al. (2017). The use of personal data can be critical and private information can be revealed easily (Chen et al., 2014). Companies need to guarantee a secured data access and a constant protection of users’ data, in order to avoid ethical challenges where loyal customers can stop trusting a company for instance (Corea, 2016).

Alharthi et al. (2017) add another challenge: Cultural Barriers. Organizational culture has a strong impact on processes, values and norms and where the overall understanding on how Big Data can improve business operations is very low and Big Data resistance exists.

Finally, some authors develop a different challenge that companies encounter. Janssen, van der Voort and Wahyudi (2017) state how a lack of communication and collaboration between the decision-makers and the people collecting the data, can lead to worse outcomes. Data collection, processing and usage are not done within a single department. Instead, data is collected by many parties and organizations might use collaborations and partnerships for acquiring the resources and capabilities for its analysis. All these actors and steps might increase the difficulty of using Big Data for decision-making.

Overall, the quality of data is not only dependent on the data itself, but also on the different ways in which it is collected and processed. Hence, the value that one can extract from the data is solely dependent on that person’s abilities and skills. The various challenges that have been outlined by the existing literature so far, seem to originate from a common issue that companies are facing nowadays. The latter appears to be lack of skills within companies.

### 2.4 The challenge “lack of skills”

With the rise of Big Data, companies seek for new skills. They try to increase their current analytical workforce with data scientists who possess a high level of IT skills and who can manipulate Big Data technologies. Many organizations realize that traditional quantitative analysts are not enough anymore (Davenport & Dyché, 2013). Since companies not only possess structured data, the difficulty is to have people able to understand and use unstructured data. (Marr, 2016). In this sense, a main challenge for companies is to find employees able to use data and turn it into valuable information. Davenport and Dyché (2013), develop this idea by explaining that companies search for people having the ability to explain Big Data outcomes
to executives. The difficulty of this task is to understand the data perfectly and to explain it in an understandable way. This ability is critically important for valuable decision making. By using the term “lack of skills”, existing literature considers the fact that companies today, struggle to find employees able to use Big Data in a sense that will allow the company to gain a competitive advantage, and who are also able to transform data analysis into actions. One key problem is that many firms struggle to explain what skills, knowledge and abilities they are looking for within job postings.

According to Stanton A.D and Stanton W.W (2016), two broad categories of skills exist that companies require.

The first skill category is related to technical competencies, those encompass software development, programming and statistical knowledge, quantitative analysis and the ability of using various statistical and analytical tools. The second category considers personal abilities. These skills are related to communication skills, collaboration and team skills, the ability to provide insights of data in a way that other team members understand it, creativity and the curiosity to solve complex problems. This is related to another issue that various articles highlight when analyzing the lack of skills within Big Data. As previously stated, the lack of communication between the different departments represents a problem for many firms. Organizations struggle, for instance, to implement close relationships between the departments addressing Big Data and the IT department who is supporting them (Davenport & Dyché, 2013).

This statement is developed by Pacino (2017), who explains that analyzing Big Data seems only to be related to the executive level within a company and that a lack in analytical training, communication and transmission of information and knowledge exists. This impacts the knowledge of many employees since knowledge between the people who analyze the data and the decision-makers is not aligned. In this regard, decision-makers often lack the necessary skills and do not understand the variables and hence make bad decisions.

Three categories of lack of skills emerged from the reviewed literature.

The first category concerns the lack of employees’ skills. Employees with little to no big data understanding, might represent a real burden to companies. As previously stated, employees either lack in technical or/and personal skills. The lack of employees’ skills can imply errors where information is not correctly placed and where important information can be lost (Alharthi et al., 2017). This idea is also supported by Russom (2011), who explains that inadequate staffing or skills for Big Data Analytics exists, and that this makes Big Data unusable for end
users. This again relates to the Veracity of the data seen previously as being one of the characteristics of it but also one of its key challenges.

The second category, which is linked to the lack of employees’ skills, is related to Managerial and Leadership abilities. In this sense, White (2011) analyzed in his article that using data effectively is not only a technical problem but also emerges from management issues. The author states managerial know-how is necessary to handle the high volume of data. Efficient leadership is required in which clear goals need to be set. Communication, general management techniques and knowledge is requested. This is particularly essential to face the communication issue that exists between the different departments and to ensure everyone in the company understands data and knows how to use it. Talent Management also needs to be taken into consideration, where expertise within Big Data must be developed to work with a large quantity of information.

In this sense, managers need to review the decision making to hire the right people with the required knowledge and understanding. If leaders do not adapt their leadership strategies to embrace this new phenomenon, they will probably lose competitive advantage. It seems clear that companies that can combine data science and domain expertise will increase their competitive advantage and surpass their rivals. That implies that companies must change their hiring patterns by focusing on data-driven individuals that are able to bring valuable decision-making information.

However, domain expertise people are often expensive, and employers tend to outsource these activities on a short-term basis. McAfee and Brynjolfsson, (2012), Mauro et al. (2017) and Corea (2016), all agree that companies need to rethink their Human Resources needs and strategies. It seems clear that a skill-set is needed to work within Big Data and competitive companies will be those who have the ability to hire the right people.

Data analytics’ jobs description seem often very vague and the required skills and responsibilities are not clearly outlined. The difficulty is now for managers to think what kind of new people they should hire and how they can increase the skills of their current human resources. Moreover, Corea (2016) insists on the fact that the job market is not yet ready for Big Data since many employers seem to not really know what exactly they are looking for when they seek to hire what are called “Data Scientists”. The author defines Data Scientists as people
with strong expertise and the willingness to perfectly understand numbers and able to provide insights based on them. He also explains that they have five main job roles which are the following: computer scientist, businessman, statistician, communication manager and domain expert. That means that these persons will request more time and resources and will cost more because of their expertise and flexibility.

Also, according to Chen, & Zhang (2014), data storage, acquisition, sharing, analysis and visualization seem to be great challenges since it requires highly skilled human resources. Training Big Data Analysts that must have mathematical abilities and professional knowledge about Big Data, seems to take many years. Without Data Analysts, companies do not have the capabilities to explore the existing data and have only limited techniques and tools to solve the different problems.

Lastly, the third identified category by the existing literature is related to technical skills, which is an overlap of the lack of employees’ and management skills. In this sense, Bird (2013) and Mauro et al. (2017) agree that organizations need to implement new analytical methods in order to transform big quantities of information into business value.

On the one hand, this requires “hard skills” where specific data and analytical knowledge is needed. On the other hand, companies need to acquire “soft skills”, where Big Data must be transformed into value creation. It seems important for companies to hire people with both hard and soft skills. Pacino (2017) adds that no stable ground exists within that field and that no improvement in skills appeared within the past few years. The implication of that lack of technical skills seems to be that the demand for people with strong Big Data analytical skills exceeds the supply.

2.5 Definitions

This part summarizes the three main definitions and concepts we have outlined earlier in our paper, and which will be of use for the reader. We have chosen to insert the definitions of the following concepts after having explained them in the literature review to ensure they are perfectly understood by the reader.

- **Big Data**: Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value (Mauro et al., 2016).
• *Big Data Analytics*: Big Data Analytics can be defined as the extraction of knowledge from Big Data and consist of a correlation between different variables (Rajaraman, 2016).

• *Lack of Skills*: We define lack of skills as the challenges companies face, related to the lack of knowledge and understanding, that translates into an inefficient use of Big Data (Marr, 2016).
Part 3: Methodology

The goal of doing research is to produce evidence and results. To do so, research uses scientific methods and follows specific rules, in order to avoid that findings are subjective and depend on the personal researchers’ views (Byrne, 2017).

On the one hand, we will explore the benefits and challenges of Big Data and use models to find a clear definition of it. Existing models are for example, the 3Vs, Velocity, Volume and Variety (Laney, 2001), 4Vs with the addition of Veracity (Corea, 2016) and 5Vs (Grobelnik, 2012) with Value. Another theory affecting the development rate of Big Data is Moore’s law as it enables companies to process information faster than ever before.

On the other hand, to gain practical knowledge about that topic we have conducted a qualitative analysis by interviewing people who work in big data-enabled companies. We want to find out how companies face this new phenomenon and how it impacts their current business strategies.

3.1 Research Design

3.1.1 Philosophical perspective

In order to conduct research, it is important to understand philosophical issues. Understanding the philosophy of business research allows to analyze the strengths and weaknesses of each, and how the philosophical assumptions may influence the judgement of the overall research quality (Easterby-Smith, Thorpe & Jackson, 2013).

The research philosophy can for instance, clarify the research designs and recognize which designs will work and which will not.

The main philosophical debates about how to conduct a research, concern ontology and epistemology (Easterby-Smith et al., 2013). On the one hand, the book defines ontology as the philosophical assumptions about the nature of reality and four different ontological views exist. We have chosen the concept of Relativism within our research, since we believe that related to our research question, many truths exist and that facts depend on the viewpoint and perception of the observer. For example, the potential lack of skills within Big Data that might exist in various companies, is perceived differently in each firm and from each manager.

On the other hand, epistemology can be defined as the way that we address the reality. Within epistemology, we have decided to adopt the concept of Constructionism. Constructionism accepts value from various data sources and allows us to make some generalizations beyond
the present sample for instance. In this sense, we believe that both, Relativism and Constructionism fit together since their starting point is to question a phenomenon and that collected data mainly consists of words. Furthermore, the analysis and interpretation of data is done through triangulation and comparisons as our statements are only based on the different perspectives and opinions we gathered. There may be many different realities, so as researchers, we had to gather the various perspectives and collect the views and experiences of diverse companies. Lastly, the searched outcome of these two concepts is that the research should generate theory (Easterby-Smith et al., 2013).

The objective of our study is to explore what happens in practice that could add something new to the theory. In this sense, we have decided to conduct an explorative study.

3.1.2 Research Approach

Concerning our research approach, we have decided to use an inductive research approach. According to Ruse (2015), with inductive theory approach, data is interpreted in order to generate findings and is compared afterwards and modified through additional data collection and analysis. Inductive approach is in this sense, a comparative analysis. Through the literature review, we have found a gap concerning the challenges that are related to the lack of skills, that various enterprises face when implementing Big Data in their business strategy. Our objective was to explore this gap and gather, understand and analyze data in order to compare it to the literature review. This allowed us to see if the gap is filled and where further research is needed. We wanted to avoid being influenced by the theory when we conducted interviews since our goal was to explore a new phenomenon.

