Aalto-yliopisto
Kauppakorkeakoulu

Big Data in Finish Companies
What has been done and the results achieved

MSc program in Information and Service Management

Maisterin tutkinnon tutkielma

Mirjamaria Petäjäniemi

2015

Tieto- ja palvelutalouden laitos
Aalto-yliopisto
Kauppakorkeakoulu

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Mirjamaria Petäjäniemi

Title of thesis Big Data in Finnish Companies – What has been done and the results achieved

Degree Master of Science in Economics and Business Administration

Degree programme Information and Service Economy

Thesis advisor(s) Merja Halme, Anu Bask

Year of approval 2015 **Number of pages** 69 **Language** English

Abstract

This study considers the topical subject of big data in the context of Finnish business environment. A need for clarifying the concept of big data is recognized. The first objective of the research is to untangle, introduce and analyse the concept of big data. Also a research gap in the area of big data utilization in Finnish companies is identified. Another objective is to empirically assess the state of big data in Finnish companies and the results that the companies have achieved by utilizing big data.

The research questions draw from the theoretical models of big data maturity and success factors of business intelligence implementations. The research questions include the following:

1. How big data mature are Finnish companies that are utilizing big data?
2. Is the big data maturity model applicable in Finnish business environment and does it succeed in differentiating the companies with different levels of maturity?
3. What kind of external factors and internal competences the companies recognize as defining their big data potential.
4. What are the factors that Finnish companies identify as contributing towards success in big data efforts?
5. Are the identified success factors aligned with the model of Yeoh and Koronios (2010)?

The method of multiple case study is applied in order to answer the research questions. The data consists of ten interviews with experts of large Finnish companies utilizing big data.

In the analysis of the interview data the maturity model is successfully applied in Finnish context. Interviewed companies are mostly on the early stages of big data maturity. However, they are using big data in versatile ways and in many different business functions.

SWOT-analysis is used as a tool to recognize external and internal factors defining the big data potential of the companies. A large share of the companies report access to large amounts of data and strong know-how in the analytics area as their strengths. Identified weaknesses include e.g. lack of organizational agility. Increasing availability of data and improving technological solutions are seen as the most predominant opportunities. Fast pace of change and fierce competition in the area are things that the companies find challenging. Other mentioned treats include crisis with privacy and changing privacy legislation.

Evidence of all other success factors of Yeoh's and Koronios' (2010) model is found in the interview data except the success factors in the technology dimension. This might indicate that the companies do not consider the technological factors as critical as factors related to the process and organization. The data also suggests that the success factor of openness to look for partners and solutions outside the company's own industry should be included in the organizational success factors.

Keywords Big data, Finland, Success factors, Maturity

Acknowledgements

As much as this study has been an individual project, there are still many people who helped to make it possible.

First of all, I would like to thank my supervisor professor Merja Halme. She gave me her full support even though she had her doubts about the plausibility of my initial idea. Assistant professor Anu Bask was another helpful person from Aalto University School of Business. I had the opportunity to observe the process of doing science while working as a research assistant in her research project during the summer 2014. This job introduced me to the world of science and lowered the bar to start working with my own project. The project also gave me the idea of a research method to use.

I express my gratitude to all the wonderful people who contributed by giving an interview. I was surprised to find out how supportive, friendly, and willing to help the Finnish business people are. Your views and experiences made this study valuable. I am very glad that you wanted to share them.

I also have to mention Pekka Puska, sales manager at Cisco Systems, who arranged for me to discuss with their expert, and Floris Grandvarlet himself. Mr Grandvarlet was very kind to offer his support and excellent advice.

There were still many others who showed their support and concern along the way. Especially, I want to mention my dear mother and amazing roommate, who were always willing to patiently listen to my thoughts and feelings about the project.

Finally, all the glory belongs to God almighty, who created me, redeemed me, and made me His own. He is forever faithful and continues to guide my steps and provide for all my needs and beyond.

Table of Contents

List of Figures

List of Tables

1 Introduction

"We believe that we live in an era of big data." (Humberty, 2015) "IoE and the era of big data are transforming our lives." (World Economic Forum, 2014) "We live in an era of big data." (Davenport, 2013) "The era of big data has begun." (Boyd & Crawford, 2012) A search with a keyword "big data" returns 2 192 articles in the Scopus database (4 November 2015). Nearly a half of the articles was published in 2015. On Google search keyword "big data" returns 52 million results. It seems evident that "big data" is something characteristic of this decade and something everyone is talking about.

However, it appears that people are not always taking about the same thing when they are talking about big data. The concept is also connected with more general changes and phenomena, like the development of new technologies and digitalization of businesses and the whole society (Rüping, 2015). One of the objectives of this research is to untangle, introduce and analyze the concept of big data – What does it really mean, how different people see big data, and what are the other concepts and phenomena related to it?

"Big data era" has also arrived in Finland. Nearly all of the biggest IT companies in Finland claim to have something to do with big data. Still, the big data utilization in Finnish companies has been studied surprisingly little. Finnish Ministry for Transport and Communications published a report in 2014 that considers the possibilities of big data for Finnish companies and the public sector. They claim that the opportunities are vast but exploitation of data is still in its infancy (Liikenne- ja viestintäministeriö, 2014). Accenture interviewed over 130 managers in Finnish companies in 2013 about digital business. One of the focus topics was big data and analytics. Their message is pretty much the same as the one of Ministry for Transport and Communications. They interpret that big data and analytics are only breaking through in Finland (Accenture, 2014). Still, the media tells us about Finnish companies that include big data as a major part of their strategy. Saarelainen (2015, January) interviewed four Finnish companies for his article "Ison datan kalastajat". He claims that actually many companies are utilizing big data but are keeping silent (Saarelainen 2015, January).

Liikenne- ja viestintäministeriö (2014) and TIVIT (2011) consider the possibilities in utilizing big data in different Finnish industry sectors. Accenture (2014), Liikenne- ja viestintäministeriö (2013) and Intell and SAS Institute (2015) surveyed representatives of Finnish organizations about their attitudes and beliefs about big data. However, these studies have not comprehensively studies the actual big data applications in Finnish companies. The previous studies do not consider the outcomes of the big data efforts in Finnish companies either.

1.1 Research objectives

The first objective of this research is to shed some light to the current state of big data utilization in Finland. Is it actually the case that everyone is talking about big data but no-one is actually doing much? Or are the leading companies already applying big data extensively but are doing it in silence?

The second objective is to assess the concrete business benefits that the companies have reached by utilizing big data. Considering the recent fuss around big data it is very likely that some companies have implemented solutions just to stay in the frontline of development. According to Boyd and Crawford (2012) there are many diverse opinions and emotions related to big data. They actually define big data as a mythology (Boyd & Crawford, 2012). Gartner technology hype cycle for year 2014 in Figure 1 illustrates how Gartner interprets that big data is currently moving from the peak of inflated expectations to the stage where the expectations start to become more realistic.

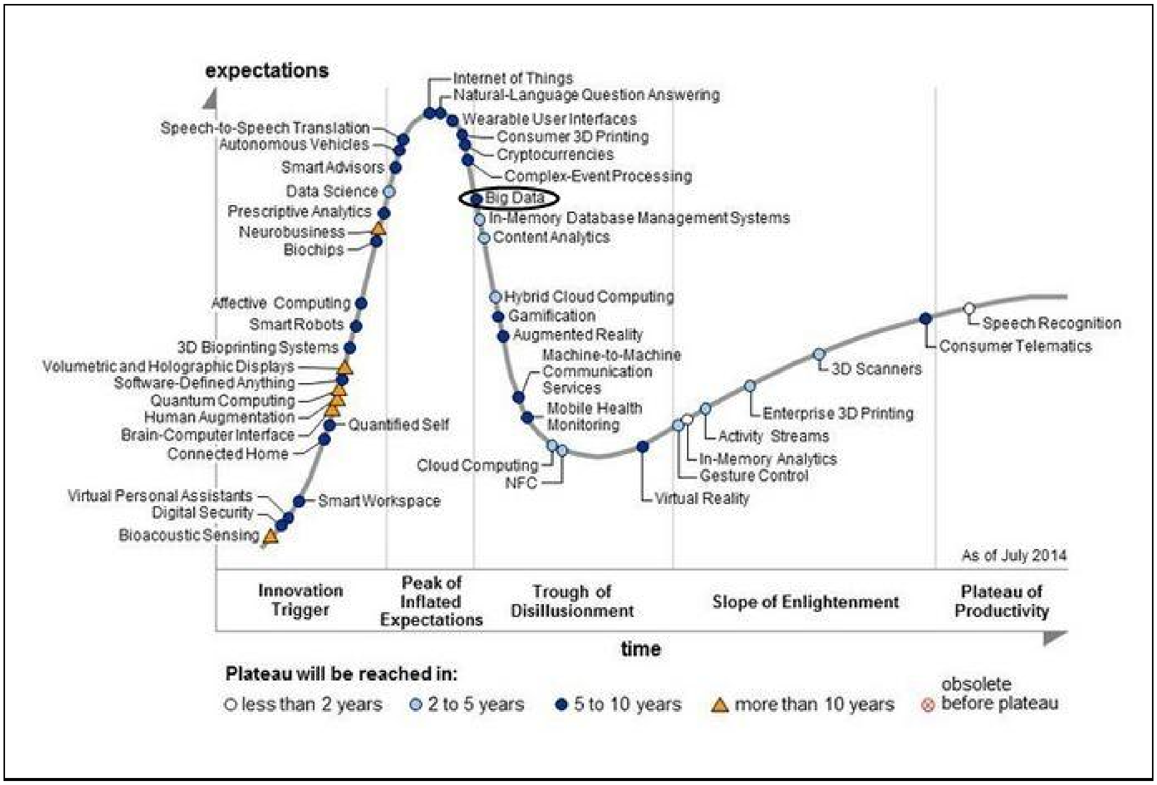


Figure 1. Gartner technology hype cycle (Gartner, 2014)

This seems to be the perfect time to assess the real results of the big data initiatives this far.

Even more interesting question than whether there are benefits from utilizing big data is: What are the different factors differentiating the companies that are successful in realizing benefits and achieving results from their big data initiatives and the companies that are not?

1.2 Research Questions and Method

In order to understand the big data projects of Finnish companies in relation to a wider theoretical framework we apply two different theoretical models related to big data. The state of big data in a company can be assessed by applying the big data maturity model first described in *The Global Information Technology report 2014* by World Economic Forum (2014). The model introduces four different stages of sophistication and maturity in the usage of big data. This study aims to apply the big data maturity model to Finnish companies that have implemented big data solutions. Research questions related to big data maturity include:

1. How big data mature are Finnish companies that are utilizing big data?

2. Is the big data maturity model applicable in Finnish business environment and does it succeed in differentiating the companies with different levels of maturity?

3. What kind of external factors and internal competences the companies recognize as defining their big data potential.

The second theoretical model considers the question of the results and success of the big data projects and the different factors contributing towards a successful initiative. Yeoh and Koronios (2010) identified seven different success factors for BI implementations based on their Delphi-study and related case-studies. This study aims to consider if these success factors are applicable also in the context of big data initiatives. The related research questions are:

4. What are the factors that Finnish companies identify as contributing towards success in big data efforts?

5. Are the identified success factors aligned with the model of Yeoh and Koronios (2010)?

The empirical part of this study intends to answer the above research questions. The method used was a multiple case study consisting of ten interviews with representatives of big Finnish companies that are already utilizing big data.

1.3 Structure of the Thesis

The structure of this thesis is as follows:

The second section starts by defining big data and discussing the different suggested definitions. Then the background of big data phenomenon as a new orientation of business intelligence and analytics is discussed, followed by the introduction of factors enabling the big data phenomenon. Next, section two briefly discusses the different application possibilities of big data solutions for selected industry sectors. Finally, the second section considers the different challenges and problems related to the utilization of big data.

The third section introduces the theoretical framework for the empirical part of this study. First a framework conceptualizing BI is introduced. The empirical part of the study is based on two related models: the big data maturity model adopted from *The Global Information Technology Report 2014* (World Economic Forum, 2014) and the BI implementations success factors framework developed by Yeoh and Koronios (2010).

The fourth section considers the previous big data research in the Finnish context. Firstly, the current state of big data in Finland is assessed. Then the different opportunities that big data creates especially to Finnish industry sectors are addressed. The section concludes by considering challenges in applying big data unique to Finnish environment.

The fifth section reports the research method and the research process of this study. It describes the different stages of the research process: method selecting, question formulation, data collection, and data analysis.

The empirical findings of the study are introduced in section six. The findings related to the big data maturity model are introduced first. Then the findings concerning the research questions related to the success factors model are explained. Finally, the results of the SWOT-analysis are explained.

Section seven *Discussion and Conclusion* repeats the research questions introduced in this section and discusses the empirical findings related to each question. This part is followed by some general discussion about the results. Finally, the implications for theory and practice and the limitations of the study are discussed, and some suggestions for further research are made.

2 Literature on Big Data

The different definitions of big data found in the literature are discussed at the beginning of this section. Big data is compared with the more traditional business intelligence approaches. Part 2.3 discusses the phenomena and technologies related to big data. After that, big data potential of selected industry sectors are discussed. Finally, part 2.5 considers the challenges the companies face in using big data.

2.1 Definitions

Big data is often referred to as a buzz word and a trend. IT companies are keen to use the most recent and trendy vocabulary in their marketing. People who want to appear contemporary also use terms like big data willingly even if they would not have a sound understanding of their meaning. It seems to be the nature of humankind that when something in the world is changing radically there is such a tumult around the phenomenon that it becomes hard to see the real essence of the matter. This is true especially where money and opportunities to gain profit are involved. Thus, it is not surprising if the term big data has become somewhat vague. Also in the scientific world many different definitions exist (see e.g. Chen et al., 2014a; Sharda, 2013, Wu et al., 2014). However, these descriptions can be seen as different angles of seeing the phenomenon rather than competing opinions of which only one is correct.

Since the term was born at the beginning of the 21st century big data has been described using the three V's: volume, velocity, and variety (Chen et al., 2014a). McKinsey Global Institute (2011) focuses on the volume in their definition of big data: "*Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse*." The amount of data collected and stored each day is growing constantly due to growing data storage capacity, lowering the costs of data storage, the development of networks, and the overall digitalization of the society. For example, Facebook's data warehouse stores 300 petabytes of data. (Wiener & Bronson, 2014) Storing this amount of data would require 408.7 million CDs. A pile of 408.7 million CDs would be approximately 490 km high. Velocity refers to the fast pace at which the data moves. In some cases there is a need for almost real time data processing. Examples of such cases include transaction fraud detection at the point of sale or automatic customization of a web page for each user (Madden, 2012). Variety means that the data comes in different shapes. It can be text, video, click-stream, or log data to name a few. The variety of big data challenges the described traditional way of storing data as it is often lacks a defined structure.

Some sources include one or more V's in addition to the original three. World Economic Forum (2014) emphasizes value which is what everyone is striving to extract from big data. Zikopoulos et al. (2015) introduce veracity that relates to the varying quality and reliability of the data. Big data can include much noise alongside the actual valuable information and the sources of the data can be unreliable. For example, it is not definite that a profile on Twitter actually represents a real person (Boyd & Crawford, 2012).

However, big data does not only refer to the nature of the utilized data. The term also reflects the developments in data storage and analysis technology and the change in how data is seen and used. For example, Sharda (2013) emphasizes the changing technology by stating that: "At the core of big data research are business concepts, technologies and statistical techniques necessary to design highly scalable systems that can collect, process, store, and analyse large volumes of both structured and unstructured data". Also Tarvis et al. (2013) refer to big data as a term that computer scientists have invented to describe the changing technology.

Wu et al. (2014) take a data management angle to big data and introduce the HACE theorem. The beginning of the acronym put the focus on sources of big data. Data comes from heterogeneous, autonomous sources with distributed and decentralized control. Businesses and other organizations now increasingly utilize data from sources outside of their control, such as social media. Big data technologies make it easier to combine data from different sources. The latter part stands for the complex (C) and evolving (E) nature of the relationships that can be found from big data. Inferential or advanced analytics seek to predict the future or find causalities and underlying patterns from the data in contrast to just describing, summarizing, and visualizing the data (Bose, 2009; Cohen et al., 2009).

Big data can also be seen as a wider phenomenon. Boyd and Crawford (2012) define big data from a sociological point of view. They define big data as "a cultural, technological, and scholarly phenomenon that rests on the interplay of:

(1) Technology: maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets.
(2) Analysis: drawing on large data sets to identify patterns in order to make economic,

social, technical, and legal claims.
(3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy."

To conclude, in business context big data is best described as a change of paradigm in analytics as, for example, Chen et al. (2012) do. Big data is often defined in relation to the conventional methods of analytics and business intelligence (BI). Chen et al. (2012) define BI and analytics as "techniques, technologies, systems, practices, methodologies, and applications that analyse critical business data to help an enterprise better understand its business and market and make timely business decisions". There has been a growing interest towards BI and analytics since the 1990s and the big data trend continues this phenomenon. Traditional BI is well established with state-of-the-art standardized software tools and best organizing and management practices for retrieving, storing, analysing, and visualizing the data. Big data is the next generation of analytics, the evolvement of which is driven by the fierce competition in the global digitalized economy.

2.2 Challenges for Traditional Business Intelligence

The main building blocks of traditional business data analysis are the enterprise data warehouse (EDW) and the business intelligence tools. A data warehouse is a database storing all data needed in decision making. This data is collected from different sources and stored into the data warehouse through extract, transform, and load (ETL) operations. The data can then be aggregated (rollups, drill-downs, slice-and-dice, pivots) and visualized (scorecards, dashboards) using different BI tools. Figure 2 gives a visual representation of a traditional BI architecture. (Chaudhuri & Dayal, 1997; Chen et al., 2012; Cohen et al., 2009)

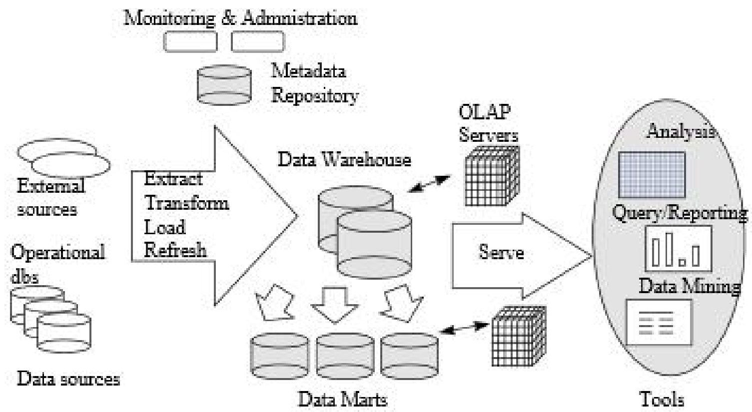


Figure 2. Data warehousing architecture (Chaudhuri & Dayal, 1997, p. 518)

The differences between traditional BI and big data are well reflected in the three Vs definition (high volume, variety, and velocity of data) and the HACE theorem (large-volume, heterogeneous, autonomous data sources and seeking complex and evolving relationships among data).

Big data volumes do not necessarily pose a significant challenge to traditional database management systems (DBMS). According to Madden (2012) most of the commercial DBMSs can handle multi-petabyte databases while many open-source systems lag behind. Jacobs (2009) notes that getting large datasets into a traditional database is not a problem. The challenge arises when you want to analyse big quantities of data. It is easier to get data in than out as the databases are often designed to handle transactions. Variety characteristic of big data is where the traditional way of storing data becomes laborious. Traditional relational databases are like collections of spreadsheets. The format of each entity and relationships to other entities and tables are carefully defined. This organization is called a schema. The data has to be structured to fit such a database. Thus utilizing unstructured data such as free text, pictures, or video is very difficult (McAffee & Brynjolfsson, 2012). The traditional systems also fall short where time is in essence in handling, especially analysing, the data (Madden, 2012). Extracting, transforming, and loading data to the system takes some time. Also, as Jacobs (2009) notes, the databases that are optimized to write rather than read do not analyse huge amounts of data quickly.

