OULUN YLIOPISTO
UNIVERSITY of OULU

TEKNILLINEN TIEDEKUNTA

BIG DATA IN FINNISH FINANCIAL SERVICES

Mikko Laurila

INDUSTRIAL ENGINEERING AND MANAGEMENT

Bachelor’s thesis

November 2017

ABSTRACT FOR THESIS

University of Oulu Faculty of Technology

Degree Programme (Bachelor's Thesis, Master's Thesis)		Major Subject (Licentiate Thesis)	
Industrial Engineering and Management			
Author		Thesis Supervisor	
Laurila, Mikko		Kropsu-Vehkaperä H, D.Sc.	
Title of Thesis			
Big data in Finnish financial services			
Major Subject	Type of Thesis	Submission Date	Number of Pages
	Bachelor's Thesis	November 2017	26 p.

Abstract

This thesis aims to explore the concept of big data, and create understanding of big data maturity in the Finnish financial services industry. The research questions of this thesis are "What kind of big data solutions are being implemented in the Finnish financial services sector?" and "Which factors impede faster implementation of big data solutions in the Finnish financial services sector?".

Big data, being a concept usually linked with huge data sets and economies of scale, is an interesting topic for research in Finland, a market in which the size of data sets is somewhat limited by the size of the market. This thesis includes a literature review on the concept of big data, and earlier literature of the Finnish big data landscape, and a qualitative content analysis of available public information on big data maturity in the context of the Finnish financial services market.

The results of this research show that in Finland big data is utilized to some extent, at least by the larger organizations. Financial services specific big data solutions include things like the automation of applications handling in insurance. The most clear and specific factors slowing the development of big data maturity in the industry are the lack of competent work-force and new regulations compliance projects taking development resources. These results can be used as an overview of the state of big data maturity in the Finnish financial services industry. This study also lays a solid foundation for further research in the form of conducting interviews, which would provide more in-depth data.

Additional Information

TIIVISTELMÄ OPINNÄYTETYÖSTÄ Oulun yliopisto Teknillinen tiedekunta

Koulutusohjelma (kandidaatintyö, diplomityö) Tuotantotalous		Pääaineopintojen ala (lisensiaatintyö)	
Tekijä Laurila, Mikko		Työn ohjaaja yliopistolla Kropsu-Vehkaperä H, TkT	
Työn nimi Big data rahoitusalalla Suomessa			
Opintosuunta	Työn laji Kandidaatintutkielma	Aika Marraskuu 2017	Sivumäärä 26 s.

Tiivistelmä

Tämän työn tavoitteena on selvittää big data –käsitettä sekä kehittää ymmärrystä Suomen rahoitusalan big data –kypsyydestä. Tutkimuskysymykset tutkielmalle ovat "Millaisia big data –ratkaisuja on otettu käyttöön rahoitusalalla Suomessa?" sekä "Mitkä tekijät hidastavat big data –ratkaisujen implementointia rahoitusalalla Suomessa?".

Big data käsitteenä liitetään yleensä valtaviin datamassoihin ja suuruuden ekonomiaan. Siksi big data onkin mielenkiintoinen aihe tutkittavaksi suomalaisessa kontekstissa, missä datajoukkojen koko on jossain määrin rajoittunut markkinan koon myötä. Työssä esitetään big datan määrittely kirjallisuuteen perustuen sekä esitetään yhteenveto big datan soveltamisesta Suomessa aikaisempiin tutkimuksiin perustuen. Työssä on toteutettu laadullinen aineistoanalyysi julkisesti saatavilla olevasta informaatiosta big datan käytöstä rahoitusalalla Suomessa.

Tulokset osoittavat big dataa hyödynnettävän jossain määrin rahoitusalalla Suomessa, ainakin suurikokoisissa organisaatioissa. Rahoitusalalle erityisiä ratkaisuja ovat esimerkiksi hakemuskäsittelyprosessien automatisointi. Selkeimmät big data –ratkaisujen implementointia hidastavat tekijät ovat osaavan työvoiman puute, sekä uusien regulaatioiden asettamat paineet kehitysresursseille. Työ muodostaa eräänlaisen kokonaiskuvan big datan hyödyntämisestä rahoitusalalla Suomessa. Tutkimus perustuu julkisen aineiston analyysiin, mikä osaltaan luo pohjan jatkotutkimukselle aiheesta. Jatkossa haastatteluilla voitaisiinkin edelleen syventää tietämystä aiheesta.

Muita tietoja

CONTENTS

1 INTRODUCTION

Big data has been causing a lot of buzz for quite some years now, and while the collection of huge data sets has faced a fair share of criticism, it has also helped create some incredible services that could not have been imagined of ten years ago. It is widely thought that big data will be one of the major forces disrupting business across industry sectors through the near future (Manyika et al., 2011), and that is why a deep understanding of the concept is needed.