3.1.3 Research Strategy

The objective of our research strategy is to adopt a plan allowing us to answer our research question.

Since our purpose is to conduct an explorative study, we aim to get opinions and perspectives to understand a common behavior from a relatively small sample, rather than gathering numerical data from a large sample. In this regard, we have decided to conduct a qualitative study through in-depth interviews, which we will analyze by reading and interpreting them as narratives.
Qualitative interviews can be defined as directed conversations based on questions and answers about a specific topic (Easterby-Smith et al., 2013).

We believe that these interviews represent an opportunity for mutual discovery, understanding and reflection on our topic. We are not only interested in understanding the respondent’s perspective and viewpoint, but also in the reason why, he or she, has this specific viewpoint. Relating the research strategy to our research question, our objective is to understand the opinions and perspectives of the various companies we have interviewed, about why a lack of skills within the use of Big Data exists in companies nowadays, and how they are handling this challenge.

In order to successfully gather qualitative data, we aimed to obtain trust from our interviewee, especially when we have never met each other before. We both believe that only if the respondent is trusting us, he/she will provide us with the requested information. Moreover, we both paid attention to adopt an appropriate attitude and language, to appear knowledgeable and sensitive.

### 3.2 Methodology

We believe that the exploratory study is the most relevant methodology for our research as we are looking to explore a new phenomenon and not verify an existing theory. We are conducting a qualitative explorative study by interviewing people with questions that are not based on our literature review. Thus, we are not looking to test an existing hypothesis or theory but rather try to build our own perspective through the exploration of this new phenomenon. In other words, we want to construct knowledge rather than simply reveal it.

### 3.3 Method

#### 3.3.1 Sampling Method

We chose to use non-probability sampling designs and more specifically purposive sampling for the selection of our interviewees.

As we have a clear idea of what sample units are needed according to the purpose of our study, we believe purposive sampling is the most appropriate for our research. The latter involves identifying and selecting individuals that are very knowledgeable about or experienced in a certain phenomenon (Cresswell & Plano Clark, 2011). Besides knowledge and experience, availability and willingness to participate was also an important aspect of our sampling method.
What’s more, it ensured these people were willing to share experiences and opinions in an articulate and reflective manner (Bernard, 2017 and Spradley, 1979).
Thus, we have identified and approached specific sample units according to our eligibility criteria. Within the purposive sampling, we chose to opt for criterion sampling as we wanted to ensure the interviewees were pertinent for our research and would provide relevant data to analyze (Easterby-Smith et al., 2013).

We defined five eligibility criteria. The first three criteria were that the respondent is working in a company at the moment of the interview to ensure he is exposed to this phenomenon, this company is using Big Data and is facing challenges within its use. The second set of criteria was more individual-related. The interviewee had to be knowledgeable about Big Data, which we defined as having worked with Big Data for more than five years, since only knowledgeable respondents will be able to provide pertinent information and have an executive position in the company he/she was working in. This guarantees all respondents have a strategic position, allowing them to have a global vision of the company’s strategy. We define executive position as President, Vice President, CEO, Managing Director, Head of Department or Department Director. (see criteria table below)
### Criteria Table

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Interview 1</th>
<th>Interview 2</th>
<th>Interview 3</th>
<th>Interview 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Company</strong></td>
<td>Stargazer</td>
<td>Techni-Contact</td>
<td>Crimtan</td>
<td>TrustCruit</td>
</tr>
<tr>
<td><strong>Usage of Big Data</strong></td>
<td>Use data to pre-select potential partners for clients and predict future trends in the influencer marketing sector. Automation of the entire collection process. Make better and faster decisions in the future.</td>
<td>Profile clients and constantly update it after every touch point and interaction. Increase sales through better understanding of clients' needs. Follow-up on prospects through online retargeting. Deeper understanding of website navigation patterns.</td>
<td>Profile customers, understand their behavior by following their online journey. Huge insight into customer need and expectations. Ability to create models, assign characteristics and target people according to many factors such as location, weather, income, age, interests.</td>
<td>Collect huge amount of feedback from candidates to help companies measure and improve their recruitment performance and processes. Standardized and efficient way to collect huge amount of information and display it efficiently.</td>
</tr>
<tr>
<td><strong>Challenges with Big Data</strong></td>
<td>Veracity of the data as some unstructured data is very hard to process. Velocity of the industry as a whole. Processes and systems are complicated to develop. Lots of data to analyze. Hard to assess the value that can be extracted by different sources of data.</td>
<td>Brand safety is very important and challenging as ads need to be displayed in a safe online environment. Lots of data, much of it is not used by Crimtan's clients.</td>
<td>Clients do not understand why they need Big Data in their recruitment process. Lots of data to analyze, some of it is not used or relevant anymore.</td>
<td></td>
</tr>
<tr>
<td><strong>Interviewee Experience</strong></td>
<td>7 years as a data engineer</td>
<td></td>
<td>25 years as CEO in the retail industry 5 years in Big Data</td>
<td>28 years in the advertising industry 10 years in Big Data</td>
</tr>
<tr>
<td><strong>Interviewee Position</strong></td>
<td>Chief Technology Officer</td>
<td>Founder &amp; Chief Executive Officer</td>
<td>Managing Director</td>
<td>Chief Executive Officer</td>
</tr>
</tbody>
</table>

*Figure 1: Criteria Table (part 1)*

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Interview 5</th>
<th>Interview 6</th>
<th>Interview 7</th>
<th>Interview 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Company</strong></td>
<td>IQVIA</td>
<td>BPN</td>
<td>Predikto</td>
<td>Sodexo</td>
</tr>
<tr>
<td><strong>Usage of Big Data</strong></td>
<td>Provide and analyze Big Data to laboratories to determine the best action plan to increase sales. Reports are made for customers to provide the best plan of action: collecting feedback (data) from doctors on a new product.</td>
<td>Get data about target audience Understand customer behavior online. Better targeting, better budget allocation, better transparency with clients Ability to predict outcome of a campaign.</td>
<td>Development of softwares to predict potential failures in the production and transportation industries. Ability to analyze thousands of data points instantly and very efficiently.</td>
<td>Create synergies between departments. Get to know better our clients &amp; Ensure deeper relationships with them. Entire marketing and sales strategy based upon Big Data.</td>
</tr>
<tr>
<td><strong>Challenges with Big Data</strong></td>
<td>Volume of the data Recovery of the data Analysis is very complex Visualization is hard to execute properly Communication flow is sometimes challenging people changing jobs. Education and communication processes took place on all fronts Digital is evolving fast Different companies are managing the digital universe: need to adapt and be proactive.</td>
<td>Massive amount of data to transport, store, analyze. Costly and very time-consuming process. Analyzing data requires extremely complex calculations. Customers do not understand the value Predikto is bringing to the market.</td>
<td>Lack of knowledge from employees -&gt; skepticism Lack of understanding Difficulty in technology adoption Hard to use all of the systems and processes. Time consuming process Not easy to assess the ROI.</td>
<td></td>
</tr>
<tr>
<td><strong>Interviewee Experience</strong></td>
<td>20 years of experience 10 years in Big Data</td>
<td>Interviewee 1: 9 years in Big Data Interviewee 2: 8 years in Big Data</td>
<td>18 years of experience 11 years in Big Data</td>
<td>20+ years of experience 5 years in Big Data</td>
</tr>
<tr>
<td><strong>Interviewee Position</strong></td>
<td>Project Director CRM</td>
<td>Interviewee 1: Vice President Interviewee 2: Head of Digital</td>
<td>Founder &amp; Chief Executive Officer</td>
<td>Head of Strategic Partnerships and CRM</td>
</tr>
</tbody>
</table>

*Figure 1: Criteria Table (part 2)*

26
3.3.2 Interview Design

As stated earlier, we aim to do qualitative interviews with a purposive criterion-based sampling method. We opted for qualitative as it allows us to have in-depth discussions and interviews with our sample units. For the purpose of our study, we aimed at the creation of primary data through the interview process. As we are doing an exploratory study, primary data is essential to our research. However, we also gathered some secondary data through existing information that is not produced by us. Company websites and LinkedIn profiles are examples of secondary data we collected.

Even if we made sure our interviewees are knowledgeable about Big Data, we wanted to be as clear and comprehensible as possible in our choice of questions and terminology used. We avoided any jargon and slang words in the questions as well as in our official communication with all interviewees. Not only does this make it simpler for the readers in the future, but it also facilities our interviewees’ understanding of the questions and made it simpler for him/her to answer in a relevant manner. We also sought to avoid leading questions or biased questions through the use of objective words and vocabulary as this could have also influenced the person’s responses.

For the development of our interview questions, we decided to conduct a pilot interview. On the one hand, this allowed us to validate our questionnaire and assess potential responses. It helped us to calibrate and reformulate some of the questions to ensure they were comprehensible and would not lead our interviewees to a certain direction. On the other hand, it enabled us to confirm the relevancy of our study since the interviewee shared the idea that Big Data challenges are an important issue in the business world today and their origin is still quite vague.

We have developed a semi-structured interview (see Appendix 3) with questions that promote open-ended answers for two main reasons. First, we have developed an interview schedule with some questions we thoroughly wanted the interviewee to answer as those would be the basis for our analysis. We have however formulated some follow-up questions based on some potential answers while always giving ourselves the opportunity to ask further questions on a particular topic of interest that the interviewee might have addressed but that we did not expect. Secondly, we want the person to communicate freely and reflect upon our topic throughout the
entire interview and semi-structured seemed to be the most flexible way to reiterate questions when needed while having a clear strategy (Easterby-Smith et al., 2013).

We followed the funnel approach to design our interview process and questions. This approach is often recommended when conducting semi-structured interviews as it begins with a broad view of the topic and then proceeds to narrow down the conversation to very specific components of the topic (Spradley, 1979).