As the HACE theorem emphasizes, big data is often combined data from different sources. The traditional approach repels new data sources as the new data has to be carefully cleaned and integrated into the existing information schema. These integrations are time and money consuming and thus new data sources are rarely added to the system after the initial launch. HACE theorem also refers to the need for more complicated analysis. Traditional BI tools only provide simple standard functionalities for analysing and aggregating the data. Thus it is common to load a portion of the database to desktop statistical packages like SAS, R, or Matlab to do more customized and sophisticated analysis. However, this process is slow and not very large amounts of data can be utilized. (Cohen et al., 2009; Wu et al., 2014)

2.3 Enablers

2.3.1 Data sources

Why is the volume of data growing so rapidly and what are all of the different data sources? After the invention and world-wide adaption of the computer and the internet the information of the society has increasingly been transformed into a digital format. People interact, work, trade, study, and entertain themselves online. Also the amount of information has grown as digital data can be created, stored, and copied automatically quickly and cheaply. The technologies are developing rapidly and data storage and analyzation is becoming less expensive and time and effort consuming. Click-streams, browsing and search histories created by websites and search engines, social media content, and data created by different sensors are frequently referred to as sources of big data (see e.g. Chen et al., 2012; Chen & Zhang, 2014; Chen et al., 2014a).

Especially, the recent progression of mobile networks has revolutionized the possibilities of data collection. People carry around smartphones every day and produce vast amounts of data by calling, messaging, browsing the internet, and using different mobile applications. It does not require big investment or complex technology to connect almost any object to the network. Everyday objects can thus collect and share data and communicate with each other. This phenomenon is referred to as the internet of things (IoT) (Xia et al., 2012).

2.3.2 Data technologies

Many companies and organizations are constantly striving to develop new and better ways to store data and tools for analyzing it. The major technologies enabling the big data phenomenon include: in-memory computing, parallel computing, cloud computing, nonrelational databases and stream computing.

In-memory computing has become popular especially in analytics as it enables fast processing of even large data volumes. The basic logic is that the data is accessed from memory (RAM) rather than on hard drive during the analysis. Some in-memory systems also allow adding new data without data preparations. (Garber, 2012) However, in-memory computing requires the data to fit into memory thus analyzing big data in memory requires a large amount of main memory that is quite costly and consumes much energy (Chen et al., 2014b). Thus it is not suitable for analyzing very large data sets.

Parallel computing or **distributed computing** works well with in-memory computing and means that the data is stored and the queries are run on multiple computers simultaneously (Indrawan-Santiago, 2012). This enables the horizontal scaling of computing power using cheap commodity hardware rather than having to have a single supercomputer to process all the data. Parallel computing has existed for a long time already but it has only recently become available to wider community by **cloud computing**. Many organizations that previously did not have access to large computing facilities can now buy services from cloud service providers (Indrawan-Santiago, 2012). Utility computing providers sell the storage space and computing power of their data centers (clouds) using pay-as-you-go pricing (Armbrust et al., 2010). Thus the users get very scalable computing services without having to make a large initial investment or having long-term fixed costs.

Nonrelational databases, frequently referred to as **noSQL databases**, have been around since the 1960s but have only recently gained significant market attention. Companies handling very large amounts of data, like Google and Facebook, have been forerunners in developing and commercializing noSQL technologies.

A relational database is optimal for storing transactional data. It prioritizes data integrity by following so called ACID (atomicity, consistency, isolation, durability) properties. Atomicity means that no update can be incomplete. Update is either performed completely or not performed at all. Consistency refers to the following of database rules. Isolation means that applications run transactions independently of other concurrent applications. Durability means that once an update is completed, it will persist. Commercial relational databases are very developed and offer a wide range of features and are usable also to nonprogrammers. However, as already discussed, relational databases are not well suited for storing varying data without a clear structure or when large amounts of data must be accessed fast. (Levitt, 2010)

Many nonrelational databases are designed for distributed computing. The data structure in nonrelational databases is simpler (Levitt, 2010) and they are not as strict as relational databases about data consistency and ACID properties (Indrawan-Santiago, 2012). These properties make nonrelational databases often faster and able to handle larger datasets than relational databases. Additionally it is easier and faster to handle unstructured data noSQL databases (Indrawan-Santiago, 2012). The three main types of nonrelational databases are key-value stores, column-oriented databases, and document-based stores (Levitt, 2010).

Key-value stores or big hash tables have a very simple structure. They only store pairs of keys and values. Each value has a unique key and the values can be different type and structured or totally unstructured. The simple data model makes key-value stores very efficient and scalable. However, the are only suitable for selected data storage needs. (Pokorny, 2013) According to DB-engines ranking (2015, September) the most popular key-value store database is Redis, which is an open source project (Redis, n.d.). Redis works in-memory (Redis, n.d.), which means that it stores the full dataset in memory rather than on hard drive. This enables much faster analysis than if the data would be retrieved from the hard drive. Redis powers, for example, Instagram's main feed, activity feed, and session system (Instagram, 2012). Figure 3 below illustrates the simple logic of a key-value store.

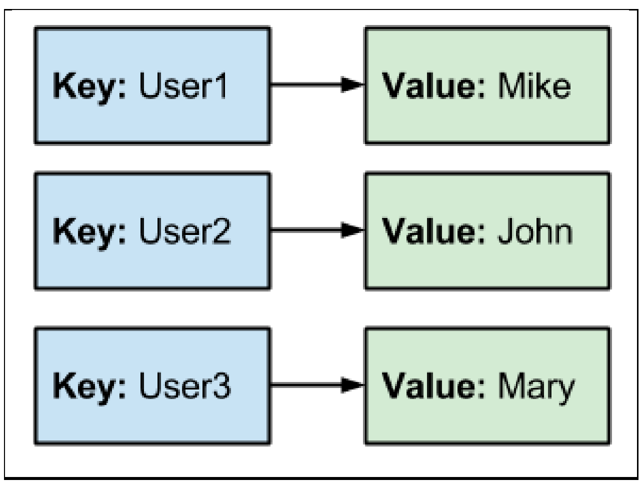


Figure 3. Logic of a key-value store (Shilov, 2014).

Column-oriented databases have a slightly more complicated structure as they store combinations of keys and values collected into columns (Prokorny, 2013). The basic storing unit in a relational database is a record i.e. a row. Correspondingly column-oriented databases store columns containing all the attribute values in a column compressed and densely packed (Abadi et al., 2009). This makes them faster in retrieving subsets of columns. Facebook developed a column-oriented distributed database management system Cassandra to power the service (Leavitt, 2010). The logic of a column oriented database is illustrated on Figure 4.

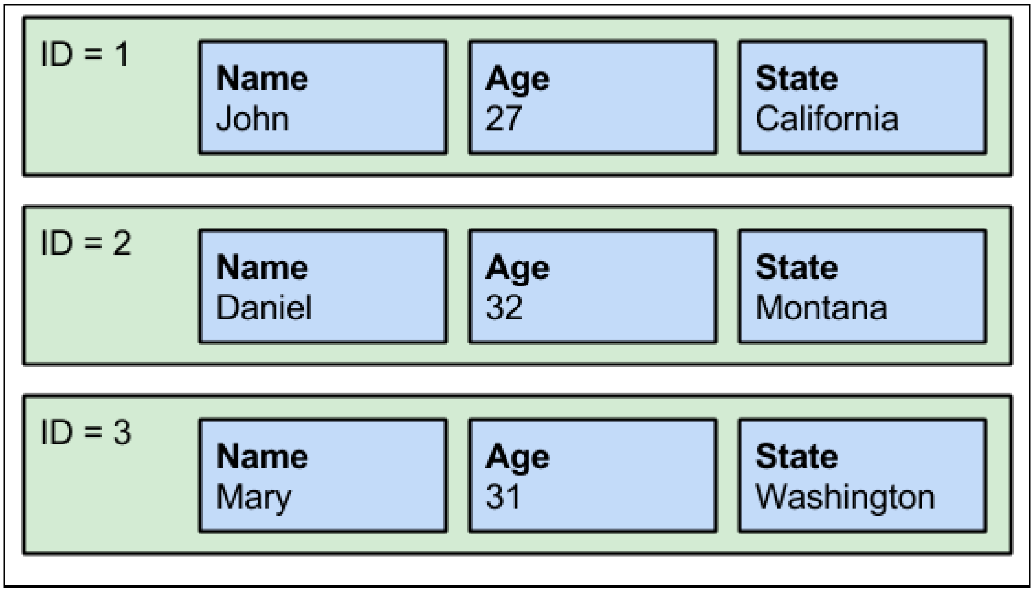


Figure 4. Logic of a collumn-oriented database (Shilov, 2014).

Document based stores store data as collections of documents (Leavitt, 2010). Document based stores set very little limitations to the type of data that can be stored in them. When updating data in document based stores, the user can add any number of fields of any length to a document (Leavitt, 2010). Mongo DB is the most popular document based database management system at the moment (DB-Engines Ranking of Key-value Stores, 2015, September). Figure 5 illustrates the logic of a document based database.

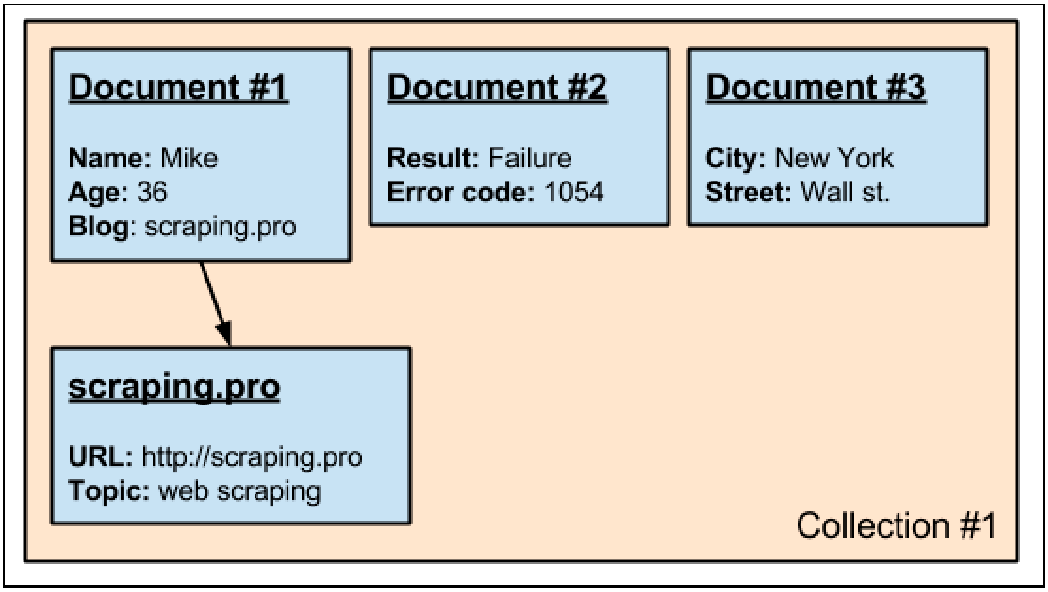


Figure 5. Logic of a document-oriented store (Shilov, 2014).

Stream computing introduces a new logic to data analysis completely different from the traditional approach. It answers the need for real-time processing and analysis of streaming data that is not necessarily stored anywhere in the long term. Traditional databases store relatively static data. The logic is that insertions, updates, and deletions occur less frequently than queries. Queries are typically run one time, when posted and the results reflect the current state of the database. In stream computing the data is handled in sequences (streams). The idea is that new data arrive at unpredictable times and are processed continuously. The long-running queries run for a period of time and return new results as new data arrives. (Golab & Özsu, 2013)

Apache Hadoop, Apache Spark, and Apache Storm are all open source projects and some of the most used big data platforms based on the previously discussed technologies. They are licensed with Apache 2.0 license. It is a permissive license that permits contributors to license their combined or derivative works under any other license (Välimäki, 2005). Derivative works can also be patented but they can not include any royalty requirements (Välimäki, 2005).

Hadoop enables storing and relatively fast analysis of vast amounts of versatile data at a reasonable cost. It was first developed by Google, developer Doug Cutting, and Yahoo! Doug Cutting named the technique after the stuffed elephant toy of his son. Hadoop is based on two major components: Hadoop Distributed File System (HDFS) and MapReduce. The distributed file system is a document based database management system that stored data in a cluster consisting of inexpensive commodity hardware. Files stored in (HDFS) are divided into blocks. Each of these blocks is replicated (three copies is common) and the copies are stored in different nodes of the cluster. The system knows where each of the blocks is located and thus some other copy of a certain block if a node is down or overloaded. MapReduce is a technique to run processing jobs in parallel i.e. utilizing several processing nodes simultaneously. When all the relevant blocks and records needed in the job have been mapped, the results from different nodes are collected and combined to acquire the wanted result i.e. reduced. MapReduce is especially useful for running jobs that require exhaustive examination of records but do not require much interaction between different nodes running the application. (Olson, 2010)

Hadoop is currently the most well-known and widely used and available big data platform. Well known users of Hadoop include, among many others, Amazon, Facebook, and Adobe (Hadoop wiki – Powered by, 2015, August 6). As Finnish Ministry of Transport and Communications (2013) points out, using Hadoop as such is not appealing to many organizations due to the amount of development work, adjusting of settings, integration, and testing involved. Companies often find it more practical to adopt a readily available distribution of Hadoop or use it as a part of a larger analytics system.

Spark is a general engine for parallel data processing first introduced by Zaharia et al. (2010). Spark is like an improved version of MapReduce that can be run on Hadoop or multiple other distributed file systems (Spark - Lightning-fast cluster computing). The main difference between MapReduce and Spark is the introduction of resilient distributed datasets (RDDs). While MapReduce can only handle acyclic data flow and the data remains on the hard drive, the use of RDDs in Spark enables in memory computing and repeated operations on the same data. Many machine learning algorithms, for example, include iterative jobs where a function is applied repeatedly to optimize a parameter. Each of the iterations can be expressed as a MapReduce job but the data has to be reloaded from the disc for each operation, which takes time. MapReduce is also not optimal for interactively running ad-hoc exploratory queries. RDDs are read-only collections of objects that are partitioned across machines. They can easily be rebuilt if a partition is lost, which makes Spark very fault tolerant. RDDs can be cached in memory across nodes and reused in multiple parallel operations. (Zaharia et al., 2010) Spark is claimed to run up to 100 times faster than Hadoop MapReduce in memory (Spark - Lightning-fast cluster computing). Spark is utilized, for example, by eBay and Yandex (Powered by Spark, 2015, September 8).

Storm has been described as Hadoop of streaming data. It is a platform for near to real time parallel data processing (Apache Storm, n.d.). A storm cluster is similar to a Hadoop cluster but while on Hadoop the user runs MapReduce jobs, on Storm the user runs topologies. MapReduce jobs are run from start to finish mainly once at a time but topologies process messages continuously until it is terminated. The basic data entity in Storm is a stream. A stream consists of a sequence of tulpes. Tulpes can be objects of any type. A topology uses spouts and bolts to process streams and transform them into new streams. A sopout is a source of streams. It may for example connect to Twitter API and emit a stream of tweets. Bolts consume input streams, process them and possibly emit new streams. They can for example join streams, aggregate them, or run functions. (Processing real-time events with Apache Storm, n.d.) Storm is used, for instance, by Groupon and Twitter (Companies using Apache Storm, n.d.).

2.4 Possibilities for Different Industry Sectors

2.4.1 Manufacturing

It is easy to see how the harnessing of big data is at the very core of the business model of many "born digital" companies like Google or Facebook. Manufacturing, on the other hand, is often seen as traditional with well-established domains and techniques. Oxford economics surveyed 363 business executives on their views about digitalization. The results show that manufacturing is not seen as a sector that will be drastically transformed by information technology (Oxford Economics, 2011).

However, we can find that in reality big data opens diverse and wide opportunities of improving manufacturing. Li et al. (2015) note that optimization manufacturing processes is vital to manufacturers in the current globalized economy. Optimization can be done by analyzing various data. RFID technologies and wireless sensor network technologies provide the means of acquiring such data of almost any manufacturing phase or domain of interest.

Lee et al. (2013) discuss and map opportunities in utilizing big data and predictive analytics in manufacturing. They suggest that analytics provide solutions for many critical manufacturing challenges. For example, manufacturing equipment can be closely monitored through sensors. This data can be used to identify and even predict failures, estimate health, degradation level and remaining useful life of the equipment. Predictions made of the sensor data enable just in time maintenance. Outside the shop floor analytics can be harnessed, for example, to manage supply chains, product life cycle, and relationships to customers. Figure

6 below illustrates the possibilities of utilizing analytics in manufacturing as suggested by Lee et al. (2013).

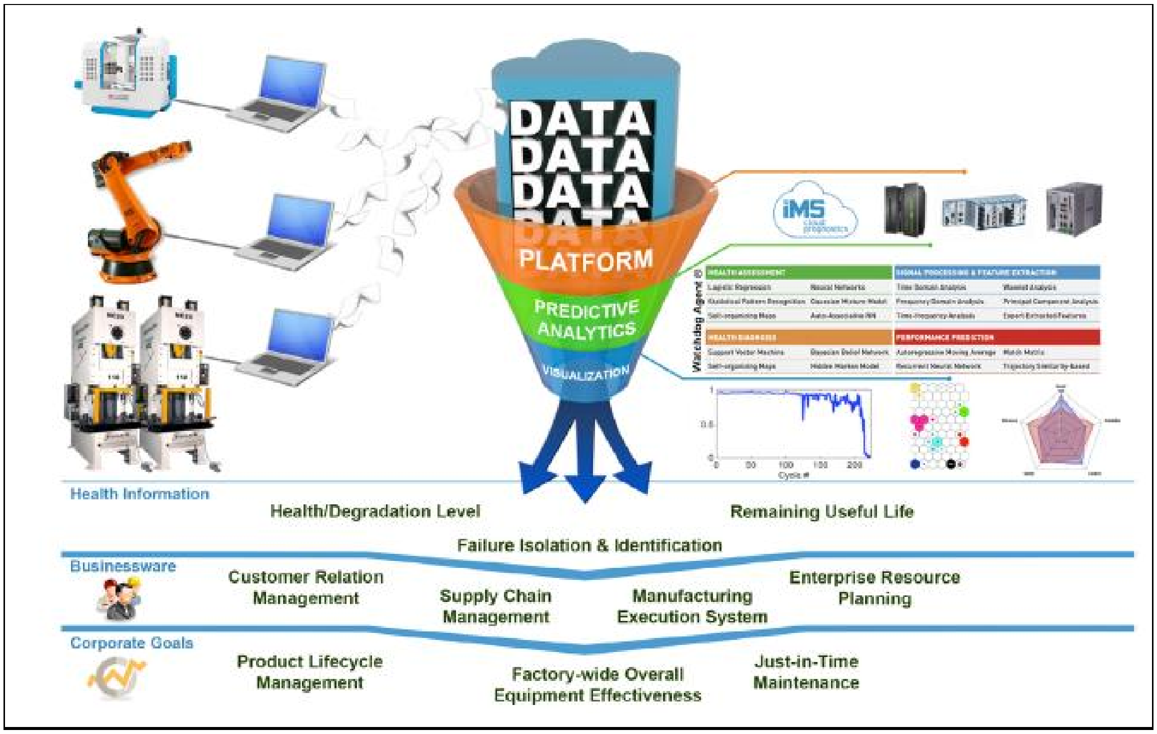


Figure 6. Utilizing big data in manufacturing (Lee et al., 2013).

Dutta and Bose (2015) list some examples of successful big data initiatives in manufacturing companies: Merck is using big data to develop vaccines faster, Volvo has a system for forecasting component failures, and Xerox is utilizing telemetric data from their customer service to improve service quality and reduce costs. Li et al. (2015) suggest scientific workflow technology to optimize shop floor workflow and scheduling.

Dutta and Bose (2015) describe in detail how a cement producer Ramaco Cements Limited implemented big data technologies to build three different analytics applications. Ramaco Geoapps uses Google map interface to capture the locations of their customers and cement outlets and link the location data to critical organizational information in the ERP database. Dutta and Bose (2015) state that this set up could provide insight of potential locations to expand the business, what kind of customers are taking their business elsewhere and why, how the sales stuff is performing and how the performance could be improved, what are the routes of the company's cement dispatches, or what the company's competitors are doing. Ramco Perform is a dashboard illustratively depicting the KPIs and achievements of the dealers. Ramaco APO used real-time analytics to plan and optimize outbound logistics, which is a major cost driver in cement business.