In this thesis, big data is researched in two different ways: first, a literature review of the definition of big data, its development and technological maturity in the Finnish industry is conducted. The focus is not on the technical details nor answering the question "how can we analyze big data sets?", but rather on reaching a better understanding of the business potential and industry disrupting implications of big data. After the literature review, publicly available data is analyzed through methods of qualitative content analysis to dig deeper into the present state and future of big data solutions in the Finnish financial services market.

While recent literature on big data is plentiful, finding scientific publications of big data in a Finnish context proved quite difficult. This may have something to do with the possible immaturity of Finnish big data –capabilities and the relatively small size of the Finnish market. It also demonstrated a clear need for a study that creates a clear summary of the big data maturity of the Finnish market. Three credible and interesting sources of knowledge were found: a report of big data, its adaptation and potential in Finland by Alanko and Salo (2013) a report laying out a big data strategy for Finland by Rastas and Asp (2014) and a report on big data related business opportunities and potential for increasing the efficiency of public services by Antikainen et al. (2016). Each of these reports were ordered by the Finnish ministry of transport and communications, so clearly the Finnish government has demonstrated an interest in the concept as well. These three sources were used to develop an elementary understanding of the big data climate in Finland.

The literature review lays a solid foundation for conducting the final study through defining the terms at hand and broadening the understanding of big data related problems and opportunities. The analysis of publicly available data offers concrete examples of big data related solutions already implemented, and a contemporary view of the future developments in big data related business in the financial services industry in Finland. Finally, the literature review and the results of the empiric study are summarized into conclusions and recommendations for further research.

2 BIG DATA

2.1 Defining big data

Even though data volumes have been skyrocketing throughout the 21st century, big data as a term became widespread as recently as 2012 (Gandomi & Haider, 2015). Confusion of the term's definition has been prevalent due to the sudden surge in interest, described by Ward and Barker (2013) as exponential since 2011, and the complexity of the concept. The fast acceptance by the public and private sectors and fast development of big data technologies meant that there was little time for the discourse to mature in the academic domain (Gandomi & Haider, 2015). Early big data literature came from numerous fields, leading to multiple definitions, which were ambiguous and often contradictory (Ward & Barker, 2013).

The complexity of big data stems from the multifaceted nature of the concept. There are two ideas that big data is predominantly associated with: data storage and data analysis. (Ward & Barker, 2013) The advance of technology has made the pace at which data is produced such overwhelmingly fast, that data sets are not usually static in size anymore, but rather dynamic. The absorption of complementary data collection, introduction of previously archived data and streamed data arriving from multiple sources cause the contents of data sets to constantly change. (Hashem et al., 2015) Additionally, big data sets no longer consist of only structured data, but also unstructured data like audio, images and video, which Gandomi and Haider (2015) describe as the largest component of big data. These factors bring up the need for advanced and ground-breaking tools for both data storage and analysis, and in general bring a change to the way we should view data.

In addition to the complexity, a problem with defining big data has been the quantification of big data. How much data is big data? With rapidly advancing processing capabilities and data storage technologies this question becomes quite meaningless: the size of a big data set today might be tenfold in two years' time. Manyika et al. (2011) dodge this problem in a clever fashion in a widely cited big data report by defining big data as "the amount of data just beyond technology's capability to store, manage, and process efficiently." This and many other definitions are found through literature that either do or do not address the quantifiable size of big data sets.

First, big data was used to describe datasets that took a lot of storage space. (Russom, 2011) With increasing interest and research on big data, the definitions used began addressing other problematic characteristics of today’s big data, such as the velocity at which the data sets are growing and the variable forms of data collected and analysis tools required. Many companies coined their own definitions of big data, focusing on different aspects of the concept (Ward & Barker, 2013).

One such definition for big data was provided by Oracle: big data is the derivation of value from traditional relational database driven business decision making, augmented with new sources of unstructured data (Ward & Barker, 2013). The core idea behind this definition is inclusion: including additional data sources for augmentation of existing operations is the core of big data. A more concrete quantification of big data was provided by Intel: big data organizations “generate a median of 300 terabytes (TB) of data weekly” (Ward & Barker, 2013). Another definition is provided by Microsoft: “Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information” (Ward & Barker, 2013). While there were and still are many individuals and organizations trying to fine-tune the definition for big data, one definition has been cited and expanded on more than any other: the three V’s.