We have developed three sections in our interviews: opening questions, key topic questions and closing questions. We aimed at having the opening questions to serve as “icebreakers”, build trust as well as set a comfortable and stress-free environment. It was also the opportunity to further inquire about the company in which the person was working and understand better the use of Big Data. The second section introduced our main topic: lack of skills. This section was split into two parts: the first being a broad question about the challenges the company was facing, and the second part dived deeper into the potential lack of skills the company was experiencing and what it was doing to cope with it.

The third section contained two questions that served as reflection questions allowing the person to talk more broadly about the subject again.

Throughout the entire discussion, we have both used the probes technique as it enabled us to confirm a saying or information the interviewee gave us earlier in the interview and ensure its validity. We also used the laddering technique to further dig into a particular phenomenon or response. By laddering up, a researcher will help the respondent move from statement of facts to start revealing the individual’s value base. We did the latter by asking “why” questions. By laddering down, a researcher seeks to obtain examples and thus gather more detailed description of a particular events. This helps explore a person’s understanding of a particular construct. We also applied the laddering down strategy when asking how the company uses Big Data but also by asking for an example of a potential lack of skills the interviewee depicted earlier in the interview.

Most of our interviews were done through Skype as we could not meet the interviewee face to face because of location-related factors. We are aware that having a discussion over Skype is less personal and establishing trust is harder, but we made sure to be as polite, accommodating and sensitive as possible. By establishing an early contact and communicating with the person
multiple times before the interview, we had already established a certain degree of trust (Easterby-Smith et al., 2013).

Remote interviews have however a multitude of advantages. First of all, convenience for both parties. We did not want the person who accepted to discuss this topic with us to invest anything more than time and Skype interviews are the most convenient process we could think of. Secondly, it allowed us to interview people in Dubai, Paris, Atlanta and San Francisco while being in Sweden which is a substantial advantage. Thirdly, it provided us with the ability to be more thorough in our eligibility criteria as we were not restricted by location or any economic investment required for travel.

We have created an Interview Design table (see Appendix 1) with all the relevant information including company-related (such as location and scope of work), person-related (such as name and nationality) and interview-related intelligence (such as length and communication method used).

3.4 Data Collection

We have conducted eight interviews in total with nine people. The interviews were done between the 21st of March 2018 and the 19th of April 2018. The length ranged from 32 minutes to 71 minutes with an average of 54 minutes per interview (see Appendix 1).

Seven interviews were done through Skype and one was done face to face. We performed three interviews in French as it was more convenient for the interviewees and five in English. As stated earlier, all interviews were recorded, with the approval of all parties involved, and transcribed shortly afterwards.

3.5 Data Analysis

After the collection of primary data through remote or face to face discussions, we applied a content analysis to our transcripts and notes. It is an approach that aims at drawing systematic inferences from qualitative data that has been structured by a set of ideas or concepts (Easterby-Smith et al., 2013).

The first step is usually to define a number of criteria and select relevant material for the analysis. The aim being to outline factors or concepts to answer the research question. Once these factors are identified, a matrix can be created and used to start drawing out connections
and easily identifiable content. This approach also gives us a great degree of flexibility as we can assign certain levels of importance per question answers. We aimed at using the indexing technique by looking for the number of occurrences of certain words or phrases in the data we gathered. This helps bringing a common phenomenon to light (Easterby-Smith et al., 2013). As we are exploring a field that has not been explored before, we also have a certain degree of flexibility as we can shape the nature of our analysis and then start drawing connections back to the frameworks we discussed in the literature review. This also stands with our more exploratory inductive approach on an overall basis.

3.6 Description of Analysis Strategy

3.6.1 Analysis Procedure

3.6.1.1 Transcripts
In order to analyze the data we gathered through our interviews, we first thoroughly transcribed each interview we conducted and reviewed the content we had written down. The transcripts made sure we did not forget important aspects discussed during the interviews and ensured we were as objective as possible with our overall analysis process. What’s more, it made our entire analysis easier. Both of us took the time to carefully read the transcripts on our own in order to get familiar with the content. Next, we brainstormed about what we have understood by reading the various transcripts and then reflected about it on our own.

3.6.1.2 Coding
We started the content analysis through the coding technique. To do so, we went through the transcripts together and started assigning codes for any content that seemed relevant to us and would help us answer our research question. After having done that, we discussed the founded codes and started to brainstorm about our coding techniques and potential improvements.

We allowed ourselves one day before finalizing the coding. Then, we discussed potential new codes, added the ones that seemed relevant and deleted the ones that were actually not necessary for our research.

Once we finalized the coding, we were ready to start the categories identification.
3.6.1.3 Identified categories

The next step was to write down all selected codes on a whiteboard in order to have a clearer overview allowing us to figure out potential connections and differences between them. This enabled us to group the similar codes into one category which became our sub-categories. Afterwards, we consolidated our sub-categories and made sure each of them was relevant. We also wanted to assess if there was a need for any additional ones. Finally, we created our categories and main categories. The latter emerged from identifying implications and common points between all of the sub-categories.

3.6.2 Creation of sub-categories

In this part, we will discuss our coding strategy and detail how we proceeded. To do so, we chose to illustrate our strategy through the table below in which we highlighted four examples that explain how we coded the content of the interviews and how we came up with sub-categories.

We will take you through our thought process and how we built the sub-category “predict the future” out of the responses from the interviewees. As you can clearly see in the samples of codes used column, “predict things”, “the outcome is more predictable” and “predict the future” all share the same idea which is “predict the future”. This common idea is the genesis of our sub-category “predict the future”.

We have applied the same strategy to build all of our sub-categories. We believe sub-categories are essential to our work as they are the first step to building categories and main categories enabling us to go deeper into the analysis and interpretation.
### Examples of coding strategy and resulting sub-category

<table>
<thead>
<tr>
<th>Respondents</th>
<th>Sentences from interview</th>
<th>Samples of codes used</th>
<th>Outlined sub-category</th>
</tr>
</thead>
<tbody>
<tr>
<td>A7</td>
<td>&quot;we predict things that are going to happen one day to 90 days in the future&quot;</td>
<td>predict things</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>&quot;The outcome, for some of the campaigns, to a certain degree, is more predictable than before&quot;</td>
<td>the outcome is more predictable</td>
<td>Predict the future</td>
</tr>
<tr>
<td>A1</td>
<td>&quot;you gather a lot of data with AI in order to predict the future&quot;</td>
<td>predict the future</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>&quot;This tool gives you the number of times this word is being used on the search bar, in real-time&quot;</td>
<td>gives you the number in real-time</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>&quot;The thing is, we also have real-time data on what our competitors are doing&quot;</td>
<td>have real-time data</td>
<td>Faster access to data</td>
</tr>
<tr>
<td>A1</td>
<td>&quot;This Application Programming Interface works constantly and updates our database continuously&quot;</td>
<td>works constantly and updates our database continuously</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>&quot;you have enough material to do conscious and well informed decisions&quot;</td>
<td>conscious and well informed decisions</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>&quot;Think of data as a tool that gives you information to make smart decisions&quot;</td>
<td>make smart decisions</td>
<td>Better decision-making</td>
</tr>
<tr>
<td>A2</td>
<td>&quot;this means that making a decision today has become much more reliable&quot;</td>
<td>making a decision has become much more reliable</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>&quot;So obviously, having analytical skills is very important&quot;</td>
<td>having analytical skills is important</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>&quot;It's great to have data, but then you need to know how to analyze it&quot;</td>
<td>know how to analyze it</td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>&quot;You need to hire or put the right people with the right skills such as mathematical skills for data scientists and all the people who are knowledgeable about data analytics&quot;</td>
<td>put the people knowledgeable about data analytics</td>
<td>Analytical skills</td>
</tr>
</tbody>
</table>
Part 4: Ethics and Trustworthiness

4.1 Ethics

In this part we focus on the ethical issues that can arise within our research. It seems clear that our thesis will not undertake studies which could harm peoples’ lives, still we want to avoid any other kind of personal or economic harm.

First of all, we want that potential respondents are fully told about the purposes of the research, what we try to find out through eventually interviewing them, their role within the research, the required time for conducting an interview, as well as potential benefits and risks. All our respondents must be given the opportunity to agree to taking part in research. To do so, we have created a consent form (see Appendix 2) which was sent to all interviewees a few days before the interview. Reading the consent form gives potential respondents the possibility to withdraw their participation at any time. In this regard, an interview only took place if the respondent has sent us back the signed document.

It seems important to inform respondents about potential risks, more specifically about the nature of risks in terms of severity and their likelihood in terms of occurrence (Byrne, 2017). In our case, we have not identified any risks for respondents when taking part in our research, apart from risks related to the confidentiality and anonymity of communicated information. To assess these risks, a whole paragraph is dedicated to confidentiality, where we want to guarantee that all communicated information that the respondent does not want to be published, will be kept confidential and anonymous. If wished so, we also guarantee that the participants cannot be identified from the information provided and send a contractual commitment that the information will be anonymized (Byrne, 2017). In this sense, once the interview was finished, we always sent an email to the participant, to confirm if we can use her/his name, or if she/he prefers to stay anonymous. In addition, for analysis purposes, we wanted to record all interviews. However, we always asked for the respondents’ agreement and no interview was recorded without the respondents’ willingness. In other words, we wanted the respondent to be aware of how we are conducting our research and the different steps we took.
Overall, we paid attention to ensure that no harm comes to participants, by not publishing any information that is likely to harm the interests of the individuals, and by protecting their privacy. Moreover, we aimed to be honest and communicated in fully transparency with them (Byrne, 2017).

4.2 Trustworthiness & quality of research

It seems clear that if we want our research to be relevant, credible and interesting to others, it must be useful to the readers (Easterby-Smith et al., 2013). The research quality is mainly determined by the concept of trustworthiness (Shenton, 2004). According to Lewis-Beck, Bryman and Liao (2004), the trustworthiness of a research is based on four criteria, namely credibility, transferability, dependability, confirmability. These criteria could be defined as the standards allowing to judge qualitative inquiries, the rigorousness of the findings, methodological concerns and the overall measure of the adequacy of study. In this sense, it is important to meet these four criteria to justify the work is worthy of attention. Hence, to describe and evaluate the quality of our research methods and explain the reason why we have employed these methods, we use the concept of trustworthiness in our study.