2.4.2 Energy production

Smart grid is the buzzword of the day in Energy production. Smart grid refers to “the next generation power grid constituted by traditional energy networks integrated with computation, communications and control for optimized generation, supply, and consumption of electric energy” (Chen et al., 2014a). As power networks typically comprehensively cover very large geographical areas, they provide interesting means of collecting data related to electricity production and consumption. Before smart electric meters measuring electricity consumption was labor intensive and it was not possible to retain real-time data. Combining detailed and timely records of power consumption with other data sources like data of electricity prices or weather provides opportunities, for example, to predict and optimize power supply and demand, gird maintenance, and electricity prices. (Chen et al., 2014a)

2.4.3 Retail

The retail market is currently in transformation. Digitalization has opened possibilities to collect massive amounts of data to optimize operations and marketing. Big global e-tail companies like eBay and Amazon are intensively utilizing data and challenging the traditional retail stores.

Pousttchi and Huffenbach (2014) note that utilization of customer data enables retailers to recognize customers’ needs in advance, make tailored offers, and give instructions in where to buy. They claim that if retailers would use data-driven personalized marketing, response rates could be increased five- to tenfold and sales by 20 – 30 percent. Tools of data-driven personalized marketing include, among others, the utilization of ratings, the analysis of items frequently bought together, and location-based marketing. These can be used to support cross- and up-selling and to provide the customer with the right offer at the right time. Combining and exploring product data, customer data, and data of campaigns and offers can be analyzed to understand buying behavior. Product ratings, service ratings, and product reviews given by customers provide valuable information to improve customer service and satisfaction. Data can also give insight to optimize pricing for maximum profit. (Pousttchi & Huffenbach, 2014)

Predictive analytics and machine-learning can also help to optimize physical operations and supply chain. Amazon patented a system that predicts the order before the customer actually places it (Bensinger, 2014). The system potentially enables Amazon to cut delivery times and to better optimize stock-levels. Being able to accurately and quickly predict trends helps to anticipate demand for different products and thus adjust the supply. This is especially important to apparel retailers that are highly influenced by trends.

2.4.4 ICT

Big data as a phenomenon creates vast possibilities for ICT -sector as the carrier of big data (Lu et al., 2013). Utilization of IoT and big data increase the demand for data collecting sensors and wireless networks carrying the data. Big data is also at the core of ICT capacity optimization and ensuring reliable and secure ICT services for the users. In addition, ICT companies can create added value to their customers by providing data and analytics based services.

2.4.5 Banking

The transactions today are mostly digital. From this it follows that banks are collecting and storing huge amounts of transactional data. By utilizing big data analytics capabilities valuable information to improve business can be attained from this data. Srivastava and Gopalkrishnan (2015) consider the possibilities of data utilization for banks. They recognize three different domains of possibilities:

1) Customer centric

Customer centric approaches include marketing and sales related tools. By collecting and analyzing customer feedback data such as customer ratings, banks can improve customer experience. Sentiment analysis provides insight into strategy management and sales forecasting. It can also help to manage leads and referrals more effectively. Data can analyzed to assess the quality of leads. Customer life events can be analyzed to find the best offers for customers, find the most profitable customers, and improve the customer lifetime value. Data can be used to micro segment offerings or find next best offers for the customers.

2) Risk management

Transactions can be analyzed to find abnormal patterns that could be caused by criminal activity like money laundering or unauthorized credit card usage.

3) Transactions

Transactions provide valuable insight into the nature of trade, future capital flows, and market sentiment. This information can be utilized to manage investments and loans better and support market analysis.

2.4.6 Logistics

Especially the opportunities of utilizing big data in supply chain management have gained considerable interest in scientific discussion during the past couple of years. Frehe et al. (2014) conducted a systematic literature review of papers addressing the implications of big data to different domains of logistics. They mention 14 different cases of big data initiatives in logistics companies. The applications include the optimization of routes, driving times, and fuel consumption to lower costs and CO^2 emissions, optimization of warehousing activities, collecting and aggregating data for reports such as CO^2 emission reports, providing data about transportation and warehousing and analysis tools for the customers as an added-value service or support pricing negotiations, and supporting risk analysis.

2.4.7 Entertainment

Spotify and Netflix are examples of entertainment companies that use data and analytics extensively. Spotify collects and analyzes user data to offer targeted advertising services. They, for example, provide the user every week with a playlist that includes new music that may suit the user’s preferences based on their previous music consumption habits. The service also suggests the user with similar tracks to the one the user is listening. Netflix provides the user with suggested movies and series based on the user’s profile and previous activity on Netflix.

2.4.8 Health

There are multiple reasons why healthcare is an especially interesting field in terms of possibilities of big data. Healthcare is a labor intensive industry and the workers are required to have extensive education, and thus the salaries are typically relatively high. Healthcare is often offered by the government as a public service to the citizens and is thus a major public expenditure. Given the current economic condition and the huge national dept of many countries, there is a need to increase the productivity and efficiency of healthcare. Big data and analytics provide possibilities to automate healthcare functions in a way that was previously not possible.

Because of the labor intensiveness, it is also hard to standardize healthcare, improve its quality, and ensure that all the patients receive high quality treatment. Human error is the major source of risk and the stakes are high as often the life or future life-quality of the patient is at stake. Automating some functions and providing support to clinical decisions could help to standardize services and lower risks.

Healthcare organizations have for some time collected comprehensive digital records of the patients. The already existing data provides huge utilization potential. However, this data is mainly unstructured, so it is difficult to analyze using traditional methods. Big data technologies provide a solution to the problem.

People are also becoming increasingly interested in their own health. They want to be more in control of their health as opposed to letting the doctors make the decisions. This results in huge market potential of health-related products that help people to track, analyze, and understand their health.

Murdoch and Detsky (2013) identify four ways how big data can help to improve healthcare quality and efficiency. Firstly, health-related knowledge can be generated more efficiently as the analysis of even unstructured data can be automated. Secondly, systems can be developed to make relevant information available to the healthcare personnel at the right time. Murdoch and Detsky (2013) note that Memorial Sloan-Kettering Cancer Center already uses IBM's Watson supercomputer to support diagnosis and selection of appropriate treatment. Third, more generalized data can be constructed by analyzing individual patient cases in the health record systems. Fourth, automated data-analysis can be used to give information directly to the patients, which enables them to better understand and take charge of their healthcare.

One of the most famous applications of big data to healthcare is the case of Google flu. To control the spreading of contagious diseases like influenza it is vital to have timely information about the spreading of the epidemic. Ginsberg et al. (2009) found that certain Google search queries are highly correlated with the number of physician visits in which a patient presents with influenza-like symptoms. Using Google search data Ginsberg et al. (2009) were able to accurately estimate the level of weekly influenza activity in each region of the United States, with a reporting lag of one day. The traditional methods for epidemic tracking typically have reporting lags of 1–2 weeks.

2.4.9 Education

As a result of the current financial state of many governments there is a need to improve the efficiency and productivity of education. Digitalization and data analytics when applied in a purposeful way help to achieve these goals.

Picciano (2012) examines the current state and possibilities of big data and analytics in American higher education. However, the same methods could largely be applied also to primary education. He describes that in higher education, analytics are used to address student performance, outcomes, and persistence and suggests that institutes data could be collected for each student transaction in a course, especially if the course was delivered online. These transactions include accesses to reading material, completions of assessments, and postings on a discussion board among others. This data could then be analyzed to recommend appropriate courses of action. Picciano (2012) also points out that big data can be applied to the administration of educational institutes. It can be utilized, for example, in admissions processing, financial planning, donor tracking, and student performance monitoring.

For example, Rio Saldo Community College in Arizona uses a learning analytics application that can predict, after the first week of a course, with 70 percent accuracy, whether a student will complete the course. Application of Northern Arizona University alerts students regarding grades, attendance, academic issues, and positive feedback. The students are provided with relevant options and resources depending on the nature of the alert. Purdue University has a Course Signals System that detects the early warning signs of a student not performing to the best of his abilities. Ball State University has developed a system to support collaborative knowledge-building. (Picciano, 2012)

2.4.10 Politics

Following the citizens' views on politics by conducting opinion polls is a vital tool for political parties in directing their communications efforts. Analysis of big data opens up a possibility to understand voters' behaviors in much more detail than before. Issenberg (2013) describes how Obama utilized big data heavily in both of his presidential campaigns. Obama's team build a database containing information about each individual voter in the country. The information contained data about all the interactions of the individual with the campaign, the results of opinion polls, data from commercial consumer data warehouses, media consumption data, and demographics. The team was able to build a model that predicted the outcome of the election very accurately. Also they were able to identify which voters were most likely to respond to different kinds of messages: which ones could be persuaded to vote, which ones to volunteer, and which ones to donate. Issenberg (2013) describes how the team used A/B -testing to find the most effective marketing message and channel to each individual. The team was also able to predict which voters were about not to vote at all and how to best mobilize them. By utilizing advanced statistical models, they were able to identify the next most likely citizen to respond to the marketing message and optimally focus the campaigning efforts. The data also reviled how different events (like the bankruptcy of Lehman Brothers) taking place during the campaign changed the voters' opinions and behaviors.

2.4.11 Urban planning

Zheng et al. (2014) studied how big data is currently utilized in urban planning and how the future intelligent city could optimize its operations based on data. They note that various kinds of data are already collected or could be collected of the cities. Some of the different data-types and sources include traffic data, weather data, data of air quality and noise levels, data of locations of people and places, land-usage data, data of public transportation usage, economic data, health data, and social media content. Data from various sources could be analyzed to optimize different city functions like planning and building of traffic and public transport connections, planning of land-usage, making economic decisions, or finding optimal borders for independent administrative areas. Also data could help to save energy, reduce pollution, or prevent accidents and crime. Links between the health of the citizens and different factors, like air quality, could be studied.

2.5 Challenges in Applying Big Data

2.5.1 Ethical issues

The most discussed ethical concern related to big data is privacy. Hull (2015) points out the problems with privacy legislation that is based on privacy self-management. He claims that people do not have the sufficient information and understanding about how their data can be used to either benefit or harm them and what is the potential value in utilizing their data when they read privacy notices and give their consent. Also if the data is sold to third parties, as for example Facebook frequently does, the data is handled according to possibly entirely different third party policies. He gives examples of how Facebook can use data of user's likes and friends to predict unexpected things with incredible accuracy. For instance, Facebook data implied that users that liked Hello Kitty were more likely to be emotionally instable than an average user. (Hull, 2015)

Boyd and Crawford (2012) also point out how social media data is utilized for purposes that the users could not imagine. No-one usually asks if they want their data to be analyzed in these studies. They refer to the case where a group of researchers gathered data from Facebook profiles of 1 500 college students. None of these students were aware that their data was being used. This data-set was made available to other researchers. Soon it was found that it was possible to deanonymize parts of the data. The case raised considerable discussion about the ethicality of using social media data in research. Boyd and Crawford (2012) state that the fact that the data is available does not justify using it for what-ever purposes without the users' permission. They underline that users of data should recognize the difference between being in public (for example sitting in a park) and being public (actively courting attention).

2.5.2 Methodological considerations

Boyd and Crawford (2012) criticize the tendency to treat big data as whole data and forget to assess the possible bias in the sample. They use Twitter data as an example and remind the users of social data that Twitter users is not synonymous with all people and Twitter accounts and Twitter users are not necessarily the same thing. Twitter users are a very specific group of people. Some Twitter users might have several accounts, several people might use the same account, or some people might cross-use each other's accounts. Also just a fraction of Twitter users are actively posting content. Twitter makes only a subset of public Tweets available through its APIs. Hardly anyone knows how the algorithms choose the Tweets made available. To conclude, Boyd and Crawford (2012) state that the size of the dataset is meaningless if the sample is skewed right from the beginning. (Boyd & Crawford, 2012)

Humberty (2015) argues that when discussing the possibilities of big data four false assumptions are frequently made: 1. By using big data we get rid of the samples and can study entire populations. 2. We know which sources of data must be examined to understand entire populations. 3. Behavior of people online closely reflects their behavior offline. 4. We can predict the future based on historical data. He points out that not all of the people use online services and even if they did, the digital world is changing so quickly that we don't know if they are still using the same services as they used yesterday. It should be taken into account that user bases of different online services are usually relatively homogenous subsets of all the people. Humberty (2015) also says that there is research suggesting that people's behavior online poorly reflects their offline behavior. Finally, he questions the plausibility of using historical data to predict the future. People change and the way they use online services change. For example, the language people use to describe the world is evolving constantly. (Humberty, 2015)

2.5.3 Management challenges

McAfee and Brynjolfsson (2012) identify five management challenges imposed by big data. First of all, there is a need for skilled leaders that can identify the opportunities of big data to the company, create clear visions, effectively communicate those visions to others and manage the process of transforming the visions into reality. The second challenge they mention is talent management. It is rare to find a person that is both skilled in handling very large data sets and able to translate the business needs into problems that big data can answer. Thirdly, the companies IT departments are facing new challenges with the big data technologies they are not usually readily familiar with. It is an organizational challenge to bring the people who understand the business problems, the right data, and the people that can handle the data and technologies together. Finally, it is a challenge to change the organizational culture from making decisions based on a hunch or instinct to making decisions based on data. According to McAfee and Brynjolfsson (2012) rather than asking itself: "What do we think?" a data driven organization asks: "What do we know?" (McAffee and Brynjolfsson, 2012).

3 Theoretical Framework

This section provides the theoretical background for the empirical part of the study. First the theoretical conceptualization of business intelligence is introduced and the concepts of maturity and success are adopted. Next models describing big data maturity and BI implementations success are introduced. The research questions of the study stem from these two models and they have been used as a basis in forming the interview questions of the case study.

3.1 Theoretical Conceptualization of Business Intelligence

Big data is still a relatively new research area and no comprehensive scientific framework has been introduced to describe the phenomenon. On the other hand, big data was earlier defined as a new paradigm in business intelligence (BI) and analytics, and BI has been studied comprehensively and multiple frameworks have been developed to describe the development of a company's BI maturity.

Lahrmann et al. (2011) define BI based on sosio-technical theory. They claim that BI is most often understood solely as an IT artifact and call for a more holistic view. They argue that also people with their capabilities and organizations with their practices should be considered. BI consists of interrelated technical and social systems.

DeLone and McLean (1992) introduced a seminal work on defining IS success. The model of DeLone and McLean (1992) divides information systems success into six interdependent categories, namely 1. system quality 2. information quality 3. use 4. user satisfaction 5. individual impact and 6. organizational impact. Figure 7 illustrates the framework. The arrows represent causal relationships between the different categories of success as suggested by DeLone and McLean (1992).

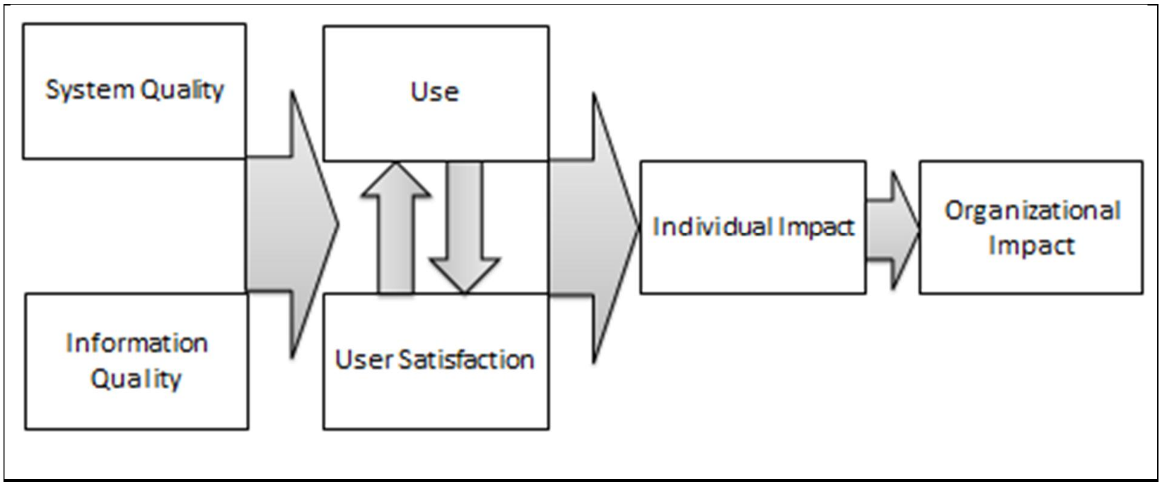


Figure 7. Information systems success (DeLone & McLean, 1992)

System quality measures relate to the processing system itself. Factors such as system efficiency, data accuracy, and query response time belong to this category. Information quality measures focus on the quality of system output. These measures can include, among others, timeliness, decision relevance, readability, informativeness, or precision. (DeLone & McLean, 1992)

Use measures describe how the recipient utilizes the output of the information system. Examples of use measures are amount of use, number of queries, and number of functions used. User satisfaction refers to the recipient response to the use of the output of the information system. User satisfaction measures introduced in the framework cover overall satisfaction, information satisfaction, software satisfaction, and decision-making satisfaction. (DeLeone & McLean, 1992)

Individual impact is defined by the effect of information on the behaviour of the recipient. The authors allocate measures like information understanding, information awareness, and effectiveness of decision-making into individual impact group. Finally, organizational impact represents the effect of information on organizational performance. Organizational impact can be measured, for example, by measuring operating cost reductions, increased revenue, or increased market-share. (DeLeone & McLean, 1992)

Several different success factors models for IS implementations have been developed based on the conceptualization of DeLeone and McLean (1992). Lahrmann et al. (2011) summarize the main idea of the different models. The core structure in such models is that t deployment of IS influences the use of IS on an organizational and individual level. This in turn leads to the organizational and individual impact of IS.

Based on this leading logic, Lahrmann et al. (2011) introduce their conceptualization of BI maturity. The framework consists of the interrelated concepts of “deployment”, “use”, and “impact”. Figure 8 illustrates the framework. Lahrmann et al. (2011) state that improved decision making is the main BI impact on the level of an individual. Improved decision making translates to better overall organizational performance. Impact is a consequence of use, which can be seen from the individual point of view (e.g. ease of use, efficiency, and effectivity) or from the organizational point of view (e.g. covered business topics and spread throughout the organization). Deployment has a direct impact on use. It includes the IT artefact – the architecture and applications, the capabilities that consist of team skills and competence of the staff, and finally the practices that are formed in development, operations and management processes related to BI.

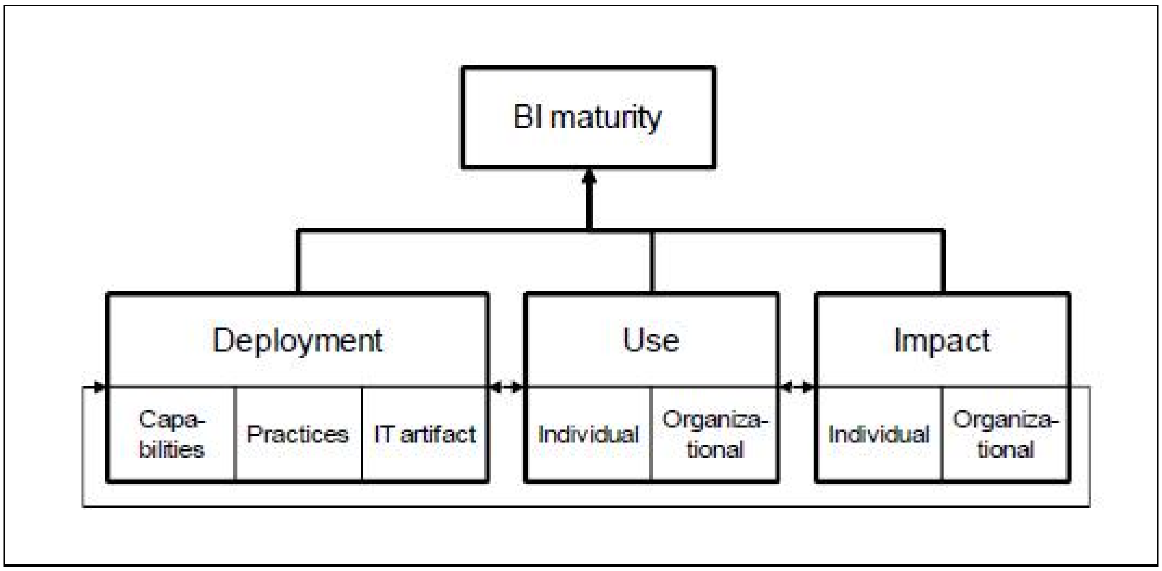


Figure 8. Theoretical conseptualization of BI maturity (Lahrmann et al., 2011)

For the purposes of this research, we apply the concepts of maturity and success in the context of big data. Two more detailed and practice oriented models of these are introduced.