4.2.1 Credibility

Credibility refers to the idea that readers and research participants should be able to understand why certain models and methods have been used, and why participants have been selected. In other words, this criterion should prove that our work is believable and that statements are accurate so that our research design makes sense for readers and participants (Given, 2008). To do so, we aimed to apply the most appropriate and coherent research method, allowing us to explore our topic in depth. Moreover, we have put much emphasis on selecting suitable participants that are relevant for our study and who are allowing us to obtain truthful responses. Therefore, we have only conducted interviews with people being in an executive position, working daily with Big Data and with a deep knowledge within that field. Because most companies are in a transition phase, we believe that interviewing people in an executive position will give us a strategic insight into the company’s vision and business strategy.
4.2.2 Transferability
Transferability takes two elements into account: the relevance of our respondents, as well as the ability to provide a complete understanding of the context being studied and ensuring that the research question is appropriately answered. In other words, it allows to prove that our study is useful (Given, 2008). In this regard, it seems important to provide a comprehensible description of the method used. Therefore, we paid attention to provide sufficient details about the data collection and method of analysis. In addition, we will provide background information about the respondents, justifying our selection choices. We believe that with that done, the reader can determine whether our method is applicable for another context.

4.2.3 Dependability
Dependability refers to the traceability and logic of the research process. It considers that researchers must be conscious of changes in the environment which can imply changes in methodology. It also refers to the reproducibility of the study for other researches (Given, 2008). Since qualitative research is conducted in changing environments, replicability is affected (Shenton, 2004). In this regard, we must be aware that the interviewee’s answers might change in the future. Especially, because in our work, we aim to get opinions and perceptions about a phenomenon. Therefore, we provided a detailed description of our method, data collection and analysis. In other words, in order to repeat a study, Given (2008) states that environmental- and context changes must be considered by researchers. We believe we provided sufficient information about our research design so that the reproducibility is possible.

4.2.4 Confirmability
Confirmability allows to certify that the collected data in the research can be verified all the way back to the original data sources, such as records or transcripts for instance (Given, 2008). We did not want to influence or lead the interviewee towards a specific direction during the conducted interviews. To do so, we always started the interview with broad questions to introduce our subject and adopted a semi-structured interview design, in order to be as flexible and open as possible. Furthermore, we wanted to avoid stressing the respondents while answering our questions, by giving them the time to reflect before asking follow-up questions. In addition to that, we have recorded every interview to facilitate the potential verification of
the collected data. Overall, we believe that our processes prove that the findings are only determined by the participants’ answers and were not modified from our side.
Part 5: Empirical Findings

By analyzing the interview transcripts, we have identified sub-categories, categories and main categories, which will be essential for the interpretation of our findings. The aim of this part of our study is to introduce and discuss the identified main categories.

As mentioned previously, we followed the inductive research approach. Thus, we believe it’s important to note that we first identified the various sub-categories, then the categories and lastly, this allowed us to come up with three main categories.

5.1 Impact of Big Data

The first main category that has emerged from the interviews is the Impact of Big Data on businesses. This main category is related to the overall changes that Big Data brings to companies and how its development affects everyone today. We outlined five categories from the transcripts.

Figure 3: Table presenting the main category “Impact of Big Data”

The first identified category is the Human Aspect that transpires from the overall evolution of technology. It encompasses the topic “Human beings and technology” that addresses the cohabitation of human beings and technology (often mentioned as robots and Artificial
Intelligence) as well as “Reluctance to change”. “They have a strong reluctance to change, they all have their old ways and habits and they wish to keep them” (A6).

The second category we defined as **Benefits** contains “Financial”, as Big Data was said to generate greater sales and revenues, and “Structural” advantages, as it brings more precision and transparency, alongside a “Faster access to data” and a larger “Volume of data”. “There are many benefits: optimize, generate return on investment, help us make smart decisions, give us access to insights that we never had before without them” (A8).

Furthermore, the **Importance of Big Data** is a theme that was always discussed along with the “Technological evolution” leading to it. Many respondents defined Big Data as being the future of business “For us, it is very clear: [Big Data is] the future of business” (A2), and at the core of many companies, including theirs.

Moreover, “Technological challenges”, such as a lack of time or understanding, and a “Forced adaptation” to Big Data generated the **Challenges** category. All respondents outlined the fact that technology is evolving very fast and that companies need to adapt quickly and cannot work without Big Data anymore. “[the employees] don’t really have a choice, they have to cope with it otherwise, we will lag behind technologically” (A6).

Finally, **Safety** is composed of “Laws and Regulations” such as the General Data Protection Regulation (GDPR) and “Privacy issues”. Many interviewees mentioned this new law that will come to effect in the end of May. The GDPR aims to regulate the way companies use and store data in order to increase data safety. Respondents highlighted the impact of this new law on many businesses since databases and the way data is handled will have to be reviewed. “It is a new regulation that will most likely revamp almost every website in Europe as everyone will be forced to rearrange and remap their databases” (A2).
5.2 Profile of the decision-maker

Secondly, this main category introduces the optimal profile of the decision-maker working in a data-enabled company. The data that we gathered led us to create three categories that each encompasses a plethora of sub-categories.

The first category, namely Characteristics comprehends firstly, the ability of the decision-maker to be proactive regarding the implementation of technology in his business, as well as the ability of being visionary. These characteristics allowed us to create the sub-category “Foresight”. Secondly, according to many respondents, the decision-maker should be passionate “you need to be passionate about data” (A3), and thirdly, he should be a “lateral thinker” (having common sense and logic for example). “I think it requires lateral thinkers as lateral thinkers can manipulate, analyze and assess data and dig deeper” (A6).

Additional respondents’ ideas were that failure is essential and inevitable but the ability to learn from failure will lead to further success. In this regard, the emerged category is the need for Experience, which is, according to many respondents, a key element within a company. “I mainly think it’s the experience which is important.” (A1).
The third identified category, **Skills**, emerged from five sub-categories:

- **Decision-making skills**: the ability of the decision-maker to choose whether to train (who and how?), to outsource certain or all Big Data Analytics activities or perform them in-house, but also the ability to develop a hiring strategy if needed. “*we don't train everybody in this company*” *(A6)*.

- **Communication skills**: the ability to communicate the Big Data Analytics results in an understandable way, being the bridge between the various departments in order to ensure a strong diffusion of knowledge as well as being an active listener. “*However, in our strategy, the IT department doesn’t know the data we need, doesn’t know our clients or their needs as it doesn’t communicate with them*” *(A8)*.

- **Analytical skills**: the ability to only select the relevant data and analyze it accordingly in order to extract the most value out of it. “*we will be able to extract beautiful and very useful data*” *(A2)*.

- **Basic skills**: In this sub-category, respondents outlined the importance for the decision-maker to have basic technical skills, synthesizing skills and a deep and thorough understanding of his business. “*You need the basic technical knowledge and capabilities*” *(A7)*.

- **Strategic skills**: this includes the capability of the decision-maker to have a clear strategy in mind about how to work with Big Data, but also being the “bridge” between the technological and business needs of the company. Some respondents highlighted that only a combination of engineering and business skills will lead to future success and prosperity of the company. “*I think the people who will be the most successful, are the ones who are able to understand the technology, but also understand the business requirements and build the bridge between what is possible technically and what is needed from a business.*” *(A7)*.

In addition, the executive level should be able to recognize potential lack of skills in order to overcome them.
5.3 Profile of the employee

After having defined the optimal profile of the decision-maker, another main category emerged from the interviews: the ideal profile of an employee working with Big Data.

The same categories as the profile of the decision-maker compose this main category: **Characteristics**, **Skills** and **Experience**. However, they are not as extensive as in the previous main category.

**Characteristics** contains a “Commitment” towards the company, a “Willingness to be trained” combined with “Passion” and “Attitude” as all of our respondents stated that only someone passionate and willing to be trained would thrive and adapt to technological change. “*we only train the ones who are willing to be trained. Some people are not interested so there is no point wasting time and resources training them, so we focus our efforts on the people who show interest*” (*A6*).
The **Skills** required for people working with Big Data are less extensive than the decision-makers’, an “Understanding of the business” and “Basic technical and analytical skills” were mentioned by our respondents.

Finally, **Experience** was not stated as being an essential feature or topic, however many respondents clearly stated that with Big Data, “Learning by doing” is a key factor. “*As long as you stay in that company, you always grow, gain experience, so understand the business needs in order for it to grow, to make a positive contribution towards the company.*” (A4).
Part 6: Analysis and Interpretation of Findings

Within this part, we will analyze the content of the interviews by digging deeper into our interviewees’ responses in order to interpret and identify causalities among and in-between them. Firstly, we will attempt to comprehend how Big Data impacts companies today and what may come in the future. Secondly, we will determine what skills are required to accommodate the implementation of this new technology within businesses and how people working with Big Data developed their skill-set around this phenomenon.

6.1 Impact of Big Data

First of all, as discussed and detailed earlier in the paper, Big Data has and is still impacting many firms no matter the industry, size, scope, location or age of the companies. This technology is omnipresent in the business realm today and companies that wish to sustain or establish a competitive advantage over their competitors need to embrace the latter to its full potential. Today, the importance of Big Data is obvious for all of our respondents. A8 stated that “Everyone is using [Big Data]. Everyone is conscious that this is the future” and A3 defined it as being “the future of business. And I believe every digital business is affected. It is really crystal clear.”

Throughout this part, we chose to investigate two categories within the main category Impact of Big Data: the human aspect with the challenges that arise from it and safety and regulations.

6.1.1 Humans and Technology

One of the main challenges Big Data is bringing to companies is the speed at which the technology is evolving. As it is still during its nascent years, Big Data is evolving at an ever-increasing rate, almost faster than Moore’s law according to A4. This directly impacts companies and employees that need to adapt quickly to the newest technologies, tools and processes should they wish to stay competitive.