3.2 Big Data Maturity Model

Lahrmann et al. (2011) note that BI requires a comprehensive overall view of the design and changing of its structures in order to be effective. BI maturity models serve this purpose by providing a framework for assessing BI in a particular company. Lahrmann et al. (2011) describe how the maturity models measure the state and performance of a company in several dimensions (exhaustive and distinctive specified capability areas, process areas or design objects structuring the field of interest). Maturity models define levels – archetypal states of maturity – for each dimension of interest.

Maturity is defined as a state of being complete, perfect or ready and argue that maturity has to reflect causes, such as BI technology employed, as well as effects, for example the realized organizational impact of BI (Lahrmann et al., 2011).

Yeoh and Koronios (2010) see BI implementation as a cycle and iterative process of continuous improvement. The same can be applied to the big data utilization of a company. Typically, the implementation starts with experimenting within clearly restricted area and small-scale applications. Finally big data can transform and define the whole business mode of the company.

El-Darwiche et al. (World Economic Forum, 2014) introduce a model for assessing big data maturity of a company. Figure 9 illustrates the four maturity stages included in the model and introduces some of the most popular applications used by the companies in each stage.

	Description	Typical use cases / applications
Stage 1: Performance management	Enables executives to view their own business more clearly through, for example, user-friendly management information dashboards. This stage typically relies on internal data, with an organization establishing key performance indicators (KPIs) to evaluate its success at achieving stated goals.	- Financial reporting - Regulatory/compliance reporting - Dashboards for management reporting - Performance measurement via key performance indicators/metrics
Stage 2: Functional area excellence	Organizations start to experiment with internal and external data to improve selected facets of their business. This may involve sales and marketing techniques such as customer segmentation and targeting, or entry-stage analytical methods for product recommendations. Advances in operational efficiency through big data, such as the efficient deployment of staff resources and the optimization of the supply chain, also reside within this maturity stage.	- Smart pricing - Targeted mailings - Customer segmentation - Customer value analysis - Choice analysis - Website clickstream analysis - Loyalty schemes - Customer satisfaction opt. - Supply chain balancing - Facility optimization - Optimization of staff utilization - Route optimization for fleet - Fraud monitoring/prevention - Patient treatment control - Traffic optimization - Crime monitoring
Stage 3: Value proposition enhancement	Organizations start to monetize big data, positioning it as a value driver of the business that offers a new source of competitive advantage beyond the mere improvement of operations or services. In many instances this involves obtaining data from external sources and deriving insights from it. This may include innovations such as customized, real-time recommendations or the personalization of services to augment the customer experience.	- Targeted advertising/customized recommendations in real time - Preventive health monitoring &disease detection - Data monetization - Online telematics services - Personalization of customer experience/products
Stage 4: Business model transformation	Big data permeates the whole organization. It becomes deeply embedded within the operation, determining the nature of the business and the mode of executive decision-making.	- Selling of data to open new revenue pools - Data-centric business models (e.g., web search, web advertising) - Quantitative management of investment funds - Crowdsourcing to augment internal data

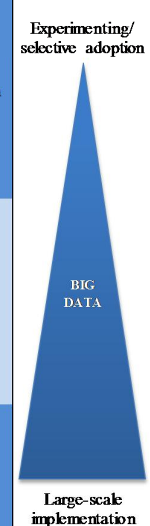


Figure 9. Big data maturity model (World Economic Forum, 2014)

The first stage of big data maturity is **performance management**. At this stage mainly internal business data is utilized to follow the performance of the business through key performance indicators. The first stage can't really be considered as big data utilization yet. Various dashboards can be developed to enable quick access to the most recent data and clear visual representation of it. (World Economic Forum, 2014)

At the second stage, **functional area excellence**, the focus is on improving performance. Internal and external data are used in experimentative and selective fashion to improve different business functions. The focus is on improving operational efficiency by for example optimizing the supply chain or the use of staff resources. (World Economic Forum, 2014)

Third stage, **value proposition enhancement**, is achieved when big data becomes a value driver and a major source of competitive advantage. Also external data is utilized and big data applications can be a major component in the product or service of the company. Revenues are generated by the use big data applications and the focus is not solely on reducing costs and improving efficiency. (World Economic Forum, 2014)

In the final stage, **business model transformation**, big data is a defining factor for the whole organization and is present in all aspects and stages of doing business. Big data also defines the decision-making processes of the organization. (World Economic Forum, 2014)

The model also considers what defines the possibilities of a company in big data utilization. The different aspects defining the big data potential of a company can be divided into external factors and internal competences (World Economic Forum, 2014). Both of these categories include factors that enable the companies better utilize big data. Lack of some external enablers or internal competences can, in turn, prevent companies from effectively harnessing big data. For this study the tool of SWOT-analysis is adopted to identify and organize these factors.

SWOT-analysis is a well-known and established tool for strategy development. SWOT-analysis is used to identify the weaknesses and strengths of the organization as well as the opportunities and threats of the environment. The internal factors of the organization, such as personnel, facilities, location, products, and services, are considered in the assessment of the weaknesses and strengths. Opportunities and threats are identified by examining organization's external political, social, technological, economic, and competitive environment. (Dyson, 2002).

3.3 Success Factors for BI Implementations

While Lahrmann et al. (2011) base their conceptualization of BI maturity on models describing information systems success in general, Yeoh and Koronios (2010) introduce a framework considering especially business intelligence implementations success. They argue that from the set of measures suggested by DeLeone and McLean (1992) system quality, information quality, and system use measures are the most appropriate to use in BI systems context. They combine these three variable groups under label infrastructure performance. In addition they introduce process performance measures that include budgeting and scheduling concerns.

Yeoh and Koronios (2010) used three-phased Delphi study to identify critical success factors for the success of BI implementations. They also conducted five case studies to verify the findings of the Delphi study. Seven critical success factors were found in the Delphi stage. The authors divided the factors into three dimensions: process dimension, organizational dimension, and technological dimension. Yeoh and Koronios (2010) also recognized that all the success factors suggested by the experts in the Delphi group indicate strong business orientation in the implementation. Figure 10 illustrates the framework.

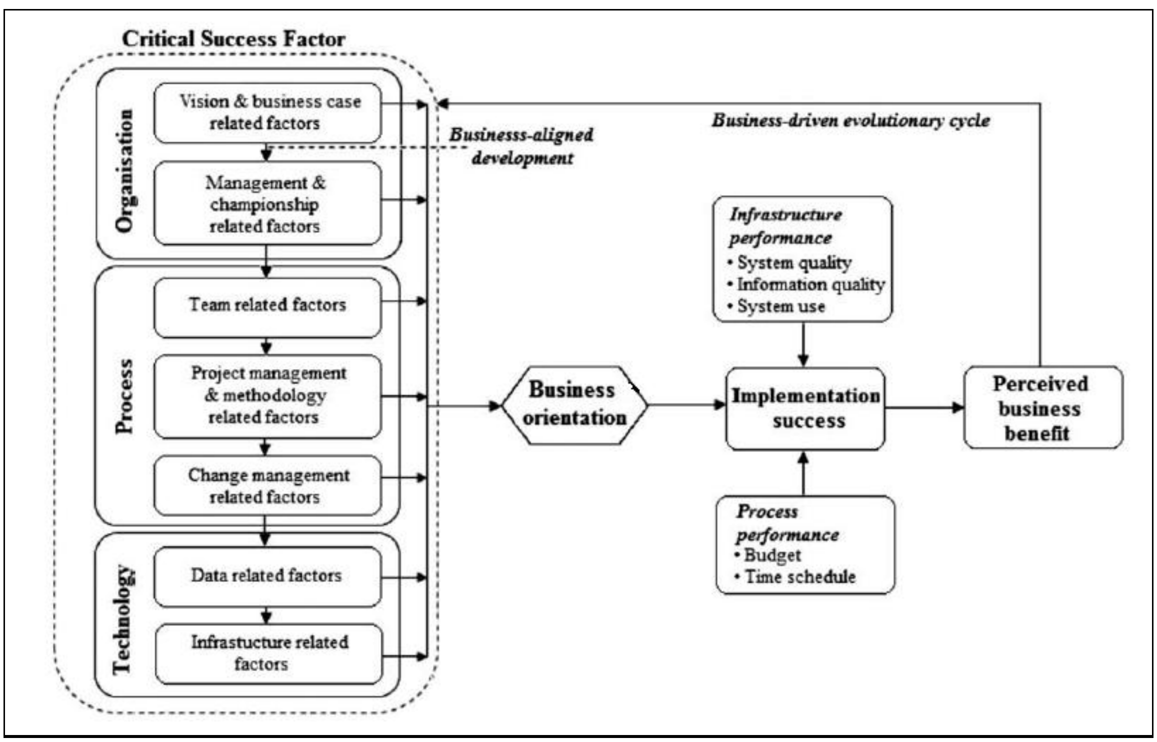


Figure 10. Critical success factors for BI implementations (Yeoh & Koronios, 2010)

The first critical success factor in organizational dimension by Yeoh and Koronios (2010) **is committed management support and sponsorship**. According to the authors, this has been widely acknowledged as the most important factor in BI system implementation. Another CSF in organizational dimension is **clear vision and well-established business case**. The implementation must be business driven and the business case must be aligned with the overall business strategy and objectives.

CSFs in process dimension include **business-centric championship and balanced team composition**, **business-driven and iterative development approach**, and **user-oriented change management**. The implementation project champion should have profound understanding of the business needs and objectives and he/she should be able to translate these goals into IT architecture. The team needs to be cross-functional and consist of both business and IT experts. The scope of the project and the project plan should be clearly defined and follow the business objectives. According to the Delphi-group members iterative approach to implementation consisting of a continuum of small steps includes less risk and is thus less likely to fail than one time "big bang" implementation. User-oriented change management implies that the end-users of the system should be involved throughout the implementation process. The users know better what they need than IT architects and developers. (Yeoh & Koronios, 2010)

Technological dimension includes two CSF: **business-driven, scalable and flexible technical framework and sustainable data quality and integrity**. The dynamic business environment should be considered comprehensively in designing the architecture. Often the systems grow larger than expected or the throughput is greater than anticipated. The technology should be scalable and flexible to accommodate increased data volume, additional data sources and attributes. Data quality and integrity are crucial concerns as unreliable data leads to poor decisions. (Yeoh & Koronios, 2010)

Yeoh and Koronios (2010) see BI implementation as a cyclical and iterative continuous process of evaluation, modification, optimization, and improvement as opposed to one-time undertaking. They also argue that the CSFs are necessary for implementation success and that the lack of any of them will cause the implementation to fail.

While this framework doesn't directly address big data endeavours but BI implementations, big data can be seen as a new phase in the development of BI and analytics (Chen et al., 2012). Thus the framework can be seen as a useful starting point in understanding the success of big data implementations. For example, Bole et al. (2015) already generalized the findings of Yeoh and Koronios (2010) to data mining projects. In this study, the organizational impact of big data implementations is of concern. It attempts to assess the success of the initiatives from an organizational perspective and consider the factors behind the success / failure.

4 Big Data in Finland

This section introduces the findings of previous research about big data in Finland. First, the current state of big data in Finland is discussed. Then the opportunities that big data offers to Finnish industries are considered. Finally, some challenges in applying big data in the Finnish environment are discussed.

4.1 Current state

The Finnish Ministry of Transport and Communications (Liikenne- ja viestintäministeriö (2013) and Accenture (2014) both studied big data in Finnish organizations. The survey of The Finnish Ministry of Transport and Communications was internet based and consultants, big data service providers, public sector organizations, educational institutes, and private companies were surveyed separately. In addition, a few interviews were conducted (Liikenne- ja viestintäministeriö, 2013). Accenture (2014) interviewed over 130 managers of the 40 biggest public and private organizations in Finland. Both of the studies draw the same main conclusion: Finnish organizations widely recognize the importance of big data as a major driver of transformation in many industry sectors and as a factor determining the competitiveness of businesses in the future. However, the impact of big data on the industries is still quite low, some people find the concept of big data difficult to grasp, and the organizations still rarely have a concrete vision, how they could utilize big data.

According to The Finnish Ministry of Transport and Communications big data is still at a very immature stage in Finland and organizations are currently mapping the possibilities and developing strategies concerning big data (Liikenne- ja viestintäministeriö, 2013). 38 % of the respondents evaluated that big data will be at the core of business of their industry in the future (Liikenne- ja viestintäministeriö, 2013). 24 % believed that big data will be a main driver of new business (Liikenne- ja viestintäministeriö, 2013). Organizations interviewed by Accenture (2014) believe that big data will start to change industries significantly in the next three years. The most big data mature industries seem to be electronics, media, and retail. 82 % of the interviewees from the electronics industry say that the impact of big data is already visible in their business (Accenture, 2014). The corresponding percentages for media and retail sectors are 71 % and 61 % (Accenture, 2014).

In the Nordic Hadoop survey conducted by Intel and SAS Institute (2015) 95 % of the Finnish respondents saw an increasing need to collect and analyze data in their company and

76 % recognized a need to collect new types of data that do not fit into traditional databases. 90 % of the respondents believed that analysis of more and new data would give their company competitive advantage. The shares of respondents (all Nordic countries) from different industries that answered yes to all three questions are illustrated on Figure 11.

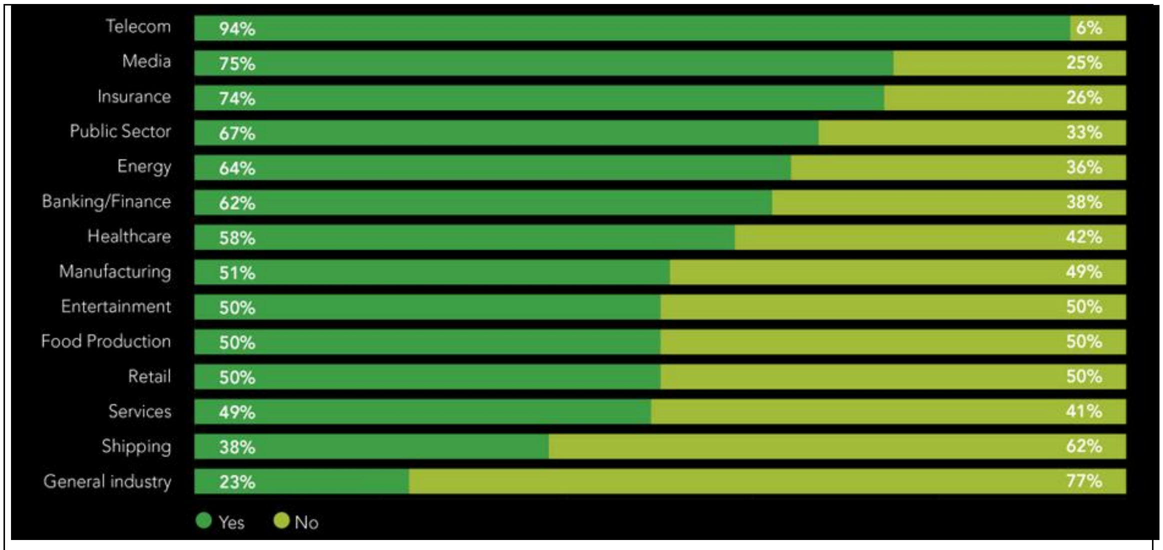


Figure 11. The shares of respondents that see an increasing need to collect and analyze data, see the need for new data storage methods, and believe that analysis of more and new data would give their company a competitive advantage by industry sector (Intel & SAS Institute, 2015).

It seems that the need for more sophisticated data storage and analysis methods is most recognized by the Telecom industry with 94 % of the respondents answering yes to all three questions (Intel & SAS Institute, 2015). However, the shares are 50 % or higher for almost all industries (Intel & SAS Institute, 2015).

The Finnish Ministry of Transport and Communications (Liikenne- ja viestintäministeriö, 2013) asked Finnish organizations about their current most important data storage solutions. The results are illustrated on Figure 12.

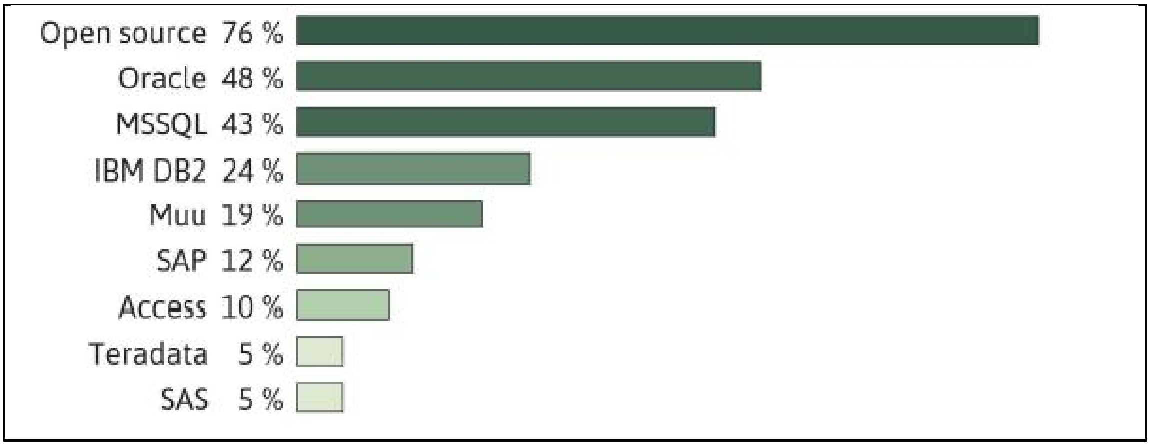


Figure 12. Most central data storage solutions in Finnish organizations (Liikenne- ja viestintäministeriö., 2013).

The most often mentioned (76 % of the respondents) data storage solution was open source followed by Oracle and MSSQL, which were mentioned by 48 % and 43 % of the respondents (Liikenne- ja viestintäministeriö, 2013). 63 % of the Finnish respondents of Nordic Hadoop Survey indicated that the current data centre infrastructure of their company is not capable of meeting the demands placed by new types of data (Intel & SAS, 2015).

Use of Hadoop was studied by The Finnish Ministry of Transport and Communication (Liikenne- ja viestintäministeriö, 2013) and in Nordic Hadoop Survey conducted by Intel and SAS Institute (2015). 38 % of the respondents of the survey of The Finnish Ministry of Transport and Communication had experience of using Hadoop (Liikenne- ja viestintäministeriö, 2013). 10 % of the Finnish respondents of Nordic Hadoop Survey (Intel & SAS, 2015) indicated that they have already implemented Hadoop and 4 % had it under implementation. The most often identified primary use case for Hadoop was customer and market intelligence, which was mentioned by 39 % of the Finnish respondents (Intel & SAS, 2015).

4.2 Opportunities

In their report *21 paths to a Frictionless Finland – Report of the ICT 2015 Working Group*, The Finnish Ministry of Employment and the Economy (2013) states that big data is integrally connected with other critical competence areas of Finland. Nokia had a significant impact of development of Finnish innovations and competence during the 1990s and the beginning of the 21st century. The telecom industry is all about processing large amounts of data and utilizing the latest ICT technologies. The Academy of Finland (2007) recognized machine learning, pattern recognition, data analysis, and data mining as particularly strong areas of research in Finland. They state that Finnish institutes are without a doubt doing world leading research in these areas (Academy of Finland, 2007).

According to The Finnish Ministry of Transport and Communications (2013) the opportunities of big data cover all industry sectors and are relevant to organizations of all sizes. In their research agenda for data to intelligence -initiative TIVIT (current DIGILE) (2011) introduce some of the most potential business verticals in terms of data utilization as a starting point for the program. Also The Ministry of Transport and Communications (2014) list possible applications for big data in different industry sectors in Finland.

One of the areas mentioned in both reports is traffic. The report of The Ministry of Transport and Communications (2014) describes how Finland has been a leader in using open transport data and how new transport systems have been studied in Finland. TIVIT (2011) states that the volume of data collected from traffic is projected to grow exponentially. The main sources of this data are intelligent vehicles and road-users carrying sensors, such as smartphones. In addition, road-side monitoring units could be used to collect data. This data creates potential to develop intelligent services based on optimization (TIVIT, 2011). The public sector could utilize data to direct the planning of traffic infrastructure. There is potential to use the data to support the reduction of CO^2 emissions and other pollution (TIVIT, 2011). The Ministry of Transport and Communications (2014) also recognize the possibility of improving traffic safety based on data.

Another sector included in both reports is healthcare and well-being. There is vast potential in utilizing the different types of data created by the healthcare organizations to conduct further research and support decision making (TIVIT, 2011). Intelligent solutions could help to tackle the pressure that ageing of the population is putting on the Finnish healthcare system (Liikenne- ja viestintäministeriö, 2014). The Ministry of Transport and Communications (2014) see potential in using big data to identify national health risks and suggest preventive measures. There is also a growing demand for electronic healthcare devices and health related mobile applications directed at consumer use rather than only for healthcare professionals (TIVIT, 2011). Some of the leading companies in the field are Finnish (Liikenne- ja viestintäministeriö, 2014). These innovations are often based on intelligence built on data.

Finnish manufacturers have developed world leading automated operation and maintenance systems in co-operation with their partners. TIVIT (2011) mentions the potential of big data technologies in further developing these systems. The Ministry of Transport and Communications (2014) introduce operations optimization, resource management, and predictive maintenance as applications of industrial internet. They introduce a vision where all the parts, machines, people, and other resources used in the production are closely monitored by sensors and the produced data is used for intelligent management (Liikenne- ja viestintäministeriö, 2014).

TIVIT (2011) describes how big data enables the companies in the retail industry to develop personalized and context sensitive services and marketing for their customers. Ministry of Transport and Communications (2014) broaden this view to all marketing and development of digital consumer services. They take Google, Amazon, Facebook, and LinkedIn as examples of leading developers of data based services (Liikenne- ja viestintäministeriö, 2014).

Finally, both of the reports mention the opportunity to combine data routinely produced by the public sector to build new and better services (Liikenne- ja viestintäministeriö, 2014; TIVIT, 2011). The Ministry of Transport and Communications (2014) also includes intelligent infrastructures and networks, research, and cleantech as potential application areas of big data in Finland.

4.3 Challenges

Three main challenges for the Finnish organizations in utilizing big data were identified in the survey of The Ministry of Transport and Communications (2013). The first one is actually a combination of two connected problems: lack of experienced data professionals and the lack of big data training. The report (Liikenne- ja viestintäministeriö, 2013) sees that this problem is due to the inflexibilities in the academic world to respond to the fast changes and the lack of training provided by organizations to their workers. It is discussed in the report (Liikenne- ja viestintäministeriö, 2013) that the current BI professionals probably have so much work in their hands that they do not have time to attend big data training and that the few big data experts are too busy to have time to teach others. Another challenge is due to fact that successfully planning and carrying out a big data strategy requires cross-functional and cross-disciplinary understanding and know-how (Liikenne- ja viestintäministeriö, 2013). It is also challenging to find good and reliable partners and consultants as there is so much hype and many different service providers intensively marketing their services (Liikenne- ja viestintäministeriö, 2013). The third challenge is due to unclear and constantly changing legislation concerning data privacy (Liikenne- ja viestintäministeriö, 2013). Especially by combining data from different sources it is quite easy to reach the boundaries of the current legislation (Liikenne- ja viestintäministeriö, 2013).

Their main message of TIVIT's research agenda (2011) is the need for horizontal co-operation in the development of data analytics capabilities in Finland. Considering that Finnish data analysis researchers are recognized as leading experts in their field, the Finnish industries have applied the developed methods surprisingly little. TIVIT recognizes a need to build co-operation between academic researchers and businesses in Finland. Also different industries should be more flexible in applying the methods and technologies from each other. Another problem in data-related research and development in Finland is that too much focus is put on technology development, while failing to identify ways to turn the new solutions into profitable products. They suggest that both of these issues should be addressed by building data utilization focused networks that bring academia and different industry sectors together. (TIVIT, 2011)

5 Research Method and Data

The process of selecting the research method and the pros and cons of the selected method are discussed. The used interview form is introduced. Finally, the data collection process and the analysis of data are described.

5.1 Selecting the Research Method

A qualitative research method was considered appropriate in this case mainly because of the nature of the research questions considered. The research questions widely consider the nature of the big data phenomenon in the context of the Finnish business world. Also why and how -questions concerning big data maturity and success factors are included. Ellram (1996) mentions that if the goal of the research is to explain a particular phenomenon qualitative research method is appropriate. More in-depth nature of a qualitative method enables the researcher really probe the how and why questions.

Qualitative methods have also been reasoned to be appropriate when the theory is in its early stage and there are no well-established models and theories or when the researched field is constantly changing in a rapid pace (Benbasat et al., 1987). Both of these characteristics describe the research related to big data phenomenon well – it is at a relatively early stage and the field is constantly changing as a result of digitalization and development of new technologies.

Yet another consideration in selecting the research method was that based on previous studies it seemed appropriate to assume that not very many Finnish companies have implemented big data solutions so far. For the purposes of answering the research questions, it was important to find companies that can already assess the results, success, and the benefits of their big data initiatives. These companies were not supposed to be many. Also it was to be expected that only a few companies would have publically shared information about their big data efforts, so relevant cases might be difficult to identify. Thus relatively small sample size was to be expected and a qualitative method, that doesn't intend to generalize the results to a wide population but rather studies a small number of cases thoroughly, was chosen.

Interviews are probably the most used data collection method in qualitative research. It is a practical way of gathering information about things that the researcher cannot directly observe like people's thoughts (Merriam, 2014). Interviewing industry experts also provides a more efficient way of gathering data of a phenomenon than direct observing – the experts can summarize the experience and knowledge they have gained during a long period of time. Merriam (2014) states that interviewing is the best technique when studying a small number of selected objects. Also the data of this study was collected by interviews.

However, completely open interviews, where the subject area of discussion is the only predefined limitation, can be difficult to interpret as the interviews might be very different from each other. For the purposes of making the interpretation easier and ensuring that all the areas of interest are covered, semi-structured interviews were conducted. Most of the questions were still open-ended to enable the subjects to express different views and add subjects that they see relevant.

5.2 Formulating the Interview Questions

After the selection of the research method, an interview form was carefully designed based on the analysis of the previous research of big data and the selected research questions, also bearing interviewee friendliness and the interpretability of the results in mind. The interview form, which included as appendix A, was organized into five parts.

The first part included background questions concerning the company. In order to gain a wide understanding of the contextual aspects, the background of the companies was investigated thoroughly. The basic information included company's name, operating industry, the number of employees, turnover, and details of the interviewee and the interview. The company's business model, ownership structure, history, market share, the current financial position, the main competitors, the main products, and the mission, vision, and values were also covered. Finally, the interviewees were asked to describe their role in the company and their link to big data and the big data efforts in their company.

The second part consisted of only one question. The interviewees were asked to define the term big data. Thus it was ensured that the interviewees' definitions of big data were not too different from each other.

One of the objectives of this study was to assess the big data maturity of the companies and to test the applicability of the big data maturity model in Finnish environment. The third part intended to investigate the current state of big data maturity in the interviewed companies. The interviewees were simply asked to describe how their company is currently utilizing big data. They were also specifically asked about the nature of the data they are utilizing and about their technology architecture as these are definitive considerations in assessing the big data maturity. In order to cover a wider understanding of different industries, the interviewees were also asked to assess the state of big data in other companies in their industry in general.

The fourth part considered the outcomes of the big data efforts of the company. To apply and challenge the success factors model (Yeoh and Koronios, 2010), it was important to identify firstly how successful the big data efforts of the interviewed companies have been and secondly what are the factors explaining the success or failure in these efforts.

The interviewees were first asked to identify how they are measuring the outcomes of their efforts. The following question of previous research by Halper (2014) was applied to big data context:

Which statement best describes the value you have seen from your big data efforts?

- We have measured positive top- and bottom-line impact.
- We have measured to-line impact only.
- We believe that we have become more effective, but can’t measure top-line impact.
- We have measured a cost reduction only.
- We believe that we have become more efficient, but cannot measure impact.
- We have gained more insight.

The interviewees were also asked to freely describe their methods and process of assessing the project outcomes.

Different potential benefits that can be achieved by utilizing big data were identified based on the review of previous research. All of the identified benefits can be found, for example, in Global Information Technology Report 2014 (World Economic Forum, 2014) The benefits included:

- Cost savings
- New sources of revenue
- More efficient processes
- New products or business models
- More fact-based decision making
- Increased forecasting accuracy
- More efficient advertising
- Improved understanding of the customers
- Improved understanding of the market
- Increased automation

- Improved quality
- Improved customer satisfaction
- Improved fraud detection and failure prevention

The interviewees were first asked to freely describe the benefits their company has seen from their big data efforts. Then they were asked to assess how significant benefits they have seen in each of these areas using a scale from one to five. The rating of one indicated that the company has not seen any benefits in the area under consideration, while the rating of five indicated very significant benefits. The interviewees were also prompted to add any other benefits they felt relevant.

In the fourth part interviewees were also asked if they would describe the big data efforts of their company as successful and what they would consider as the most important factors explaining the success or the failure. Finally, they were prompted to assess their satisfaction with the results they have seen so far on a scale from one to five and asses the aspects most affecting their satisfaction or dissatisfaction. One indicated that the interviewee was very dissatisfied with the achieved results and five corresponded to being very satisfied the results.

The final part of the interview form asked the subjects about the future plans of big data utilization in their company. The questions intended to identify whether the companies are expected move to a higher stage of big data maturity in the future or are the companies satisfied with their current level of big data utilization. The subjects were asked especially if they can identify new potential data sources or future development possibilities of other kind, and whether they already have plans for future development.

To identify external and internal factors enabling the companies to utilize or prohibiting companies from utilizing big data more extensively, a well-known tool of strategy development (Hill & Westbrook, 1997), SWOT-analysis was used. The subjects were asked to identify the major internal strengths and weaknesses of their company and the major external opportunities and threats related to their future plans.

5.3 Collecting the Data

Personal contacts, the snow-ball approach, as well as articles and other online and offline resources related to the use of big data in Finnish companies were utilized in finding appropriate persons to interview. Fifteen companies were contacted via e-mail or phone and asked if they were interested in taking part to the study. As a result, representatives of 10 large Finnish companies that are already utilizing big data were interviewed between May 6th and October 14th 2015. The interviewees were people highly involved in implementing, using, and developing analytics in their organization with titles ranging from Data Scientist to Chief Development Officer.

The interviewees were given the opportunity to familiarize themselves with the questions beforehand as suggested by Tuomi and Sarajarvi (2009). The interviews were conducted either face-to-face with the interviewees in their company premises or on the phone via Skype whatever was preferred by the interviewee. The durations of the interviews were between 25 and 75 minutes. As background information on the interviewed companies was generally publicly available via the company's web pages, and to save time these questions were not discussed with the interviewees.

The interviewees and the companies preferred to stay anonymous so the companies are referred to as Company A, Company B, Company C, and so forth, following the chronological order of the interviews. Table 1 provides some background information on the interviewed companies. To anonymize the companies, they were divided into three size categories based on the number of employees and three size categories based on the turnover of the company in 2014. The categories were defined so that they provide relevant information on the interviewed companies to support the interpretation of the results but also are wide enough so that individual companies cannot be identified. The categories for the number of employees are 100 – 400, 700 – 2 000, and 4 000 – 20 000, and the turnover categories are 50 – 700, 1 500 – 2 500, and 7 000 – 10 000 in million euro.

Table 1: Background information of the interviewed companies

Company	**Employees**	**Turnover**	**Industrial classification** (TOL 2008)
Company A	700 – 2 000	50 – 700	Computer programming, consultancy and related activities
Company B	100 – 400	50 – 700	Publishing activities
Company C	4 000 – 20 000	1 500 – 2 500	Manufacture of machinery and equipment
Company D	100 – 400	1 500 – 2 500	Arts, entertainment and recreation

Company E	4 000 – 20 000	1 500 – 2 500	Publishing activities
Company F	4 000 – 20 000	7 000 – 10 000	Manufacture of computer, electronic and optical products
Company G	100 – 400	1 500 – 2 500	Computer programming, consultancy and related activities
Company H	700 – 2 000	50 – 700	Land transport and transport via pipelines
Company I	4 000 – 20 000	7 000 – 10 000	Wholesale trade
Company J	700 – 2 000	50 – 700	Arts, entertainment and recreation

After the collection of data, the interviews were transcribed and analysed following the process suggested by Tuomi and Sarajärvi (2009). The process is as follows:

1. Decide what you are interested in the material and make a strong decision.
2. Go through the material separating and marking the issues related to the area of interest. Everything else is left out from this study.
3. Collect the marked parts together and separate them from the remaining material.
4. Classify the material.
5. Write a summary.

(Tuomi & Sarajärvi, 2009). The process of classifying the material was straight forward as the interviews were structured and thus the structure had been defined already in the interview form design phase.

6 Empirical findings

This section introduces the empirical findings of the study. The order of the topics discussed follows the order in which the things were discussed with the interviewees. The first questions stem from the big data maturity model (World Economic Forum, 2014) and the last from the BI implementations success factors model (Yeoh & Koronios, 2010). Finally, the results of the SWOT-analysis reflect the environmental factors and internal competences that shape the big data potential of the companies as described by the big data maturity model. A summary of the interview data question by question is provided on Appendix B.

6.1 The Interviewed Companies and Big Data

6.1.1 Definitions of big data

To investigate how the concept of big data is understood in the companies and to ensure that the interviewer and interviewee understand the concept similarly enough, the interviewees were asked to define big data. The identified definitions reflect the different views suggested by academic literature.

Companies A, B, C, and E mention the three Vs (volume, velocity, variety) definition. Companies F and G emphasize the volume of data by stating that the volume of big data is so large that it becomes difficult to manage by using the traditional approaches. Company D sees the variety of data and the lack of rigid structure as the defining elements of big data.

Companies C and E adopt a technology-centered view and emphasize the new technologies related to the process of acquiring, storing, and refining data. Companies H and J mention the change in the nature of the stored data: raw data is stored as such to be used also for future unknown analytics needs. Company I emphasizes the potential of data in creating value for the business.

It was agreed that the most important factors in differentiating between big data and traditional analytics approaches is the utilization of large amounts of data and the usage of unstructured or semi structured sources of data alongside the well-structured data from the organizational IT-systems.

6.1.2 Company A

The core function of company A is to produce and market mobile applications for consumers. They function also in other business areas but the utilization of analytics is mainly related to application business. The person interviewed is the head of analytics in company A and leader of the business intelligence team consisting mainly of data scientists.

Vast amounts of data of users’ actions in the applications are manly utilized in guiding and further development of applications. The data helps to optimize applications to better engage users, and encourage the users to make in-app purchases. The developers have certain hypotheses of how the users use the applications. The data supports them to validate these hypotheses. They can, for example, see whether the users are using some features, whether they are using them as expected, or whether the application needs more content. Company A also does AB-testing of their applications. AB-testing is a process of presenting different users with a different version of the same application. Analytics are used to determine which of the versions performs best and should be used in the future.

It is important for company A to get the overall picture of all the applications and their users. This data is utilized in customer relationship management, for example in sending personalized and targeted push-notifications to the users. The applications can be cross promoted, or inactive users re-engaged or persuaded to use some other application.

Company A utilizes outside analytics services to get information about the effectiveness of their advertising in other applications or web-pages. Combining information from outside analytics services, their own advertising analytics, and data of the value of their users, company A can optimize advertising volume on different channels.

Finally, the interviewee mentions that their market research also utilizes quite large quantities of data. There are hundreds of thousands of different applications on app stores and a there is quite much data of them also publicly available.

6.1.3 Company B

Company B has evolved from publishing activities to providing online data banks and search tools for consumers and businesses. In addition, they offer business to business marketing and sales solutions that include analytics and other digital services. The interviewee’s title is head of architecture and he is responsible for research and development functions in the organization. He is also responsible for handling big data and the related technologies.

Company B utilizes many different kinds of data from their own systems, their partners, and customers. The main application areas include: 1. Reporting, supporting management decision-making, and optimization of internal processes 2. Developing data-based online services for consumers and businesses 3. Offering marketing and sales optimization services for business customers.

6.1.4 Company C

Company C is a global producer of industrial machinery. Their senior research architect was interviewed for this study. He is leading technology research in the organization and is currently managing research related to a strategic industrial internet initiative.

Company C is in a unique position to have access to data of the production, sales, usage, and maintenance of their products globally. The interviewee points out that collecting and analysing sensor data is something the companies in the industry have been doing already for decades. What is new is that the advanced technologies enable combining this data with other data sources and conducting more sophisticated and larger-scale analysis.

Data utilized by company C is mostly well structured data from the organization's IT-systems and from sensors in the machinery. They have developed what they call a next generation enterprise resource planning system for maintenance that stores data of all maintenance actions of the company's products globally. The company is able to assess the life-cycle of each product and predict problems in quality and demand of specific spare parts. The interviewee describes how they have been able to optimize quality according to which products are used heavily and which are not used very frequently. Quality of less used products can be lower, which creates savings. Improving the safety of the products is also one of the important application areas of data. The condition of critical parts can be monitored through sensors and potential safety problems can be addressed early on.

6.1.5 Company D

Company D provides entertainment services for consumers both online and at physical customer service points. The interviewee works as an ICT architect in the company and is responsible for the technical side of the company's ongoing big data initiative.

The company is currently building big data capability and renewing their data warehouse. The main function of the renewed platform is to collect data of users' actions in the online services and usage of products in customer service points. However, the platform is designed to be able to store and handle also versatile data from outside sources, for example social media content or open demographics data. The main benefits that they expect to gain by the new system include automating especially data-management related processes, gaining new understanding of the customer and the products to support management decision making, service development, and optimize marketing.

6.1.6 Company E

Company E operates widely in the media industry. The core of their business has traditionally been in publishing but they are increasingly focusing also on other business areas. The interviewee from this company is the vice president of customer insight and analytics. She is the leader of the global big data initiative in the company covering all business areas including digital and non-digital products.

The interviewee comments that data that can be called big data is mainly related to the digital services. Websites and mobile applications produce click-stream data of the users' actions. There is also data related to online advertising, for example data of impressions and clicks. Social media data is occasionally used. This online data can be combined with more traditional business data, like sales data, customer data, and market research data.

Raw data can be refined to calculate, for example, different attributes like propensity of risk to churn or estimated lifetime value for the customer profile. In addition to customer relationship management, data and analytics are used to customize digital products in real time based on the needs of individual customers, provide targeted options for advertisers, and automate processes.

6.1.7 Company F

Company F operates globally in different business areas tightly related to data and information technology. Their major function is to build, maintain, and develop infrastructure for transferring data and provide supporting services to the customers and end-users. Also the person interviewed – head of big data and analytics – is the most familiar with this business area. She works in a business unit called technology and innovation that conducts research of the new technologies and builds the technology strategy.

Company F has access to huge amounts of signalling data. Another major data-type is data related to maps and location. They utilize signalling data to manage infrastructure and its capacity in real time. They can, for example, predict congestions in advance. The second application area is customer experience management. Company F analyses the signalling data of individual users to compose a real-time happiness indicator for the user. Thirdly, these masses of data can be utilized to get information about the society, for instance to support urban planning.

6.1.8 Company G

Company G develops online applications for consumers. Company G operates largely on the same market as company A and companies A and G could be described as competitors. One of the company's data scientists was interviewed for this study. He is responsible for analytics related to a specific application and part of the application development team.

Outside the development teams of specific applications, the company uses data for reporting and creating different kinds of future scenarios for company management. Marketing and advertising also use data but these functions are based outside Finland. The main application area of big data is the use of behaviour data created by the users in the applications to support application development.

The interviewee says that the developers often like to have a data point to support their view. They can follow how the applications are used and if there are any changes. For example, they can identify features that the users are not utilizing and recognize that no more resources should be used to further develop those features.