In the 21st century, a common belief and fear amongst people from a plethora of industries is that Artificial Intelligence, robots or algorithms will replace them. This was confirmed by some of our respondents in our discussions “Many people are scared that they will be replaced by a robot or assisted by a robot for some tasks, these people, directly or indirectly, will think we are taking their jobs away” (A8) This creates an atmosphere of skepticism, fear and translates
into a reluctance to change in many companies “They have a strong reluctance to change, they all have their old ways and habits and they wish to keep them” (A2). People are scared not to understand their role in the company anymore, scared to be outsmarted by technology or do not wish to be pushed out of their comfort zone. Implementing Big Data in a company means new processes, new people, new skills, new platform and sometimes revamped business model. This was the case for A1 when he stated, “Big Data is orienting your business at some point”. What Stargazer is also doing is “responding to a large demand from a client by using technology. We work with technologies, not with humans” (A1). This directly corroborates why people are scared to lose their jobs to machines, more specifically in this case to machine learning and Artificial Intelligence.

However, it is critical to understand that technological evolution will happen nevertheless, as we have seen previously, with all our respondents claiming that, in our case, Big Data is the future. Change is inevitable in most industries and companies will be forced to adapt or they will “fail and fall” (A6). A2 adds “honestly, they don’t really have a choice, they have to cope with it otherwise, we will lag behind technologically. We have to, jobs are changing drastically”.

A significant aspect of this phenomenon is that all of our respondents still claim that there will always be a need for human interaction and input for technology to work efficiently. “Even though you’re using machine learning and AI, the human input is still very important in this” (A3). This will however not change people’s compulsory adaptation and development of new skills.

The generation Y, often referred to as Millennials, defines people that were born between the late 1980s and the late 1990s. We believe, because of its specific characteristics, this generation might be more keen or open to change. What’s more, Millennials tend to be more at ease with new technology as they are digital natives and are always looking to make an impact in the activities they are involved in (Sinek, 2011).

This generation might be able to cope more easily, but also more willingly to changes in the workplace, changes in jobs and changes in skills needed to work with new-technologies. One explanation is the fact that this generation grew up with the rise of internet and new-technologies, so it has mastered its use generally better than others. This also means Millennials
will have the instinct to look for information on search engines or develop a new skill through tutorials on video streaming websites or learn something by joining Massive Open Online Courses (MOOC) for instance. This generation can diminish the challenge of technological evolution and reluctance to change because it does not only develop its skills through experience but also through technology.

Finally, it is key to understand how technology such as Artificial Intelligence, machine learning and Big Data will affect the way we work but also how companies fit in the ecosystem. In order to ensure humans and machines cohabit in the workplace, we should embrace technology for what it can do better and faster while still putting the emphasis on the human control, creativity, talent and experience.

6.1.2 Data Safety

We will now focus this second part of the analysis on the safety and regulations surrounding data, both business intelligence and personal data. Firstly, all of our respondents mentioned at least once data safety as a challenge for their companies. In terms of both ensuring the safety of the warehouses and storage facilities, but also making sure the data is used ethically by companies and third-parties. “We need to evaluate how we share this data and how we use it. In my business, it is not dangerous but with more personal data such as healthcare, it can be very critical.” (A2). There is a large risk of abuse that could compromise the users’ privacy and confidentiality. This directly translates into an increased need for transparency between companies, their partners and users in order to reestablish trust in the entire value chain.

However, transparency does not seem to be sufficient as five of our respondents mentioned the need for regulations and more specifically the GDPR in our discussions. In May 2016, the General Data Protection Regulation was published in the EU Official Journal four years after the initial proposal to update the current data protection regulation by the European Commission which dates back to 1995. The GDPR will become fully enforceable throughout the European Union on the 25th May 2018. It was designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens data privacy and to reshape the way organizations across the region approach data privacy (“EU GDPR Information Portal”, 2018).
A2, A8, A4 and A6 agree on the fact that, regulations are needed. A2 states that “this is a good start and this law will benefit everyone in the EU”, A6 agrees and adds that “someone needed to regulate this industry” as “there is definitely a lot of abuse and misuse”. One scandal that supports the need for regulation and misuse is the Cambridge Analytica in March 2018 in which the data from up to 87 million Facebook users was shared improperly with Cambridge Analytica for inappropriate use ("Facebook scandal 'hit 87 million users'", 2018).

In contrast to the prompt technological change and forced adaptation in the business world, it seems that laws and regulations are deeply lagging behind. One might ask if the GDPR is not somehow already obsolete in some respects, considering all the factors we mentioned earlier in our study. Nonetheless, regulations on such an important topic as data safety is a key factor in reassuring people and securing information. As A8 mentioned, it is key to “hire people specialized in data warehousing, privacy and security to ensure [they] could face any potential hack”.

It is important to note that the GDPR legislations is only a European initiative. However, we believe in the next few years, the GDPR will be refined and many more laws will emerge all over the world to secure the ever-increasing quantity of data that is generated every day. We also believe that policy makers and lawyers might need to develop new expertise and knowledge regarding technological advancements as they are the individuals creating, shaping and applying these legislations.

6.2 Profile of the decision-maker and employee

Within this second part, we will analyze all findings related to the skill-set required to work with Big Data. It seems interesting to analyze the main category Profile of the decision-maker alongside with Profile of the employee and see how they continuously impact each other. Both main categories are clearly linked since the executive level of a company will have an impact on the employee’s skills. Employees can only be knowledgeable about the technological evolution and more specifically, about the use of Big Data, if the management is aware of this phenomenon and decides to implement various strategies to facilitate its understanding.
6.2.1 Characteristics and Experience

First of all, we can notice that skills are not the only criterion that is required to work with Big Data. Three of our respondents outlined the need for previous experience "I mainly think it’s the experience which is important" (A1), as well as the value of personal characteristics. On the one hand, experience within the same company allows to have a better understanding of the business and its needs. Two respondents talked about the importance of the people able to represent the “bridge” between the technological and the business needs of a company. New technologies will only be useful for a company if they solve a problem or improve the current structure. The difficulty here for the executive level is to identify where Big Data is needed and how to use it in an efficient manner. What matters is not to perfectly understand the technological side of Big Data but being able to fully comprehend how to use it effectively from a business-decision point of view, while maximizing the return on investment. Ultimately, the only criterion for a successful implementation of Big Data in a company will be the business performance that results from it.

What’s more, contrary to popular belief, attending classes about Big Data Analytics and data mining for instance, is not enough. Respondents agreed that practical experience is more important in ensuring someone’s success and improving someone’s performance. Big Data is continuously evolving and “learning by doing” seems to be the most efficient way to develop an expertise within that field. People need to practice, fail and learn from failure.

On the other hand, personal characteristics were said to be essential for both the decision-maker and the employee. For the decision-maker, these characteristics include the ability to be proactive and far-sighted as well as looking for what might come next in order to adapt quickly. Hence, it is clear that with the constant technological evolution, only a proactive person will ensure the prosperity of his company. Therefore, understanding the company’s needs and reacting to change is not enough in the business world today. One has to be “fast and know the global market” (A6) and where his company is positioned while always looking for future trends and innovations.

Employees need to show commitment and a willingness to be trained on Big Data processes and tools. Some respondents also used the word “passion” since some CEOs believe that only people passionate about Big Data and numbers will be able to work with Big Data. These people will thus continue to add value to their companies. People who do not wish to get out of their
comfort zone and adapt to technology, will ultimately be expendable as technology will continue to evolve and people “don’t really have a choice, they have to cope with it” (A2). We could classify these characteristics into the required soft skills, since they represent abilities that a person does not learn but which are part of someone. We could say it is a certain “kind of person” that is required to work with Big Data and that education is not enough.

6.2.2 Skills

However, personal characteristics and experience do not seem to be enough to understand and master Big Data. A combination of different skills is required.

First of all, we believe the base for a strong decision-maker profile is to have an intricate understanding of his business. This enables him to base every decision on a thorough knowledge of his business, its needs and shape the company’s strategy accordingly. When discussing the skills needed with our respondents, all of them mentioned the necessity to have a strategy in mind, a plan on how to implement Big Data as it would most probably impact the future of the company, be it failure or success. In this sense, the decision-maker needs to have strategic skills and analytical skills to encompass how technology and, in this case, Big Data, will impact his business in the short, medium and long term.

Thus, the decision-maker will need to have analytical skills in order to fully develop and use his strategic skills. What is important to note here is that both analytical skills, which we defined as data selection and data analysis, are linked to the understanding of the business and are both required to master synthetization skills. For instance, if the decision-maker does not understand the needs of his company in terms of data, he will not be able to select what information is relevant. What’s more, he will ask his teams to follow the wrong route and analyze irrelevant information which will lead to establishing the incorrect strategy. Also, he will be able to ensure the selected data is the needed data and avoid wasting time and resources. Once data has been selected and analyzed, synthetization is required to spread information in an understandable and efficient way to all concerned employees.

This understanding of the business, although on a different level, must be shared by all the employees as it will enable them to understand their roles and their purpose in the company. This automatically fosters trust and relationships between company and employees. What’s
more, it would help the decision-maker to establish himself as a bridge between departments and ensure the knowledge and strategy is understood but shared by everyone in the company.

In general, as stated above, the company’s performance with Big Data seems to rely primarily on the executive level. It is the management’s responsibility to inform stakeholders about changes in processes, and adoption of new tools and systems. This implies the need for proficient internal communication and the aptitude of the decision-maker to be a bridge between people.

Having proficient internal communication combined with an ability to be an active listener results in having downwards and upwards communication channels in a company. Developing such channels is key for a decision-maker as he will have constant feedback over his decision-making and he will be aware of a potential lack of skills within his firm. Having access to this information will improve his strategic skills by allowing him to identify challenges and develop strategies to overcome them.

An interesting point to develop is that all our respondents explained that they are facing challenges with Big Data. However, few of them knew where these challenges originated from and did not believe there was a lack of skills in their company. Nonetheless, by asking further questions about their Big Data issues, the lack of skills revealed itself as being the main source of those challenges. We believe that having these issues is not critical, the real challenge for companies is to understand that there is a lack of skill within their business and where these deficits come from. Decision-makers need to be aware of that to react and adapt their strategies accordingly.

Furthermore, timing and planning were constantly mentioned as being essential in any business decision since technologic growth moves at a very high velocity.