6.1.9 Company H

Company H is a public transport operator. Their technical director, that was interviewed, he is responsible for the vehicle fleet, real estates, and ICT in the organization. All the cars have data loggers that collect sensor data and location data every second. This data is then combined with the data in the ERP system and reports of the function of each vehicle and driver are constructed. The reports are used to support management decision making and, for example, to measure the performance of the vehicles, assess the maintenance needs of the vehicles, and to investigate customer reclamations.

6.1.10 Company I

Company I is a retail chain that operates in three different markets. Customer data and analytics team provides analytics services for partnering companies and subsidiaries in all the markets. The person interviewed is the leader of this team.

The company's loyalty program covers significant part of the Finnish market and produces large amounts of traditional customer data. The company's systems also store receipt line data of the sales of each store. The websites, online stores, and mobile applications produce data of the users' actions. There is customer satisfaction and market research data. Finally, there is data from outside sources, for example open data has sometimes been combined with other data sources.

The interviewee states that the analytics needs in each market are different. The team provides each store with customer data specific for that store to help them to optimize pricing, the assortment, and to improve the service in other ways. The stores can identify the sources of their current sales and customer segments with potential to increase sales. Company I is also starting to use data in customizing and targeting marketing.

6.1.11 Company J

Company J is a competitor of company D and provides customers with entertainment services both online and at physical service points. CDO of company J tells that their big data initiatives are mainly still on the planning stage. They are currently using data to optimize server usage in their data center and in mobile applications to alert the users that seem to be using the services too frequently.

Main data sources of company J include the online services that produce click-stream data of the users and the company's service systems. They want to store all of the data produced by their systems to create a data lake that can be utilized for future unknown analytics needs. The interviewee also mentions that it would be interesting to combine data form their own systems with some external data sources. They see that the main potential benefits of developing big data capabilities include improving customer experience and overall customer understanding, optimizing marketing, and predicting maintenance needs of their systems.

6.2 Big data maturity

Based on how the interviewees described big data utilization in their companies and the big data maturity model of World Economic Forum (2014) introduced in the fourth section, the big data maturity of each interviewed company was assessed. Table 2 presents an overview of data utilization in the interviewed companies. The maturity model seems quite well applicable to the interviewed companies. However, some companies present many other characteristics of higher maturity stages but are not utilizing external data sources as much as defined by the maturity model.

None of the companies seem to be at the first stage of big data maturity. In the first stage the company is not yet utilizing big data at all. The intention of this study was to examine companies that have already implemented big data solutions. Thus none of the companies are on the first maturity stage any more.

Companies C, D, H, I and J appear to be at the second stage of maturity, *Functional area excellence*. The extent to which these companies utilize big data varies significantly but they all analyse of both internal and external data to optimize production, resource utilization, maintenance, pricing, or assortment, and to gain understanding about their customers. However, big data is still applied only to selected business domains and the companies are not monetizing big data.

Four companies, A, E, and G, are at stage three, *Value proposition enhancement*. Big data is clearly a value driver and a competitive advantage to these companies. They provide targeted advertising services, personalizing services in real time for the individual customer's preferences or big data is an integral part of their product or service development.

Companies B and F seem to be at the highest maturity level, *Business model transformation*. They are utilizing big data extensively, data is a defining factor in their business model, and they are providing services that are entirely based on data.

Table 2: Overview of data utilization in the interviewed companies

Company	**Types of data**	**Uses of data**	**Maturity stage**
Company A	Event data from users' actions in online applications Data from servers (e. g. advertising data)	Application design Customer relationship management Online advertising Market research	3
Company B	Event data from users' actions in online applications Search engine queries Browsing data Location data	Service development Reporting Analytics services for customer companies Process control Data-based products	4
Company C	Sensor data Claims data Maintenance data Customer data Sales data Spare parts data Machinery usage data	Production planning Automating and optimizing maintenance Quality optimization Reports to the customer companies of the performance of their machinery Improving safety Optimizing sales activities	2

Company D (in the future)	Event data from users' actions in online applications Usage data of offline products Social media content Open data (e. g. demographics)	Service development Automating processes	2
Company E	Customer data Market research data Click-stream data Online advertisement data Social media content	Customer relationship management Development of services Personalization of services Optimizing and automating advertising Optimizing sales activities Automating processes	3
Company F	Signaling data Location data	Optimization of infrastructure and capacity Reports and analytics to primary customer companies, other companies, and public sector Predicting and preventing quality and capacity problems End-user experience management Data-based products	4
Company G	Event data from users' actions in online applications	Application design Reporting and future scenarios to management	3
Company H	Sensor data Location data	Management decision making support Performance measurement and improvement of vehicles and drivers Assessing maintenance needs Reclamation investigation	2
Company I	Customer data Transactional data Location data Market research data Click-stream data Open-data	Optimization of pricing Improving customer service Assortment optimization Future sales potential Optimization and personification of advertising	2
Company J	Server log data Event data from users' actions in mobile applications	Server usage optimization Service development Misuse detection	2

6.3 Big data architecture

While not included in the big data maturity model (World Economic Forum, 2014), the used technologies also tell something about the big data maturity of a company.

Some of the interviewed companies did not want to disclose the details of their big data architecture, while others described it quite comprehensively. A common approach seems to be to combine Amazon's data storage and cloud computing services with open source big data tools. More specifically, three companies mention that they have adapted this approach (companies A, B, and G). Different open source tools mentioned include: Hadoop (companies A, C, D, E, and F), Kafka (companies D and F), Storm (company F), Pig (company G), Spark (company G), and open source scripting languages (companies A and G). Company F has developed customized big data tools, for instance a stream processing engine, together with their partners. Companies H and J are still relying on more traditional approaches of Microsoft SQL server and Cognos.

Six of the interviewed companies have outsourced data storage and computational capacity. Companies A, B, and J are both doing analytics in-house and buying analytics services from outside service providers, while five of the companies want to develop analytics know-how and keep it in-house. Company H mentions that it is working in close collaboration with its many partners in building big data competence. Company H seems to have chosen quite different approach in comparison to the other companies: the company owns a data center but is buying all analytics services from outside.

6.4 Overall state of big data in each industry

To get some insight to the overall state of big data in the industries where the interviewed companies are operating the interviewees were also asked about their views on how their competitors are using big data. Interviewees from companies C, E, and H believe they have the leading big data solutions in their industry. Representatives of companies A, B, D, G, and J believed that their competitors have pretty similar approaches to analytics as them, while companies A and G still believe that they have some competitive advantage in this area.

6.5 Benefits from Implementations and Future Plans

6.5.1 Measuring outcomes

In order to understand implementations success in the interviewed companies, the level of successfulness of the projects should first be assessed. To understand how the companies are actually measuring the results of their big data initiatives the interviewees were asked which of the following statements best describes the value their company have seen from their big data efforts:

1. We have measured positive top- and bottom-line impact.
2. We have measured top-line impact only.
3. We believe that we have become more effective, but can't measure top-line impact.
4. We have measured a cost reduction only.
5. We believe that we have become more efficient, but cannot measure impact.
6. We have gained more insight.

Several interviewees protested that projects are very different from each other and it is hard to choose one description for all big data efforts. However, companies D, E, F, and H state that they have measured both positive top- and bottom-line impact. Company A has been able to measure top-line impact. Companies I and J chose the option "We believe that we have become more effective, but can't measure top-line impact. Companies B and C have measured only cost reduction. The representative of company G chooses the option "We have gained more insight."

Practices in assessing and measuring the results of big data projects seem to vary considerably, which of course is not surprising as the interviewed companies operate in very different markets and are in different stages of big data maturity. For example, the representative of company D tells that the metrics are project specific. They follow, for example, changes in margins, sales, the number of customers, or the number of identified customers. In company E, the analytics team also has a sales quota like the sales team. For company H there simply has to be savings in the fuel consumption or maintenance of the vehicles.

6.5.2 Achieved benefits

The successfulness of the big data efforts of the companies was also assessed by considering the achieved benefits. Before introducing the interviewees with the predefined list of the potential benefits of big data utilization, they were asked to freely explain the most important benefits they have reaped this far.

Three companies (company E, I, and J) mention improved marketing. They explain that big data enables more relevant and personalized marketing message. Cost savings are also achieved through better-targeted marketing. Companies D and F state that they have been able to make their internal processes more efficient. Companies F and J mention improved customer experience. Companies I and J perceive improved understanding of the customers as one of the most important benefits.

Other mentioned benefits include:

- Improved customer retention (company A)
- Increased turnover per application (company A)
- Ability to handle diverse data and large data volumes and capability to conduct more sophisticated analysis (company B)
- Improved understanding and visibility to the life-cycle of the products (company C)
- Ability to better optimize quality and predict demand (company C)
- Improved safety of the products (company C)
- Ability to respond to changes in real time (company E)
- Ability to personalize online services for each customer (company E)
- Ability to measure customer satisfaction in real time and even predict it (company F)
- Ability to predictively manage infrastructure (company F)
- Streamlining of architecture (company F)
- Ability to focus application development activities more efficiently (company G)
- Ability to optimize assortment and pricing (company I)

The interviewees were also asked to rate a predefined set of potential benefits according to how significant benefits their company has seen in this area. The results are illustrated on Figure 13. Most of the companies have achieved significantly or very significantly more fact-based decision making. The understanding of the customers and advertising efficiency have improved significantly. The two next most significant benefits include increased forecasting accuracy and more efficient processes. Companies have also seen improved fraud detection and failure prevention, improved customer satisfaction, improved quality, cost savings, and new products or business models.

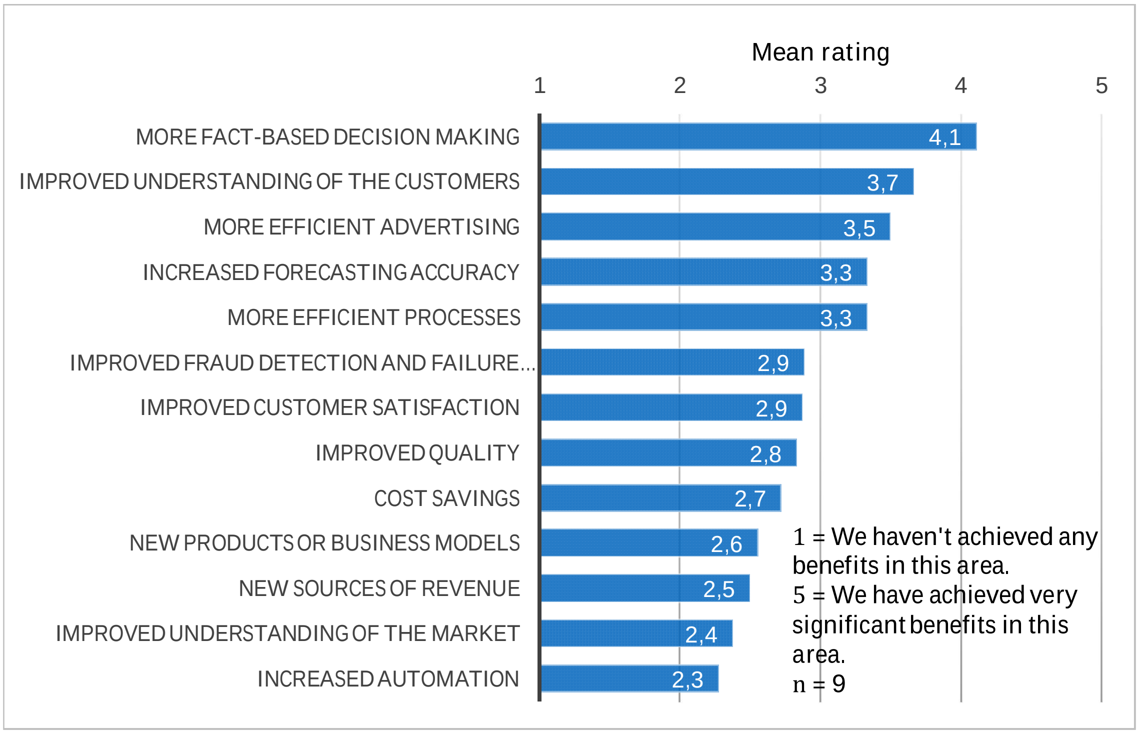


Figure 13: Significanse of benefits achieved from big data initiatives by the interviewed companies

6.5.3 Project success and success factors

The interviewees see their big data efforts overall as successful. However, many of them say that they are still only at the beginning of the journey in utilizing big data. Companies B and G complain that they have almost always had problems with young technologies at the beginning of projects.

Four interviewees (companies A, B, C, and G) see that having talented people with the right mentality is the key factor contributing towards success. Support from the top management is identified as a success factor by companies C and E. Companies C, D, E and H mention strong support from the business side. These factors are also strongly emphasized by Yeoh and Koronios (2010).

Other success factors identified by the interviewees include:

- Finding good partners (companies C and F)
- Having a digital product that makes it easy to collect data (company A)
- Successfully training people to use new technology (company B)

- Having an experimental approach at first (company C)
- Looking for applicable approaches outside the company's own industry (company F)
- Starting to investigate the opportunities of new technologies and approaches in an early stage (company F)

The interviewees seem to identify also open and outward-oriented approach as a success factor in contrast to only focusing on things happening inside the company or a single industry. They say that finding good partners and looking for solutions outside their own industry has contributed to the success of the company's big data efforts.

6.5.4 Satisfaction to achieved results

To further understand the perceived successfulness of the big data initiatives, the interviewees were asked to assess their satisfaction to the results their company has achieved this far from their big data efforts. The mean satisfaction score retained is 3.1.

The main source of dissatisfaction seems to be that the interviewees see potential benefits that have not been achieved yet or that their company has not even started to build big data competence in full scale. One of the interviewees states:

"We are on a marathon and have proceeded only a kilometer this far."

Also problems with new technologies are causing dissatisfaction. Representative of company G describes the situation:

"The tools are still quite immature. Using them is tedious as there are bugs. They require much technical know-how. Getting things working has sometimes taken more time than hoped because we haven't had enough knowledge of the tools and how to use them, or because a tool has still been under development. In 2012, as we started to work with these things, the situation was still much worse. Nowadays you can quite easily start using the tools."

Some interviewees feel frustrated because it is hard and time consuming to change the way people do things – the practices and procedures. Several of them say that it takes time to change the course of a big ship. In addition, representative of company E describes how sub-optimization in the product-centred business units sometimes eats up the benefits from the efforts of the more customer-oriented analytics team.

6.5.5 Future plans

It seems probable that the interviewed companies will continue to develop their big data competence and move to the higher levels of big data maturity in the future. All of the interviewed companies are actively developing their big data competence and have concrete plans for the future. Sources of data that the companies are not yet utilizing but see as potentially valuable include social media (companies E and G) and sensor data created by the increasing number of different objects connected to the internet (company F). Companies D and H simply mention that they want examine the possibilities in using external sources of data. Company C wants to be able to combine data from different sources more flexibly.

Three companies mention that they want to introduce more sophisticated analytics: Companies A and H are planning to develop more predictive models. Companies G and H intend to start analyzing data in real time. Company H also aims to look into the possibilities of utilizing machine learning algorithms. Representative of company G mentions that they are looking forward to the development of technological tools that enable better analysis of correlations in very large data-sets.

Some of the mentioned future plans concern organizational change and the development of new products. Company B works to further develop its digital offering and company E is planning to introduce products based profoundly on data. Company D looks into finding good analytics partners and developing co-operation with them. Company F focuses on taking the already developed competence into use throughout the company.

6.5.6 SWOT-analysis of the future plans

The big data maturity model (World Economic Forum, 2014) introduced the concepts of external enablers and internal competences in explaining the big data potential and the resulting big data maturity of the company. SWOT-analysis of the future plans in developing big data competence shed some light on the external and internal factors either enabling the companies to become more big data mature or hampering their development.

Access to large amounts of data (companies C, E, F, G, and I) and strong know-how (companies A, B, F, and G) are things that several companies see as their internal strengths concerning the future possibilities in big data. Company C sees harmonized and standardized processes and supporting IT-systems as their core strength. Also Company A considers good technological tools as a key enabling competence. Being a born-digital company makes it easy for company B to utilize big data. They also consider their ability to productize analytics as a critical strength. Company D states that their existing data analytic culture significantly improves their opportunities in harnessing big data.

Lack of organizational agility is seen as an internal weakness by four companies (D, E, F, and I). Correspondingly becoming left behind in the development is identified as a major threat by three companies (D, F, and I). Company A considers too strong process orientation as their major weakness. As strong know-how and functional technology are seen as strengths, not having the required know-how in-house (company H) and legacy IT-systems (company J) are seen as important weaknesses in further developing big data competence.

Most of the identified opportunities relate to the improved availability of data and possibilities offered by new technologies. Company D directly mentions the availability of data as a major factor creating possibilities for further development. Company C describes how their market position and global IT-systems can give them global visibility to data. Company F sees that the industry is becoming increasingly virtualized, which means that it becomes easier to collect data. Technological development, in turn, is directly mentioned by company J. Companies A, B, F, G, and I describe more concrete application areas of improved technology. Especially, the ability to better combine data from different sources is mentioned by two companies (A and I).

Companies A, C, D, F, and I identify treats related to organizational agility, the change of culture, and business processes. Three companies mention crisis with data safety or privacy (companies B and G) or stricter privacy legislation (company B) as a major factor threatening the further development in big data. People are also seen as a critical factor and thus as a source of uncertainty. Companies H and J are concerned of the possibility that people with critical know-how would leave the company, and company J mentions the difficulty of finding good data scientists. Company H is also concerned of the possibility of partner lock-in.

7 Discussion and Conclusion

This part discusses the findings of the study reflecting the theoretical frameworks of big data maturity and BI success factors. The implications for practice and theory are discussed. Finally, the limitations of the study and possibilities for further research are considered.

7.1 Discussion of the Findings

7.1.1 Findings related to big data maturity

The objective of this study and the research questions were introduced in the first section. The research questions related to the big data maturity model introduced by World Economic Forum (2014) included the following:

1. How big data mature are Finnish companies that are utilizing big data?

2. Is the big data maturity model applicable in Finnish business environment and does it succeed in differentiating the companies with different levels of maturity?

3. What kind of external factors and internal competences the companies recognize as defining their big data potential.

How the interviewed companies are using big data was explained in part 6.1, where the interviewed companies were introduced and their big data utilization was discussed. The identified applications cover versatile facets of business. The most predominant application areas seem to be marketing and customer relationship management related processes and product or service development, especially when the product or service is digital. The companies are also utilizing different kinds of data from different sources. However, the data produced by the company's own systems still seem to be the most utilized source of data. Data from external sources is still utilized relatively little. The most common source of big data among the interviewed companies is click-stream and event data of user's actions produced by different online services. Sensors of various types are another important source of data.

The stages of big data maturity of the interviewed companies was discussed in part 6.2. Five of the interviewed companies seem to be on the second stage of maturity: functional area excellence, three on the third stage: value proposition enhancement, and one on the final stage: business model transformation. There seems to be quite much variability in the big data maturity of the interviewed companies, but most of the companies are still on the early stages of maturity. The results also indicate that the model can successfully be applied in Finnish context and that it efficiently differentiates companies in different stages of maturity.

According to the results of the future plans -part of the interviews, all of the companies are actively striving to utilize big data more widely and efficiently. All interviewees identify several future possibilities and say that they already have plans for development.

The results of SWOT-analysis part of the interviews provide some insight into the environmental factors and internal competences that define the big data potential of a company. The main environmental factor the companies perceive as enabling more efficient use of big data in their company is the growing amount of accessible data. The utilization of data requires internal capabilities. Many of the interviewed companies identify strong know-how as their main capability related to big data. A large share of the interviewed companies see the lack of organizational agility as their main weakness. This is understandable when the large size of the interviewed companies is considered. Data privacy problems and changes in legislation are environmental factors that are seen as potentially restrictive to the future development of big data capabilities.

7.1.2 Findings related to big data success factors

The first section introduced the following research questions related to the success factors model by Yeoh and Koronios (2010):

4. What are the factors that Finnish companies identify as contributing towards success in big data efforts?

5. Are the identified success factors aligned with the model of Yeoh and Koronios (2010)?

All of the interviewees perceive the big data efforts of the company as successful so far. They indicate of being on average somewhat satisfied with the results. However, two interviewees indicate problems with technology in the beginning on new efforts. Four interviewees state that their big data efforts are still in a very early stage and the final successfulness of them still remains to be seen.

The companies indicate of achieving several significant benefits from their big data initiatives. The most important benefit areas include more fact-based decision making, improved understanding of the customers, and more efficient advertising. The companies have seen on average at least some benefits in all the studied areas.

The first success factor in the model of Yeoh and Koronios (2010), **committed management support and sponsorship** is mentioned by two interviewees. The success factor of **clear vision and well-established business case** is not explicitly mentioned as a factor explaining the success in the big data efforts by the interviewees but it can be identified implicitly in several discussions. Four companies mention the importance of having project teams consisting of skilled people with the right mentality and two companies emphasize the support from the business side. Company H mentions that the need and motivation for the efforts came from the business side. These views can be seen to support the success factor of **business-centric championship and balanced team composition.** Underlining business side support can also be seen as evidence of the success factor of **business driven and iterative development approach**. Company C also mentions experimental approach and not taking too big bites at once as one of their success factors. This can be seen as synonymous with iterative development approach. Being able to successfully train people to use the technology explains success big data efforts according to company B. This provides support for **user-oriented change management** as a success factor. Evidence supporting the success factors of **business driven, scalable, and flexible technical framework** and **sustainable data quality and integrity** is not found in the interview data.

Two companies also mention that finding good partners was an important factor in defining the successfulness of their initiatives. Company C mentions that they looked for applicable approaches outside of their own industry. These findings suggest that openness to look for solutions outside the industry and operate with partners could constitute an additional success factor.

7.1.3 General discussion

The interviewed companies have found different ways to utilize big data that are appropriate to their industry, while marketing and product or service development are the most dominant areas of application. It is apparent that companies with digital products find it easier and more natural to utilize big data, but also companies operating in traditionally non-digital industries have found ways to support their business with data-based services. It is promising that most of the interviewed companies also seem to understand analytics and big data as a widely concerning all the different business functions and with versatile opportunities to gain benefits, not just as an IT thing or way to optimize processes to get cost-savings, for example.

The most common architectural solution seems to be to buy the data storage space and the computational capacity from a service provider and build the analytics technology on different open source solutions. Most of the companies also seem keen to develop analytics know-how inside the company while using partners appropriately.

The companies have already reaped concrete benefits from utilizing big data and are generally satisfied with the results. The benefits in changes in the decision-making process and improved customer understanding are perceived most significant. Some dissatisfaction also exists mostly due to the fact that the companies understand that unrealized potential still remains and that there is room for further development.

It seems that despite of the varying levels of big data maturity of the interviewed companies, all of them are actively and intensively developing their big data competence further. They recognize that the analytics field is developing fast and changing entire industries and they want to keep up with the development. The companies recognize analytics as a way to gain competitive advantage against their competitors. Simultaneously, they are concerned that new born-digital operators will find innovative ways to make data work for them, transform entire industries, and make the traditional ways of making business obsolete.

What should be pointed out, however, is that often it is not relevant to the companies whether their analytics solution is big data or not. Their major concern is how they can improve their business by utilizing data and analytics. The technological considerations and the sizes of data sets are based on what best responds to the requirements of the different analytics use cases in the company. Thus it is not relevant to the businesses to define big data or differentiate between small and big data analytics.

7.2 Implications for Theory

The data seems to confirm the critical success factors model (Yeoh & Koronios, 2010) to some extent. At least implicit evidence supporting all the five success factors in organizational and process dimensions was found. However, the success factors included in the technology dimension were not mentioned by the interviewees. This could suggest that technological factors are not considered as important or critical in defining big data implementations success in comparison to other success factors.

The interviewees also brought up the importance of alignment of analytics and the overall business that the success factors model underlines. They emphasized the importance of top-management support and support from the business side. It was seen preferable that the need and initiative for developing analytics would come from the business-side rather than from the analytics function.

When describing the success factor of **business-centric championship and balanced team composition** Yeoh and Koronios (2010) emphasize that the teams are cross functional. The interviewees did not mention the cross-functionality of the teams as a success factor. Instead they underlined the skill-set and the attitude of the individual members of the team. Perhaps, cross functionality is not a vital thing when composing teams for big data projects. Balanced team composition can rather be achieved by considering the skill-sets and attitudes of the individuals.

Some of the interviewees mentioned the importance of finding good partners and looking for solutions outside the company's own industry sector in developing analytics successfully. This is something that is not included in the model of Yeoh and Koronios (2010). The evidence suggests that a success factor of openness to look for solutions and partners from outside organizations is also relevant to big data implementations.

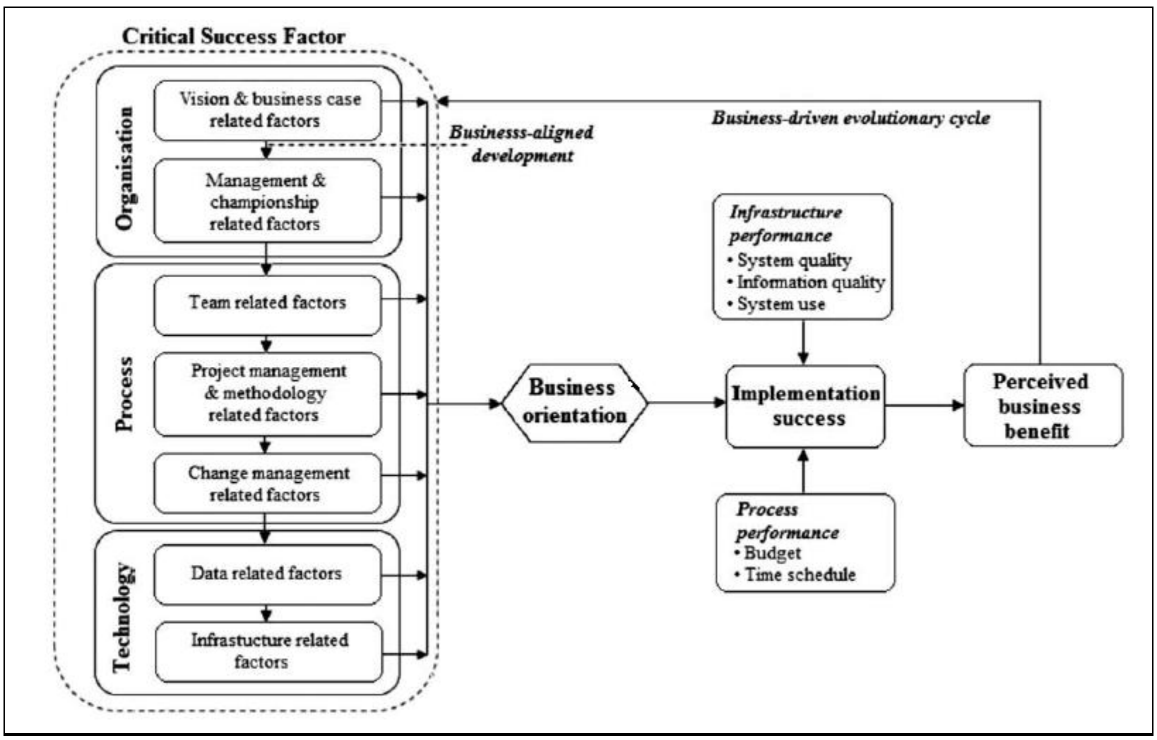


Figure 14. Critical success factors for BI implementations (Yeoh & Koronios, 2010)

Figure 14 visualizes the components of the original model. The new success factor is related to vision, business case and management. Thus it could find its place inside the organizational dimension.

The big data maturity model (World Economic Forum, 2014) seemed to be well applicable to the interviewed companies. All the different maturity stages were found. The majority of the companies were identified to be on the maturity stages 2 and 3, while two companies seemed to be on the first stage of big data maturity and on the final maturity stage. This finding also supports the model's applicability and usefulness in efficiently differentiating companies in terms of big data maturity.

However, the maturity model does not consider the differences in technological solutions and levels of complexity of the utilized methods. These factors can also be seen as differentiating the interviewed companies on different levels of big data maturity. Also the assumption that companies utilizing external data are more big data mature than companies utilizing mainly internal data from the company's own systems can be questioned. The internal systems of some companies can produce massive amounts of unstructured and semi-structured data.

7.3 Implications for Practice

The results of the interviews provide valuable insight for Finnish companies currently evaluating the potential of big data for their company. The study includes plenty of examples of how big data can be applied in different industry sectors. The potential challenges, problems, and threats that the companies should consider are discussed. Also the opportunities and challenges specific to the Finnish industries are discussed. Finally, the identified success factors can be used in determining how the companies should go about when they start to implement big data solutions.

The study provides insight for Finnish companies that operate in the same industry sectors as the interviewed companies and are interested in how their competitors are utilizing big data. This can help the companies in benchmarking themselves in terms of analytics function and identify where they might have competitive advantage against their competitors are where they are lagging behind the leading companies.

Also the Finnish government and the public sector can gain insight of the big data maturity of Finnish companies from this study. This can help them in determining how to best support the competitiveness of Finnish industries in the global market.

7.4 Limitations of the Study

Morse et al. (2002) state that: *Without rigor, research is worthless, becomes fiction, and loses its utility*. They introduce five verification strategies for ensuring the validity and reliability of qualitative study. These strategies include: methodological coherence, appropriate sample, concurrent data collection and analysis, theoretical thinking, and theory development. The validity and reliability of this study will be discussed by assessing each of these factors.

According to Morse et al. (2002), the aim of methodological coherence is to ensure congruence between the research question and the components of the method. This study is unusual in a sense as some of the research questions conserving the state of big data in Finland are factual in nature and thus would have supported a selection of a quantitative method. A qualitative method was chosen because of the nature of the research questions concerning the success factors model and because of the expected restrictions to collecting a large sample of data. Perhaps, if a quantitative approach had been selected, more generalizable results would have been attained.

Appropriateness of the sample can be assessed by considering the quality and quantity of the sample. As the interviewees were all industry experts, it seems reasonable to state that the participants of the sample best represent and have knowledge of the research topic. The main limitation of this study is its narrow scope in terms of size of the sample. When choosing a qualitative research method, the researcher always trades the scope of the study and generalizability of the results to gain deeper understanding of the studied phenomenon. The results can by no means be said to reflect the overall state of big data in all Finnish companies. The interviewed companies were very large in comparison with the majority of Finnish companies. Also many of them are regarded as the leading Finnish companies in analytics utilization.

The interviewees seemed to be motivated to talk about big data and to contribute towards reliable research results. The large share of people that wanted to contribute by giving an interview is one indication of the motivation. 15 people were contacted in total and 10 of them were willing to give an interview. The interviewees also explicitly stated that they thought that the research topic was actual and interesting and expressed interest in familiarizing themselves with the results of the study.

Some researchers also suggest that when the sufficient number of people have been interviewed, the point of saturation is reached so that the results start to become repetitive and new ideas and opinions are not mentioned any longer (Tuomi & Sarajärvi, 2009). There is no evidence of reaching the point of saturation in this case as new views and ideas were still identified in the last interviews. It has been suggested that 15 interviews would be sufficient to reach the point of saturation in most cases (Tuomi & Sarajärvi, 2009). The sample of this study consisted of 10 interviews which is considerably less than the suggested 15.

The aim of collecting and analyzing data concurrently was not achieved as in this case the data was first collected and then analyzed. Additional considerations related to the data collection method include the effect of the interviewer in influencing the responses and the trustworthiness of the results as it cannot be verified that the subjects are actually telling the truth. The interviewer should remain as neutral as possible and avoid influencing the interviewee's thoughts. The nature of interviews as a two-way human interaction is both a strength and a weakness of the method. Also the need for the interviewees to be careful of not disclosing their trade secrets to the competitors should be considered. Because of this, the results might not reflect the full reality of the state of big data even in the interviewed companies.

Theoretical thinking is indicated in the way of which the results of the study are discussed in relation to research questions and the theoretical framework. Theory development is defined as deliberatively moving between the micro perspective of the data and macro-level conceptual understanding. The success factors framework and the big data maturity model have been challenged and verified by the empirical data. Also the introduction of a new success factor has been suggested to further refine the success factors model.

7.5 Further Research

The considered research questions cannot be comprehensively answered in the scope of this study. They rise many interesting further research possibilities.

The big data maturity of companies would be best assessed by a questionnaire tool developed to assess different aspects of big data maturity. As several different maturity models already exist, there is a need to compare these models to find and verify the most appropriate framework for assessing big data maturity. Then a questionnaire tool could be developed. This tool could be used also in a quantitative setting further assessing the big data maturity of Finnish companies.

The external enablers and internal competences that the report of World Economic Forum (2014) mentions to explain big data maturity potential of a specific company were not comprehensively covered in this research. It would be interesting to know more about the aspects that enable or restrict Finnish companies in utilizing big data. These factors would be best studied by using a qualitative research method.

The recognized new success factor of openness to look for solutions from other industry sectors and finding partners should be studied further and verified. Also a completely separate success factors model could be developed specifically for big data efforts by following, for example, the research process introduced by (Yeoh & Koronios, 2010). First it should be discussed does the definition of success of information systems implementations apply also to big data efforts.

References

Academy of Finland. (2007). Computer science research in Finland 2000–2006.

Accenture. (2014). Kaikki liiketoiminta on digitaalista – mutta mitä sen suhteen pitäisi tehdä? Kurvio, M., Kaario, K., Kelloniemi, J., Riekki, J.

Apache storm. (n.d.). Retrieved from: https://storm.apache.org/index.html

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Petterson, D., Rabkin, A., Stoica, I. & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM, 53(4),* 50–58.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly, (11(3),* 369–386.

Bensinger, G. (2014, January 17). Amazon wants to ship your package before you buy it [Web log post]. Retrieved from: http://blogs.wsj.com/digits/2014/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/

Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*, *109(2)*, 155–172.

Boyd, D & Grawford, K. (2012). Critical questions for big data – Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15(5),* 662–679.

Chaudhuri, S. & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod Record, 26(1),* 517–526.

Chen, H., Chiang, R. H. L., Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36(4),* 1165–1188.

Chen, C. L. P., Zhang, C-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences, 275(2014),* 314–347).

Chen, M., Mao, S., & Liu, Y. (2014a). Big data: a survey. *Mobile Networks and Applications, 19(2)*, 171–209

Chen, Z., Lu, Y., Xiao, N., & Liu, F. (2014b). A hybrid memory built by SSD and DRAM to support in-memory big data analytics. *Knowledge and Information Systems, 41(2),* 335–354.

Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. Greenplum. (2009). MAD skills: New analysis practices for big data.

Companies using Apache Storm. (n.d.). Retrieved from: https://storm.apache.org/documentation/Powered-By.html

Davenport, T. H. (2013). Keep up with your quants. *Harvard Business Review, 91(7/8),* 120–124.

DB-engines ranking of key-value stores. (2015, September) Retrieved from: http://db-engines.com/en/ranking/key-value+store

DeLone, W. H. & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3(1),* 60 – 95.

Dutta, D. & Bose, I. (2015). Managing a big data project: The case of Ramaco Cements Limited. *International Journal of Production Economics, 165, 293*–306.

Dyson, R. G. (2002). Strategic development and SWOT analysis at the University of Warwick. *European Journal of Operational Research, 152(2004),* 631–640.

Ellram, L. M. (1996). The use of the case study method in logistics research. *Journal of Business Logistics, 17(2),* 93–138.

Garber, L. (2012). Using in-memory analytics to quickly crunch big data. *Computer, 45(10),* 16–18.

Gartner. (2014, August 14). Gartner's 2014 hype cycle for emerging technologies maps the journey to digital business. Retrieved from: http://www.gartner.com/newsroom/id/2819918

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009) Detecting influenza epidemics using search engine query data. *Nature, 457(19),* 1012–1015.

Golab, L. & Özsu, M. T. (2013). Issues in data stream managmenet. *ACM SIGMOD Record, (32)2,* 5–14.

Hadoop wiki – Powered by. (2015, August 6). Retrieved from: http://wiki.apache.org/hadoop/PoweredBy

Halper, F. (2014) Predictive analytics for business advantage. *TDWI Research.*

Hill, T. & Westbrook, R. (1997) SWOT analysis: It's time for a product recall. *Long Range Planning, 30(1),* 46–52.

Hull, G. (2015). Successful failure: what Foucault can teach us about privacy self-management in a world of Facebook and big data. *Ethics of Information Technology, 17(2),* 89–101.

Humberty, M. (2015). Awaiting the second big data revolution: From digital noise to value creation. *Journal of Industry, Competition and Trade, 15(1),* 34–47.

Indrawan-Santiago, M. (2012) Database research: Are we at a crossroad? Proceedings from NBiS: *15th International Conference on Network-Based Information Systems.* Melbourne, VIC: IEEE.

Instagram. (2012). What powers Instagram: Hundreds of instances, dozens of technologies [Web log post]. Retrieved from: http://instagram-engineering.tumblr.com/post/13649370142/what-powers-instagram-hundreds-of-instances

Intel & SAS Institute. (2015). Do you perceive your company as data driven? –The 2015 Nordic survey about big data and Hadoop. Retrieved from: http://nordichadoopsurvey.com/

Issenberg, S. (2013). How president Obama's campaign used big data to rally individual voters. *MIT Technology Review, 116(1),* 38–49.

Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM, 52(8),* 36–44.

Lahrmann, G., Marx, F., Winter, R., & Wortmann, F. (2011). Business intelligence maturity: Development and Evaluation of a Theoretical Model. Proceedings from IEEE: *44th International Conference on System Sciences.* Hawaii: IEEE

Leavitt, N. (2010). Will noSQL databases live up to their promise? *Computer, 43(29),* 12–14.

Lee, J., Lapira, E., Bagheri, B., & Kao, H. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters, 1,* 38–41.

Li, X., Song, J., & Huang, B. (2015). A scientific workflow management system architecture and its scheduling based on cloud service platform for manufacturing big data analytics. *International Journal of Advanced Manufacturing Technolgy,* 1–13.

Liikenne- ja viestintäministeriö. (2013). Big data Suomessa – Keskustelualoite.

Liikenne- ja viestintäministeriö. (2014). Big datan hyödyntäminen.

Lu, T., Guo, X., Xu, B., Zhao, L., Peng, Y., & Yang, H. (2013) Next big thing in big data: the security of the ICT supply chain. Proceedings from ASE/IEEE: *International Conference on Social Computing.* Washington D.C.: IEEE.

Madden, S. (2012). From databases to big data. *IEEE Computer Society, 3(16),* 4–6.

McAfee, A. & Brynjolfsson, E. (2012, October) Big data: The management revolution. *Harvard Business Review.*

Merriam, S. B. (2014) *Qualitative Research: A Guide to Design and Implementation.* San Francisco, Wiley.

Ministry of Employment and the Economy. (2013) 21 paths to a Frictionless Finland – Report of the ICT 2015 Working Group.

Murdoch, T. B., Detsky, A. S., The inevitable application of big data to health care. *The Journal of the American Medical Association, 309(13),* 1351–1352.

Morse, J. M., Barret, M., Mayan, M., Olson, K., & Spiers, J. (2002) Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1(2),* 13–22.

Olson, M. (2010) HADOOP: Scalable, flexible data storage and analysis. *International Journal of Electronic Commerce, 1 (3),* 14–18.

Oxford Economics. (2011). The new digital economy.

Pokorny, J. (2013) NoSQL databases: a step to database scalability in web environment. *International Journal of Web Information Systems, 9(1),* 69–82.

Pousttchi, K. & Huffenbach, Y. (2014). Engineering the value network of the customer interface and marketing in the data-rich retail environment. *International Journal of Electronic Commerce, 18(4),* 17–41.

Powered by Spark. (2015, September 8). Retrieved from: https://cwiki.apache.org/confluence/display/SPARK/Powered+By+Spark

Processing real-time events with Apache Storm. (n.d.). Retrieved from: http://hortonworks.com/hadoop-tutorial/ingesting-processing-real-time-events-apache-storm/

Rüping, S. (2015). Big data in medicine and healthcare. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 58(11),* 794–798.

Saarelainen, A. (2015, January). Ison datan kalastajat. *Talouselämä*.

Shilov, M. (2014, Jun 16). Where is NoSQL practically used? [Web log post]. Retrieved from http://scraping.pro/where-nosql-practically-used/

Spark - Lightning-fast cluster computing. (n.d.). Retrieved from: http://spark.apache.org/

TIVIT. (2011). Strategic research agenda for data to intelligence. Myllymäki, P., Ahtikari, J., Puolamäki, K., Carlsson, C.