In order to ensure the relevance of the chosen strategy, one needs to create synergies between the departments even though the latter do not communicate with each other on a daily basis. Establishing a bridge between the IT and the sales department for instance will benefit the decision-maker but also the employees as data scientists and salespeople will share new insights with each other. The term “bridge” seems to be relevant here since it is part of the management’s role to minimize gaps and ambiguity. Seven out of eight respondents talked about how important communication is but also how it represents one of the biggest challenges in companies today. We could even say that a lack of communication can be the origin of a lack
of skills in companies nowadays. If communication is only going downwards, knowledge development and transmission will be weak if not broken. This is why one of the tasks of management is to communicate information easily but also be an active listener. In other words, the difficulty that many firms seem to encounter is that the few people able to understand and process the data, are rarely able to communicate their knowledge to other employees, and this creates a gap. This will lead many companies to invest money and resources into a process or phenomenon they do not fully understand or need.

It is important to note that not everybody in a data-enabled company needs to be knowledgeable about data analytics or data mining for example. However, we believe they should understand what data analytics bring to the company, how it is changing the business and how it will impact further business decisions. In this regard, a key aspect on which companies should focus on, is the internal communication. Here again, the executive level should make sure that a constant communication exists in order to make sure that, wherever it is needed, knowledge is aligned and available.

There are many ways for business leaders to implement Big Data in their company and deal with potential issues, drawbacks and challenges along the way.

Before developing this subject, we would like to reiterate the importance of basic, strategic and communication skills and how they impact the decision-making ability of business leaders.

The most prominent decision is training the employees. Training is a key element for companies since all eight respondents talked about the importance of regular trainings and skills development. The role of the management team is then to decide how employees should be trained (videos, seminars, workshops, IoT devices), how often (weekly, monthly, quarterly, yearly) and who should be trained. Indeed, it does not make sense to train people on a system they will never use. It also seems irrelevant to train people who do not work directly with Big Data. This would be, in theory, a bad business decision. However, in some cases, some employees, committed, eager to learn and passionate, might be worth training as they could potentially fill in an open position at some point in the future. This is again related to the decision-maker’s ability to understand the current and future needs of his company and allocate resources accordingly.

Others, as we’ve seen previously, might be reluctant to change or fear that new technologies such as Big Data might replace them. Training and, in this case, educating employees when
they are skeptic and anxious might change their mindset and reassure them as we believe transparency is essential in any team and company.

Ultimately, the main objective is to train employees to become more data-driven and more at ease with the technology used in the company.

However, training is not the only option available to decision-makers. The decision-maker’s awareness of a lack of skills in his company can be solved by developing a hiring strategy that meets current and future needs. Hiring is an opportunity to renew or complete the current workforce skills, implying of course that the right people are hired for the right position. As stated numerous times by our interviewees, “you need to hire the right people” (A8) but also “we did hire some developers and data scientists” (A2) and “we hired one person with high expertise in machine learning and AI” (A1).

Having the ability to recognize value-adding applicants is another skill developed through communication and basic skills but also met with experience and foresight. Today, there is a “large pool of talent” (A6) and the main challenge is simply to hire the right person for the position.

Finally, outsourcing is the last option decision-makers can opt for. This occurs after recognizing that the current workforce cannot sustain the current needs of the company, training might be too costly or too time consuming and hiring would result in a huge allocation of resources for a potentially unknown outcome. All of the respondents agreed on the fact that decision-makers need to “recognize they don’t have the skills to do this, they don’t have the people. So, they need to buy them, so either hire people or bring in a consultant” (A3). Working “with experts” (A8) is a reliable way to develop knowledge about Big Data and ensure the resources allocated towards technological change are well spent and will generate a viable return on investment.

It is not an easy task for decision-makers to know which strategy is best, or which strategies to combine in order to increase business performance and meet the needs and vision of the firm. We believe decision-making skills to be the most critical ones as they are fueled by all the other skills the person has acquired, combined with the person’s experience and characteristics.

Finally, the rise of new technologies, and Big Data in this case, seems to have had a huge impact on the daily job of CEOs and people on the executive level of a company. One respondent
pointed out that, with the rise of intelligent technologies and the amount of data points and information, his job has become entirely data-driven. He stated he was “flooded with charts, [he has] maybe 600 charts which is completely crazy... [he] cannot digest and go through so much data, it is simply not possible”.

This shows how much the job of the decision-maker has become more precise, less relying on gut-feeling but almost solely on data. The skills needed are thus very different from the ones necessary only a few years ago. It is important to note that the executive management had to go through a colossal and fast learning process to be able to continue managing their companies and building relevant and viable strategies.
Part 7: Discussion

The aim of this part is to review our findings in the context of the literature and the existing knowledge about the subject, as well as interpreting the study’s results and analyzing how they can help to answer our research question. We will first discuss the empirical findings that have completed existing theories in the literature review, before discussing the findings that have added new insights to the theory and that have not been mentioned by scholars yet.

7.1 Findings correlated to the literature

Beforehand, it seems clear that companies are nowadays facing challenges with the implementation of Big Data. Both, the theory through Kaisler, Armour, Espinosa and Money (2013) and our findings, demonstrated that, because of the massive volume of data that is produced every second, many firms encounter difficulties in processing it. Companies struggle in knowing which data to select, how to use it efficiently, how to visualize it and understand it, and how to communicate its outcome in an understandable way to other company’s stakeholders. Sivarajah et al. (2017) share the same idea regarding the struggle for individuals and companies to visualize data efficiently so as to enable fast decision-making.

The first challenges category we outlined in our literature review discusses “Data Challenges”. Those are directly related to the five main characteristics of Big Data. Four of them have been validated through our interviews: Volume, Velocity, Variety and Visualization as many of our respondents are experiencing challenges due to these. However, Veracity, which implies that some sources are inherently unreliable, has not been mentioned by our respondents but this does not refute or contradict the theory.

Many reasons for these Big Data challenges exist. Here again, the conducted interviews agreed with the existing literature about the second challenges category namely “Process Challenges”. Alharti et al. (2017) for instance, explain that many firms do not have a clear strategy in mind on how to address the various sources and formats of data and how this data will be useful for the company’s future performance.
By comparing the theory outlined by Pacino (2017) with our findings, a general lack of understanding about Big Data can be highlighted and Big Data seems still, to be a very vague topic for many people. Various reasons exist why knowledge about Big Data is so low, but according to our conducted interviews, the existing lack of skills can be related to a lack of resources, such as time (to train employees for instance), money and adequate technologies. As stated by Alharti et al. (2017) and Russom (2011), as well as by some of our respondents, using new technologies implies investing in new platforms and softwares for instance, but also in people specialized in that domain, able to train other company stakeholders. Our interviewees consider infrastructure readiness to be a challenge as implementing new databases and software is costly in time and finances.

Moreover, Stanton and Stanton (2016) share the same idea as our respondents about the fact that, basic technical skills such as programming, statistical knowledge, quantitative analysis and the ability of using various statistical and analytical tools, seem to be necessary to work with Big Data. However, these data specialists are often very expensive. What’s more, the computing power needed to analyze, and model data efficiently has not been discussed by our respondents. This does not refute the theories we mentioned earlier in this paper. Besides, warehousing capabilities have been highlighted but not seen as a challenge. It was simply seen as part of the process to become a data-enabled company.

The third identified challenges category, dubbed “Management Challenges”, has been related to all issues related to data safety and control. However, through our findings, management challenges encompass the managerial decisions and human resources aspects. For instance, Janssen, van der Voort and Wahyudi (2017) and Pacino (2017) shared the same idea as our interviewees about the importance of internal communication. Knowledge is not aligned between the different departments and communication is a real issue impacting the lack of skills within Big Data. Only few people know how to analyze and use data efficiently within a company. That is why the management department should allow constant communication between different team members in order to align knowledge. In this regard, we believe a lack of employees’ skills is the result of bad management. On the one hand, managers need to have good management skills allowing a constant communication, taking the right decisions concerning hiring, training or outsourcing for instance. On the other hand, managers need also to have some basic technical skills about Big Data.
7.2 Findings adding new insights to the literature

The second part of the discussion concerns all the findings that add new insights to our literature review.

First of all, it seems important to note that personal characteristics and previous experience are two elements that were not mentioned when reading about the required skills to work with Big Data. The theory mainly outlined skills that people learn through education and trainings. The conducted interviews added that commitment and passion are two important elements, and that it is not something a person can learn at school. In addition, all eight respondents agreed on the idea that experience is key and that the best way of learning how to work with Big Data is to practice. Contrary to our beliefs, having a clear definition of Big Data and having some basic technical skills, seems to be sufficient in the first place. In this sense, we believe that attending courses about this subject is really not enough and that to have a clear understanding, a person needs to firstly understand how Big Data will be used in this particular business and how it will affect it.

A second aspect which was not fully developed by scholars, is the relationship between technology and humans (Boyd & Crawford 2012). Many people are reluctant to use new technologies and scared about the impact it will have on their jobs in the future. However, there will always be a need for humans since technology only works in combination with humans supervising and controlling processes.

A third interesting aspect that has not been highlighted within the literature review, is the importance of decision-making skills. Taking the right decisions, according to the business and technological needs of a business, is essential for deciding whom and how to train within the enterprise. They should also reflect if outsourcing should be an option or if hiring data scientists should be considered.

Finally, the term “bridge” has been used by two of our respondents and seems to be a very interesting aspect to our research. We can link this term to two different ideas:

The first idea is related to the need for people to be able to understand which technologies need to be used within the company. Many firms want to use new technologies but do not know how technologies and in particular, Big Data will solve a problem in the company. Technology only
makes sense if it answers a business need. In other words, the people who are able to become the bridge between technology and business, and have both technological and business skills, will be the ones who will succeed. A person needs to have a clear understanding of the business in order to efficiently implement and use new technologies.

The second idea is related to the importance of internal communication. Bad communication is one of the main reasons for a general lack of understanding about Big Data. Indeed, an important aspect that transpired from our discussions is the fact that communication should not always be related to “informative communication” where the executive level communicates to employees. In other words, communication should not always be downwards, but also upwards. Decision-makers should thus be active listeners, able to pay attention to their team member’s requirements and needs. This will allow them to understand where potential challenges exist and what strategy should be adopted to overcome them. We believe it is essential that managers understand how employees feel in order to make the right decisions. The human resources play an important role and good communication is key. In other words, a company requires people to be the “bridges” between different departments and team members so that knowledge and information is aligned, and no gaps exist.

To conclude, the implementation of Big Data within companies does not only bring new business opportunities.

Many challenges come with it and our findings together with the existing theory, clearly outlined that these challenges arise from a lack of skills whereas the theory alone considered the lack of skills as being one of the challenges and not the origin.

There is no contradiction between them, but our findings allowed us to add a more practical insight to what the literature provided us with. The interviews provided us with different opinions on what skills are missing nowadays and how the executive level of a firm should adapt its strategy to develop the requested skills and knowledge.
Part 8: Conclusion

In this part, we will reflect on our paper by summarizing how we examined the literature and answered our research question through our findings.

As Big Data is still a very recent phenomenon, numerous scholars, companies and people discourse about it. What’s more, we have shown that it is still a widely unknown phenomenon for many people as the technology is still in its nascent phase.

It is clear now that Big Data, if used properly, brings a lot of benefits and advantages to companies choosing to invest in it. In our digital era, everyone creates data constantly and more and more sensors gathering data, are being installed every day. This puts Big Data at the epicenter of business intelligence and directly impacts business performance. However, it is also undeniable that numerous challenges arise from its implementation and many companies will have to cope with those to benefit from the technology. That is why we have chosen to investigate in our paper how the implementation of Big Data is impacting companies today, the challenges that arise from it and whether they originate from a lack of skills.

To answer our research questions we have read, analyzed and compiled the existing literature about that topic and compared it to the empirical findings gathered from the in-depth interviews we conducted. This enabled us to determine that several of those challenges originate from a lack of skills within companies. As most firms tend to not have a strategy in mind on how to address Big Data, due to what we have shown to be a lack of understanding, challenges arise. We believe a lack of internal communication to be the genesis for most, if not all, of the issues data-enabled companies are facing and will face in the future. These are mainly composed of a lack of understanding, lack of knowledge, communication and knowledge gaps between departments and employees and a lack of strategy alignment between the latter. In order to solve these issues, communication skills should be mastered by the executive level and communication channels established through all levels of the company both upwards and downwards, as everything depend on the decision-makers’ ability to gather information from every level of the company, analyze it and establish decisions based on that intelligence.
We believe having a deep understanding of the company’s structure, systems and procedures are essential to make thorough and sound decisions, but internal communication is as significant. Fundamentally, establishing the right systems to gather feedback and insights from employees is key in promoting open discussions and addressing the issues the company is facing. Essentially, every company could establish those. However, having the ability to integrate those issues into the company’s strategy is a skill we believe is critical for decision-makers to develop and perfect, especially when dealing with the implementation of Big Data. Indeed, working with new technologies implies a constant learning curve and for this to work efficiently, establishing strong communication channels and skills is key.

We trust that we delved into a topic that is not mature yet as it is still under investigation and vigorous debate by scholars. We strongly consider our paper to bring new insights into the literature as various reasons for the lack of skills in Big Data within many companies have been discussed and detailed in our interviews.

The next chapters of our paper describe firstly, the managerial implications of our study, in which we provide further suggestions on how decision-makers can adapt their Big Data strategy in order to overcome all the challenges related to a general lack of skills. Secondly, the limitations of our research are outlined and lastly, ideas for further research are highlighted.
Part 9: Managerial implications

By conducting our research, we had the chance to gather very interesting insights, both related and not related to our research question.

9.1 Further managerial implications related to our research question

Today, it is critical for the executive level to comprehend that skills are missing and that it needs to either work with external experts, hire new people, or train their existing employees. In other words, it is key to be able to understand and admit that changes are required and sometimes, the current workforce does not have the necessary skills to support a new strategy. To do so, based on the theory and findings of our study, we suggest companies to go through different steps when implementing Big Data in their business activity. These steps originate directly from the answers our respondents provided in the part 4 of our questionnaire, which focuses mainly on reflections and suggestions regarding Big Data implementation (see Appendix 3). We believe that going through the following steps will allow companies to overcome potential lack of skills and use Big Data in an efficient manner.

The first step is to thoroughly develop the knowledge of the executive level towards Big Data. It is key to make sure the decision-makers understand Big Data, how it will be used in their company and how it could increase the business performance. Only if and when the executive level has this understanding, knowledge transmission to other team members can be possible.

The second step is to clarify where Big Data can be implemented and where it cannot or does not make any sense. Here again, it is critical to understand where there is a need for Big Data in the company as it only makes sense to become data-enabled if you know where to use it and towards what purpose.

Thirdly, a Big Data strategy has to be implemented as planning and timing are essential. When building the strategy, it is critical to decide if the company should outsource its data activities, train its existing employees or hire data-skilled people for instance. It is also important to take
into consideration where the company needs to invest in new databases and software for instance, how much and for how long.

Fourth, defining measurable outcomes is vital, many firms have a strategy and vision but do not assign quantifiable milestones to it. Big Data, as with every new technology, allows businesses to make astonishing things as we have seen earlier, and data engineers and scientists, when passionate about their work, will continuously create new software, interfaces and so forth. However, as with every investment, it needs to make business-sense and provide a sustainable return on investment. One needs to always shape his strategy based on the intelligence gathered, being technological, financial or else, to ensure the prosperity of the company.

Fifth, one needs to establish strong communication channels between the decision-maker and the employees but also between the various teams with the aim to align knowledge and facilitate the communication flow internally. It is also important to guarantee upwards communication in which the executive level should listen to the employees’ requests, suggestions and questions.

Finally, the company should always have in mind that technology moves extremely fast and by the time one has implemented Big Data and new processes, they might already be obsolete. That is to say, a company needs to constantly re-evaluate its business needs and seek innovation and improvements. As technology and Big Data are in a permanent evolution cycle, constant learning and adjustments are required.

To conclude, companies need to adapt to the technological evolution to ensure their relevancy. All companies could use Big Data, but it is important for them to have a clear understanding of the technology and the ability to identify whether they need it and if so, where changes need to be made in their current structure. Then, it is crucial to map out where a lack of skills exists in order to overcome this challenge.
9.2 Further implications not directly related to our research question

On another hand, we also obtained interesting insights that are not directly related to our research question, but which allowed us to increase our overall understanding of Big Data and enlightened us about its role and impact on business.

The first insight that we obtained through our interviews is related to the overall impact of new technologies on businesses. Our study allowed us to understand how business processes have changed and how this has impacted the required skills for a specific job, which have evolved drastically during the last years. Jobs have changed as new platforms and tools are used which require a consistent and fast adaption from all company’s stakeholders. Many interviewees highlighted the idea that companies are forced to adapt to technology such as Big Data. This has sometimes, negative consequences on employees who are scared to change their habits and feel overwhelmed by new technologies and the speed at which they are developed. CEOs explained how Big Data and Artificial Intelligence are often associated to the idea that human beings are becoming less important or even expandable, and that existing jobs will be replaced by robots and computers for instance.

Another interesting insight we obtained, concerns the definition of Big Data. On the one hand, all respondents defined Big Data in a different way, according to their specific industry and business. This outlined the idea that Big Data is still a recent phenomenon that does not have a universal definition. On the other hand, all eight respondents agreed by saying that Big Data provides them with more information that when used correctly, allows to predict the future and increase business performance by improving the decision-making process.

Lastly, an important aspect emphasized by five respondents, of which we had no knowledge before, concerns the General Data Protection Regulation that will be enforced in Europe on May 25th, 2018. It will impact every business handling personal data. Data safety and privacy issues are very important aspects that have gained more attention within the past few years especially for businesses dealing in a business to consumer market.
In a word, our study allowed us to gain more knowledge about Big Data and recognize what changes companies can adopt to overcome potential lack of skills that arise from this phenomenon.
Part 10: Limitations

This part of the paper highlights the various issues we encountered as researchers while writing our thesis. These limitations can be separated into three categories. The first category comprehends the inherent challenges encountered before even conducting our interviews. The second category comprises the limitations related to methodical challenges while conducting the interviews, and the third category explains all the challenges related to communication challenges during the interviews.

10.1 Inherent challenges

When conducting our qualitative study, one of the first challenges we encountered was the fact that not everyone we wanted to interview, agreed to participate. Our objective was to obtain approximately ten hours of interviews (so ten interviews lasting one hour). We both contacted all the companies that seemed to fit our criteria but not everyone responded or had the time to help us with our research. As a result, we gathered short of ten hours of interviews.

A second limitation we encountered can be related to the language barrier. We were both limited to four languages, which are French, German, English and Arabic. In other words, someone not speaking one of these languages did not have the ability to participate in our study. This reduced our “research environment”.

Furthermore, we had the opportunity to interview companies from different countries. On the one hand, this allowed us to conduct an international study but on the other hand, it required more adaptation to different cultures, visions and work processes.

Since we decided to not focus our research on one particular industry, the analysis of the answers seemed to be more complex at some points. That is to say, each interview was very different, and every company had a different vision and definition of Big Data, related to its industry. Thus, we had to step back and see the bigger picture when connecting the dots.

Finally, another limitation was that, by being two students, we were aware of the potential given impression of being unskilled and that our inquiry is not very important and might be a waste of time for the respondents.
10.2 Methodological challenges

Secondly, we encountered multiple challenges related to our decision to do qualitative interviews to answer our research question.

First of all, doing qualitative interviews implies having an in-depth discussion with our respondents. However, since we contacted some people in the United States of America (USA), France and the United Arab Emirates (UAE), we could only offer to do the interviews through Skype or through conference calls. This solution was very time consuming as we communicated multiple times with potential respondents to find suitable timings for our interviews but also means of communications suitable to both parties. This was particularly challenging considering different time-zones as well as the fact that for example, the UAE does not apply Daylight Saving Time which occurred during our interview time-frame. Thus, when communicating, we always had to make sure to use both Swedish time zones and the one from the respondent’s country to avoid misunderstandings.

Furthermore, remote interviews implied communicating Skype usernames or phone numbers beforehand, making sure our internet connection was working properly and having a back-up plan should it not work as expected (i.e. a phone number). Finally, we felt like remote interviews were not as personal as meeting someone in real-life and this sometimes made the interviews somewhat harder to kickstart especially when encountering some minor problems with microphones or sound quality in the beginning.