Tuomi, J. & Sarajärvi, A. (2009). *Laadullinen tutkimus ja sisällönanalyysi*. Helsinki, Tammi.

Sami Sahala, Rauno Saarnio and Pentti Kurki

Välimäki, M. (2005) *The rise of open source licensing - A challenge to the use of intellectual property in the software industry*. Helsinki, Turre Publishing.

Wiener, J. & Bronson, N. (2014, October). Facebook's top open data problems [Web log post]. Retrieved from https://research.facebook.com/blog/1522692927972019/facebook-s-top-open-data-problems/

World Economic Forum. (2014). The global Information technology report 2014 – Rewards and risks of big data.

Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*

Yeoh, W. & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of Computer Information Systems, 50(3),* 23–32.

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010) Spark: Cluster computing with working sets. Proceedings from *2nd USENIX conference on Hot topics in cloud computing.* Berkley, CA: USENIX Association.

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology, 5(3),* 38–55.

Zikopoulos, P., deRoos, D., Bienko, C., Buglio, R., & Andrews, C. (2015). *Big data beyond the Hype –A Guide to conversations for today's data center.* New York, USA: McGraw Hill Education

Interviews

Company A (Head of Analytics). 6.5.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company B (Head of Arcitecture). 8.5.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company C (Senior Research Engineer). 21.5.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company D (ICT Architect). 26.5.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company E (VP, Customer Insight and Analytics). 5.6.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company F (Head of Big Data & Analytics). 24.6.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company G (Data Scientist). 17.8.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company H (Technical Director). 24.8.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company I (Director of Customer Knowledge). 17.9.2015, interviewed by Mirjamaria Petäjäniemi. The recording is in the writer's possession.

Company J (Chief Development Officer). 14.10.2015, interviewed by Mirjamaria Petäjäniemi. The notes of the interview are in the writer's possession.

Appendix A: The Interview Form

Aalto University
School of Business

Interview form

27 April 2015

Big Data projects in Finnish companies

1. Introduction

Company	
Industry	
Number of employees	
Turnover	
Interviewee	
Position	
Date and time	

(Asking for permission to record the interview)

(Introducing the study)

(Discussing the confidentiality issues)

1.1. Background information of the company

- Business model
- Ownership structure
- Mission, vision, values
- History
- Market share
- Current financial position
- Main competitors
- Main products

1.2. Please, describe your role in the company and your link to big data and the big data projects in your company.

2. Defining big data

2.1. How would you define the term big data?

3. Describing the previous / current projects

Aalto University
School of Business

Interview form 27 April 2015

3.1. Could you describe how you are using big data in your company?

3.2. How about your competitors, is this approach common in your industry?

3.3. What kind of data are you utilizing?

3.4. Please, describe the big data architecture you are currently using?

3.5. Why have you chosen to build this kind of architecture?

3.6. Is it common in your industry to use these technologies?

3.7. What parts of the data processing and analysis process have you outsourced and what are you doing in-house?

4. Outcomes of the previous / current projects

4.1. Which statement best describes the value you have seen from your big data efforts?

We have measured positive top- and bottom-line impact.	
We have measured top-line impact only.	
We believe that we have become more effective, but can't measure top-line impact.	
We have measured a cost reduction only.	
We believe that we have become more efficient, but cannot measure impact.	
We have gained more insight.	

4.2. How are you assessing and measuring the results of your big data projects?

4.3. What kind of benefits have you reaped this far?

4.4. How significant benefits have you achieved in the following areas (scale: 1 – 5, where 1=no benefits and 5=very significant benefits)

Benefit	Score
Cost savings	
New sources of revenue	
More efficient processes	
New products or business models	
More fact-based decision making	
Increased forecasting accuracy	
More efficient advertising	
Improved understanding of the customers	

Aalto University
School of Business

Interview form 27 April 2015

Improved understanding of the market	
Increased automation	
Improved quality	
Improved customer satisfaction	
Improved fraud detection and failure prevention	
Other:	

4.5. Overall, would you describe the big data efforts of your company as successful?

4.6. What would you say are the most important factors explaining your success (/failure)?

4.7. How satisfied on a scale from 1 to 5 (1=very dissatisfied and 5=very satisfied) you are with the results of your big data efforts?

4.8. What aspects most affect your satisfaction/dissatisfaction?

5. Future plans

5.1. Can you identify data related to your company that you are not utilizing yet but that could potentially be valuable?

5.2. Can you identify other kind of future possibilities in big data utilization for your company?

5.3. Do you already have some plans for further development?

5.4. Let's make a SWOT-analysis considering the further possibilities for your company in utilizing big data. What do you recognize as the most important strengths and weaknesses of your company? What about the main external opportunities and threats?

Strengths	**Weaknesses**
Opportunities	**Threats**

Appendix B: Summary of the results

Aalto University
School of Business

25.10.2015

Big data in Finnish companies

Summary of the interviews

1. Background information of the interviews and companies

Background information of the interviews

Number of interviews	10
Data collected between	6.5. – 14.10.2015
Durations of the interviews	25 – 75 min

The interviewed companies in chronological order

Company	Employees (2014)	Turnover (2014)	Industrial classification (TOL 2008)	Title of interviewee	Types of data	Uses of data
Company A	700 - 2 000	50 - 700	Computer programming, consultancy and related activities	Head of Analytics	Event data from users' actions in online applications Data from servers (e. g. advertising data)	Application design Customer relationship management Online advertising Market research
Company B	100 - 400	50 - 700	Publishing activities	Head of Architecture	Event data from users' actions in online applications Search engine queries Browsing data Location data	Service development Reporting Analytics services for customer companies Process control Data-based products

Aalto University
School of Business

25.10.2015

Company C	4 000 - 20 000	1 500 - 2 500	Manufacture of machinery and equipment	Senior Research Engineer	Sensor data Claims data Maintenance data Customer data Sales data Spare parts data Machinery usage data	Production planning Automating and optimizing maintenance Quality optimization Reports to the customer companies of the performance of their machinery Improving safety Optimizing sales activities
Company D	100 - 400	1 500 - 2 500	Arts, entertainment and recreation	ICT Architect	Event data from users' actions in online applications Usage data of offline products Social media content Open data (e. g. demographics)	Service development Automating processes
Company E	4 000 - 20 000	1 500 - 2 500	Publishing activities	VP, Customer Insight and Analytics	Customer data Market research data Click-stream data Online advertisement data Social media content	Customer relationship management Development of services Personalization of services Optimizing and automating advertising Optimizing sales activities Automating processes
Company F	4 000 - 20 000	7 000 - 10 000	Manufacture of computer, electronic and optical products	Head of Big Data & Analytics	Signaling data Location data	Optimization of infrastructure and capacity Reports and analytics to primary customer companies, other companies, and public sector Predicting and preventing quality and capacity problems End-user experience management Data-based products
Company G	100 - 400	1 500 - 2 500	Computer programming, consultancy and related activities	Data Scientist	Event data from users' actions in online applications	Application design Reporting and future scenarios to management

Aalto University
School of Business

25.10.2015

Company H	700 - 2 000	50 - 700	Land transport and transport via pipelines	Technical Director	Sensor data Location data	Management decision making support Performance measurement and improvement of vehicles and drivers Assessing maintenance needs Reclamation investigation
Company I	4 000 - 20 000	7 000 - 10 000	Wholesale trade	Director of Customer Knowledge	Customer data Transactional data Location data Market research data Click-stream data Open-data	Optimization of pricing Improving customer service Assortment optimization Future sales potential Optimization and personification of advertising
Company J	700 - 2 000	50 - 700	Arts, entertainment and recreation	Chief Development Officer	Server log data Event data from users' actions in mobile applications	Server usage optimization Service development Misuse detection

2. Defining big data

How would you define the term big data?

- 3V's: volume, velocity, variety (4: company A, company B, company C, company E)
- Capability to combine data from different sources (2: company C, company I)
- New technologies to efficiently acquire, store and analyse the data (2: company C, company E)
- Data volume is so large that it becomes difficult to manage with traditional approaches (2: company F, company G)
- Storing raw data for future unknown analytic needs (2: company H, company J)
- Management of data with high variety, especially semi structured data (company D)
- Extracting value from data to do better business (company I)

3. Competitors architecture and outsourcing

How about your competitors, is this approach common in your industry?

- I believe that our competitors are doing pretty similar things (5: company A, company B, company D, company G, company J)

Aalto University
School of Business

25.10.2015

- I believe that this area is a competitive advantage for us (5: company A, company C, company E, company G, company H)
- I would say that our solution is a big data solution but our competitors' have traditional solutions (company H)

Please, describe the big data architecture you are currently using?

- Hadoop (5: company A, company C, company D, company E, company F)
- Amazon's cloud storage and computing services combined with open source (3: company A, company B, company G)
- Centralized enterprise data warehouse (3: company B, company C, company D)
- Kafka (2: company D, company F)
- Open source scripting languages for automated parallel computing (2: company A, company G)
- Storm (company F)
- Customized tools developed together with partners e.g. a stream processing engine (company F)
- Pig (company G)
- Spark (company G)
- Microsoft SQL server (company H)
- Cognos (company J)

What parts of the data processing and analysis process have you outsourced and what are you doing in-house?

- Outsourced computational capacity, analytics both in-house and buying services from outside service providers (3: company A, company B)
- Outsourced computational capacity, analytics mainly in-house (3: company D, company E, company G, company I)
- Mainly trying to keep data in own hands, occasionally using analytics consultants, outsourcing maintenance and development of the systems (company C)
- Developing solutions together with partners, certain control points (company F)
- Own servers; outsourced development, control and maintenance of architecture; buying analytics as services (company H)
- Doing analytics together with consultants (Company J)

Aalto University
School of Business

25.10.2015

4. Outcomes of the previous / current projects

Which statement best describes the value you have seen from your big data efforts?

	A	B	C	D	E	F	G	H	I	J	Nb
We have measured positive top- and bottom-line impact.				X	X	X		X			4
We have measured top-line impact only.	X										1
We believe that we have become more effective, but can't measure top-line impact.									X	X	2
We have measured a cost reduction only.		X	X								2
We believe that we have become more efficient, but cannot measure impact.											0
We have gained more insight.							X				1

How are you assessing and measuring the results of your big data projects?

Company A: We can measure everything we want extremely precisely as everything is digital and quantified.

Company B: There are a couple of faces in the process that we measure. For example, how quickly we get the new data into use. Then we are always measuring investment compared to project benefits.

Company C: Projects vary a lot. I can't give a single metric that we use. We are measuring these as any research and development initiatives.

Company D: Output from the platform is measured by the development projects that use it. Metrics are project specific. They commonly concern, for example, improving margins, increasing sales, increasing number of customers, or increasing number of identified customers.

Company E: Metrics vary. We have common goals. The analytics team also has sales quotas like the sales team.

Aalto University
School of Business

25.10.2015

Company F: We build a business case for each use case in a very early stage. There has to be some kind of metric right from the start. There has to be an idea of how we can save in some internal process or much more importantly how this can create benefits for our customers.

Company H: There has to be savings in fuel consumption or maintenance of the vehicles.

Company I: It varies a lot from feedback from the managers to return on investment quantified in euros. Often there are also cost savings if we can automate or improve the process.

Company J: We brought people from different business functions together and discussed what kind of benefits we could get and metrics for these.

What kind of benefits have you reaped this far?

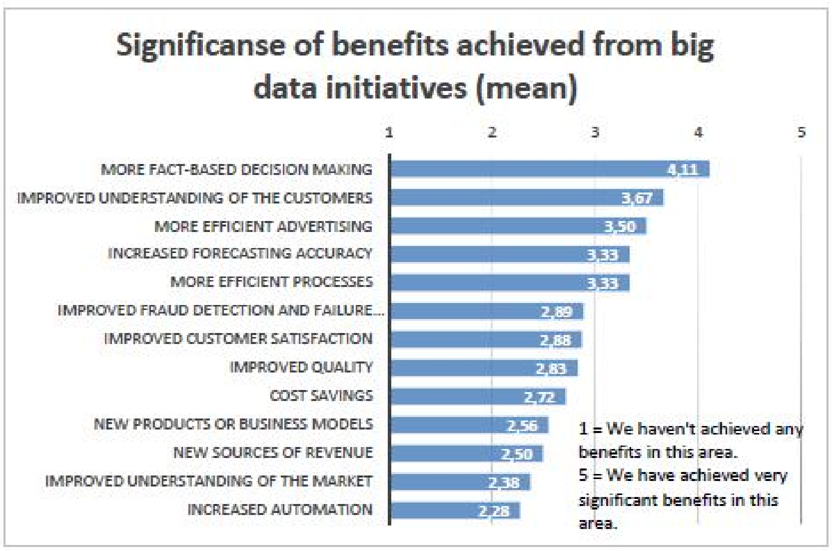
- Improved marketing (more relevant, personalized, cost savings) (3: company E, company I, company J)
- Improved efficiency of internal processes (2: company D, company F)
- Improved user experience (2: company F, company J)
- Improved understanding of the customers and becoming more customer-centric (3: company I, company J)
- Improved customer retention and turnover from applications (company A)
- Able to handle diverse data and large data volumes, capability to more sophisticated analysis (company B)
- Improved understanding and visibility to the life-cycle of the machine, optimization of quality, capability to predict demand, improved safety of the products (company C)
- Ability to response to changes in real time, ability to personalize services for each individual customer (company E)
- Ability to measure customer satisfaction in real time and even predict it, ability to predictively manage infrastructure, streamlining of architecture (company F)
- Ability to focus application development activities more efficiently (company G)
- Ability to optimize assortment, ability to optimize pricing (company I)

Aalto University
School of Business

25.10.2015

How significant benefits have you achieved in the following areas (scale: 1 – 5, where 1=no benefits and 5=very significant benefits)

Benefit	A	B	C	D	F	G	H	I	J	Mean	Sd. Error
More fact-based decision making	4	4	3	5	4	5	5	2	5	4,11	1,05
Improved understanding of the customers	5	4	3	5	3	5	1	2	5	3,67	1,50
More efficient advertising	5	4,5	NA	5	NA	4	1	1	4	3,50	1,76
Increased forecasting accuracy	4	3	2	4	4	4	4	1	4	3,33	1,12
More efficient processes	4	2	3	3	4	5	3	2	4	3,33	1,00
Improved fraud detection and failure prevention	2	3	3	1	4	3	2	3	5	2,89	1,17
Improved customer satisfaction	3	NA	2	4	4	1	4	1	4	2,88	1,36
Improved quality	2	3	2,5	3	3	2	4	2	4	2,83	0,79
Cost savings	1	3,5	3	2	3	1	5	2	4	2,72	1,35
New products or business models	2	4	2	4	3	1	1	2	4	2,56	1,24
New sources of revenue	1	3,5	2	4	2	3	1	2	4	2,50	1,17
Improved understanding of the market	3	3	NA	2	2	3	1	2	3	2,38	0,74
Increased automation	1	1,5	1	4	4	3	1	1	4	2,28	1,44



Aalto University
School of Business

25.10.2015

Overall, would you describe the big data efforts of your company as successful?

Company A: Yes, but I don't believe that we have done anything revolutionary compared to our competitors. We have achieved the reference level. The goal should be to get advantage compared to others.

Company B: Yes and no. We have completed the projects we have started ok, but we have always had problems at first with the technologies.

Company C: We have gained internal benefits and in some cases there have been also benefits to customers. I would say that we will at least be successful in the future.

Company D: Surprisingly successful, we have been able to make the business-side people interested of these areas.

Company E: I would say that it is starting to look like success. The benefits don't come immediately.

Company F: The efforts have been successful this far. The scale could go up. In my opinion we haven't done enough.

Company G: The projects have almost always gone wrong at first. The tools are still quite young. They require a lot of technological know-how and we haven't always had skilled enough people available. It has been learning by failing and sometimes things have taken more time than we would have hoped.

Company H: I'm satisfied. You would always hope that you had done better but we have achieved something.

Company I: The customers have not benefited yet. The best indicator of incipient internal success is that the demand for analytics is higher than the supply. We have started a marathon and run perhaps a kilometre but we are well on track.

Company J: Yes, in what we have done this far.

What would you say are the most important factors explaining your success?

- Skilled people with right mentality (4: company A, company B, company E, company G)
- Support from top management (2: company C, company E)
- Support from the business side (2: company C, company D)
- Finding good partners (2: company C, company F)
- Being able to successfully train people to use the technology (company B)
- Digital product makes acquiring data easy (company A)

Aalto University
School of Business

25.10.2015

- Experimental approach, not taking too big bites at once (company C)
- Looking for applicable approaches outside own industry (company F)
- Starting to investigate the opportunities in an early stage (company F)
- The need and motivation came from the business (company H)
- Being able to successfully implement new technology (company I)

How satisfied on a scale from 1 to 5 (1=very dissatisfied and 5=very satisfied) you are with the results of your big data efforts?

A	B	C	D	E	F	G	H	I	J	Mean
4	3	3	NA	3,5	3	4	4	1,5	2	3,11

What aspects most affect your satisfaction?

- Doing things in a modern way (company A)
- Ability to openly share information and open culture (company A)
- We have been able to quickly ramp this up (company E)
- Strong demand for analytics from the business people (company E)
- We have been able to raise this to the corporations strategy (company I)

/dissatisfaction?

- Having not yet achieved everything (2: company C, company F)
- Difficulty of changing the ways of doing things (company D)
- Having not achieved as much as could have been possible (company E)
- Suboptimization and product centeredness in the organization eating up the benefits (company E)
- Difficulties with the young technologies (company G)
- Time consuming to develop competences in a big company (company F)
- Not putting full resources and focus on analytics development (company J)

5. Future plans

Can you identify data related to your company that you are not utilizing yet but that could potentially be valuable?

- Data from sources outside the company (2: company D, company H)
- Social media (2: company E, company G)

Aalto University
School of Business

25.10.2015

- Data from different items and devices connected to the web in the future (IoT) (company F)

Can you identify other kind of future possibilities in big data utilization for your company?

- Developing more predictive models (2: company A, company H)
- Real time analysis (2: company H, company G)
- Further development of the digital offering (company B)
- Further development of analytics and internal process control (company B)
- Combining data from different sources (company C)
- Finding good partners and developing co-operation with them (company D)
- Introducing data-based products (company E)
- Scaling of the current competences throughout the company (Company F)
- More elaborate analytics as the technologies develop (company G)
- Machine learning (company H)

Let's make a SWOT-analysis considering the further possibilities for your company in utilizing big data. What do you recognize as the most important strengths and weaknesses of your company? What about the main external opportunities and threats?

Strengths

- Wide access to large amounts of data (5: company C, company E, company F, company G, company I)
- Strong know-how (4: company A, company B, company F, company G)
- Good tools (company A)
- Born-digital company (company B)
- Ability to productize analytics (company B)
- Harmonized and standardized processes and supporting IT-systems (company C)
- Excising data analytic culture (company D)
- Large amount of historic data that can be utilized in training and testing models (company H)

Weaknesses

- Not being agile enough in development and implementation (4: company D, company E, company F, company I)

Aalto University
School of Business

25.10.2015

- To strong process orientation (company A)
- We are only in the beginning (company C)
- Not having the required know-how in-house (company H)
- Legacy IT-systems (company J)

Opportunities

- Combining data from different business functions (company A)
- Automated decision making (company B)
- Developing more data-based products (company B)
- Global visibility to data through market position and global IT-systems (company C)
- Availability of data (company D)
- Wide opportunities to add intelligence to all products and solutions (company F)
- Virtualization of the whole industry (company F)
- Utilization of text analytics (company G)
- Combining browsing data with the current data sources (company I)
- Technological development (company J)

Threats

- More agile competitors are faster to develop competence (3: company D, company F, company I)
- Crisis with data safety or privacy (2: company B, company G)
- People with critical know-how leaving the company (2: company H, company J)
- Analytics becomes too focused on small details while the whole industry changes (company A)
- Changes in privacy legislation (company B)
- Not able to invest enough resources to see things through properly in long term (company C)
- The capabilities are successfully developed but the processes remain similar (company D)
- Partner lock-in (company H)
- Difficulty to find good data scientists (company J)