Finally, a very important methodological issue we encountered, originates from the limited time we had to write our thesis. We would have liked to conduct more interviews but being limited by the time, we were able to only interview eight different companies. In this regard, it seems important to notice that our findings cannot be generalized. That is to say, our interpretation and hypotheses can be limited to a certain degree.

10.3 Data collection challenges

As we mentioned earlier, we discussed our topic with multiple people around the world, most of them were acquaintances or contacts of ours but more importantly, they were people we contacted because of their relevance to our study. Interviewing them was particularly tough as our interviewees were experienced Executives and we were about to dig into the challenges
their companies are facing and how they view a potential lack of skills in their workplace which is not a joyful subject.

What’s more, we asked open questions in our questionnaire meaning the interviewees had some control over the interviews. Indeed, they had the ability to not only avoid our questions but could also talk about a topic not related to the questions at all, willingly or not.

Big Data being a recent phenomenon also signifies that it is a source of competitive advantage for companies who use this technology efficiently. This translates into a potential reluctance to discuss this topic with us, compelling us to reformulate or deduce some information from the answers.

Also, it is possible that the people we interviewed are not aware or not willing to admit they have a lack of skills in their company, as it can be regarded as a weakness or flaw in one’s strategy or vision.

Finally, one of the most important limitations of our study, and one which would necessitate further research, is the fact that we only interviewed people with positions at the executive level of their company. Thus, it is important to note that when we defined the ideal profile of the employee, we only gathered the insights and perceptions of the executive level and not of the person concerned. In other words, we only have a limited perspective of the profiles we built.
Part 11: Further Research

This paper has examined the main challenges that arise with the implementation of Big Data within companies and how these challenges originate from a general lack of skills. Since Big Data is a recent phenomenon, further studies can be made on that subject.

In this regard, we believe it could be interesting to conduct the same research but on a larger scale in order to make generalizations possible. To do so, the research should be conducted over a longer time period.

In addition, a case study could be conducted in which the same research question is related to one particular company. This would allow the researchers to have a deeper understanding of the company’s business strategy and concrete suggestions could be made based on more intricate and in-depth observations. Furthermore, each lack of skill that we have identified within our research paper, such as the lack of communication, understanding or time, could be developed. The research question could focus on one particular lack and in this sense, analyze where it comes from and create a framework on how to overcome it.

Furthermore, further studies can be done on the importance of having engineering and business skills. This aspect has been mentioned by various respondents and we believe, it requires a deeper analysis since it seems to be the key to success when working with Big Data.

Finally, the same study could be conducted by interviewing only employees in order to gain another perspective of the research problem and profiles built earlier in the study and compare both views.
References


ccSINTEF. (2013). Big data, for better or worse: 90 % of world’s data generated over last two years. ScienceDaily, May 22, 2013.


Appendices

Appendix 1: Interview Design Table

<table>
<thead>
<tr>
<th>Interview Design</th>
<th>Interview 1</th>
<th>Interview 2</th>
<th>Interview 3</th>
<th>Interview 4</th>
<th>Interview 5</th>
<th>Interview 6</th>
<th>Interview 7</th>
<th>Interview 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent’s code</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
<td>A5</td>
<td>A6</td>
<td>A7</td>
<td>A8</td>
</tr>
<tr>
<td>Name of the interviewee</td>
<td>Gilles Ivetot</td>
<td>Frederique Stumm</td>
<td>Mark Eid</td>
<td>Simon Werner Zankl</td>
<td>Gilbert Salhani</td>
<td>Antonio Boulos Yvette Madi</td>
<td>Mario Montag</td>
<td>Delphine Genin</td>
</tr>
<tr>
<td>Nationality of interviewee</td>
<td>French</td>
<td>French</td>
<td>Lebanese</td>
<td>Swedish</td>
<td>French</td>
<td>Lebanese Lebanese</td>
<td>American</td>
<td>French</td>
</tr>
<tr>
<td>Company</td>
<td>Stargazer</td>
<td>Techni-Contact</td>
<td>Crimtan</td>
<td>TrustCruit</td>
<td>IQVIA</td>
<td>BPN</td>
<td>Predikto</td>
<td>Sodexo</td>
</tr>
<tr>
<td>Size of the company</td>
<td>Start-up</td>
<td>Small</td>
<td>Large</td>
<td>Start-up</td>
<td>Large</td>
<td>Large</td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>Scope of the company</td>
<td>International</td>
<td>National</td>
<td>International</td>
<td>National</td>
<td>International</td>
<td>International</td>
<td>International</td>
<td>International</td>
</tr>
<tr>
<td>Location of the company</td>
<td>San Francisco, USA</td>
<td>Paris, France</td>
<td>Dubai, UAE</td>
<td>Jönköping, Sweden</td>
<td>Paris, France</td>
<td>Dubai, UAE</td>
<td>Atlanta, USA</td>
<td>Paris, France</td>
</tr>
<tr>
<td>Communication method</td>
<td>Skype</td>
<td>Skype</td>
<td>Skype</td>
<td>Face to face</td>
<td>Skype</td>
<td>Skype</td>
<td>Skype</td>
<td>Skype</td>
</tr>
<tr>
<td>Length (in minutes)</td>
<td>55</td>
<td>53</td>
<td>71</td>
<td>55</td>
<td>41</td>
<td>65</td>
<td>32</td>
<td>59</td>
</tr>
</tbody>
</table>
Appendix 2: Consent Form

Consent Form  
Research on Big Data Challenges

Please consider this information carefully before deciding whether to participate in this research.

**Purpose of the research:** To understand more about Big Data, its implementation in companies, the challenges that might arise from it. More specifically, we want to investigate the lack of skills in companies today.

**What you will do in this research:** If you decide to volunteer, you will be asked to participate in one interview. You will be asked several questions. Some of them will be about Big Data, others will encompass the challenges regarding its implementation and finally a lack of skills in companies today. With your permission, we will tape record the interview. You will not be asked to state your name on the recording.

**Time required:** The interview will take approximately one hour.

**Risks:** No risks are anticipated.

**Benefits:** This is a chance for you to give us your outlook on Big Data and share your experience of that particular phenomenon. It is also an opportunity to reflect on this fast-growing topic. You will also have the opportunity to receive our final report if you wish to.

**Confidentiality:** Your responses to interview questions will be kept confidential. At no time will your actual identity be revealed. You will be assigned a random numerical code. The recording will be destroyed on the 21st of May 2018. The transcript, without your name, will be kept until the research is complete.

The key code linking your name with your number will be kept in a locked file cabinet in a locked room, and no one else will have access to it. It will be destroyed on the 21st of May. The data you give us will be used for our research paper and may be used as the basis for articles or presentations in the future. We won’t use your name or information that would identify you in any publications or presentations.

**Participation and withdrawal:** Your participation in this study is completely voluntary, and you may refuse to participate or withdraw from the study without penalty or loss of benefits to which you may otherwise be entitled. You may withdraw by informing the experimenter that you no longer wish to participate. You may skip any question during the interview but continue to participate in the rest of the study.

**To Contact the Researcher:** If you have questions or concerns about this research, please contact:  
Patrick Ishac: patrick.ishac@kegdebs.com, +33 6 98 92 34 28  
or  
Hannah Dussoulier: Hannah.dissoulier@kegdebs.com, +33 6 33 56 71 31

**Agreement:**  
The nature and purpose of this research have been sufficiently explained and I agree to participate in this study. I understand that I am free to withdraw at any time without incurring any penalty.

Signature: ___________________________ Date: 20/03/2018

Name: ___________________________
Appendix 3: Interview Template

The questions highlighted in bold are our main questions. The ones in brackets are potential questions we could ask if the person does not answer our questions appropriately. The questions in yellow, purple and blue in italic are potential follow-up questions.

Date:
Person interviewed:
Company:
Position:
Experience:
Duration:
Location:
Communication model:

- PART 1: COMPANY

The first part of the interview has an introductory purpose, in which the company explains how and why Big Data is used in its business processes. The various questions of this part will allow us to justify that the chosen respondent is relevant for our research.

How would you define Big Data? If you had to choose 3 words, which one would you use to define it?

How do you use Big Data in your company and what do you use it for?

How much of an impact has Big Data created?

How does the company in which you are working currently performs in the use of Big Data?

How much have your processes changed to accommodate the use of Big Data?
or
What makes your internal processes big-data enabled?
PART 2: CHALLENGES

Within the second part of the interview, we will focus our questions on the challenges the respondent may encounter while using Big Data. This part is of high relevance since we will be able to compare our hypothesis based on the empirical work with a more practical point of view.

What are the main challenges that you encounter within Big Data?

PART 3: LACK OF SKILLS

The third part of our questionnaire represents a key element within the interview since it is directly related to our research question. Within this part, we will seek to understand where the challenge “lack of skills” originates from, as well as what companies do to resolve this challenge and what set of skills are required. This will allow us to increase our knowledge about the processes that companies need to implement in order to use Big Data efficiently.

Which skills do you believe are required to work with Big Data considering you’re not a Big Data analyst?

Would you say there is a lack of skills in your company today? Please elaborate.

Yes. Why? What are the skills that are missing?

No, then why isn’t there one? What did you do to avoid this issue/challenge?

Do you believe other companies are facing the same challenges or other additional ones? What are the reasons for this lack of skills?

1.1 In your opinion, why is there a lack of skills related to Big Data in other companies today? What are the reasons for the lack of skills? (education, training, technology adoption)

How did your company react to the Big Data implementation from a human resources point of view?

If you trained, what skills did you develop/expand within your current workforce and how? Have you felt that this training was sufficient? outcome?

If you hired, what skills were you looking for?

If you outsourced, what were the factors that influenced your decision?
• PART 4: SUGGESTIONS + REFLECTIONS

Finally, we will end the interview with a reflection part in which the interviewed will be able to provide some future suggestions and recommendations within the use of Big Data (for his own company but also in general). This part will conclude the interview.

What would you suggest to companies to increase their understanding of Big Data?
(What do you think is needed to successfully implement / use / understand Big Data?)

What would you do differently if you had to do it from scratch?