2.1.1 The three V’s

Today, the most common way of defining big data is through the three V’s, data *volume*, data *variety* and data *velocity.* First introduced by Doug Laney in 2001 in his research note “3D Data Management: Controlling Data Volume, Velocity, and Variety”, the three V’s has become what is closest to the industry standard way of defining big data. The reason that the three V’s has become such a prevalent way of defining big data, is that it helps you understand that it is not only the size of the data set that constitutes big data. (Russom, 2011)

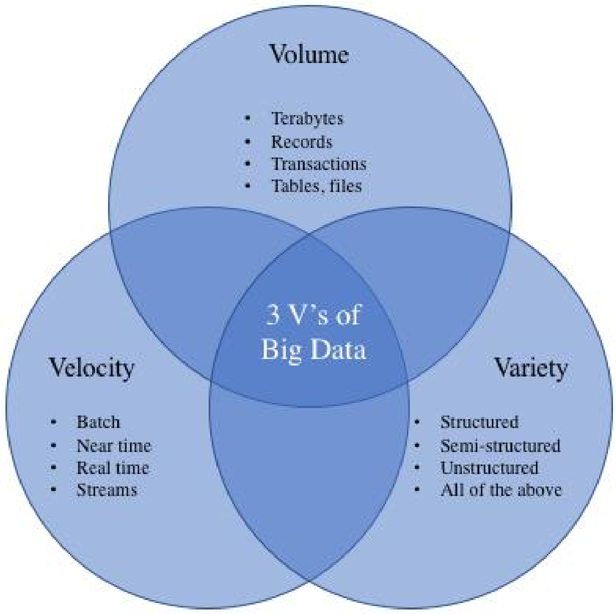

Figure 1. Characteristics of big data (modified from Russom, 2011)

The first of the three V's stands for data *volume*, the amount of data available to an organization (Kaisler et al., 2013). According to Russom (2011) it is obvious that the volume of data is the primary defining attribute to big data. On the other hand, with today's plentiful storage capacity and computing power, one could argue that it is not the volume of data that causes the most problems. Two data sets of the same size might require different types of advanced analysis technologies, e.g. video data versus tabular data. (Gandomi & Haider, 2015)

According to Gandomi and Haider (2015) big data datasets were reported in the sizes of terabytes and petabytes, (1024 gigabytes and 1024 terabytes, respectively.) Kaisler et al. (2013) classified data volumes in the range of exabytes (1024 petabytes) and beyond as big data. Clearly there is no consensus on the exact volume threshold a data set must cross to be considered big data. A boundary one could propose would be the capabilities that a private person has at one's disposal, so that if the volume of a data set is too big for a personal computer to handle, it could be considered big data.

The second V of the three V's, *variety*, is used to describe the structural heterogeneity in a dataset (Gandomi & Haider, 2015). There is a greater variety of data sources than ever before, for example logs, clickstreams and social media (Russom, 2011). It follows that data sets that are analysed no longer consist solely of structured data, like spreadsheets and relational databases, but also include more complex semi-structured and unstructured data, like text, images, audio and video (Gandomi & Haider, 2015).

Obviously semi-structured and unstructured data have always been around, but the new technologies that enable leveraging such data for analysis is where new innovations arise (Gandomi & Haider, 2015). The problems such technologies need to address have to do with incompatible data formats, inconsistent data semantics and non-aligned data structures (Kaisler et al., 2013).

The third of the three V's, the *velocity*, can be thought of as the data generation or delivery (Russom, 2011). Data velocity is not only a problem in terms of bandwidth management, but also in terms of ingesting the data (Kaisler et al., 2013). According to Gandomi and Haider (2015), the main driver of the increase in data velocity has been the proliferation of digital devices such as smartphones and sensors. Retailers can be processing more than one million transactions per hour, which generates incredibly big datasets for real-time analysis (Gandomi & Haider, 2015).

The three V's listed here give a pretty simple but useful description of big data and its characteristics. Thinking of big data in terms of the three V's helps you grasp the complexity and gravity of the concept. Widely considered as the standard definition for big data today, the 3V's is the way we think of big data and how we define it throughout the rest of this thesis.

2.1.2 Other measures of bigness of data

Like the term big data, the 3V's is still evolving and experiencing expansions in terms of its definition. Since the 3V's became commonplace, many new V's have been proposed to describe characteristics/problems of big data, such as value, veracity and variability.

Data *value* is a defining attribute of big data introduced by Oracle. Big data often has a low value relative to its volume. Creating high value is based on analysing large volumes of such data. (Gandomi & Haider, 2015) Hashem et al. (2015) define data value as the process of discovering huge hidden values from large datasets with various types and rapid generation. They also claim that value is the most important aspect of big data. Data *veracity* is a big data V coined by IBM. Data veracity is used to refer to the inherent unreliability in some sources of data. (Gandomi & Haider, 2015) This is especially evident in data collected from social media. For example, customer sentiments, being based on human judgement, are uncertain in nature, but still contain valuable information. Another more recently surfaced V is data *variability*. Data variability is an additional dimension of big data, introduced by SAS. (Gandomi & Haider, 2015) Big data velocity is often not consistent, and variability is used to refer to this variation in the data flow rates. These more recent V's seem quite specific and case-sensitive. Interesting, but not so important for grasping the general concept of big data.

2.2 Analyzing big data

Obviously just storing big data sets is not of great value without conducting meaningful analysis. The ability to analyze big data to develop actionable information, is how the value arises from big data. (Kaisler et al., 2013) According to Hashem et al. (2015), scalable analysis algorithms are required to obtain timely and useful information from large amounts of data. Kaisler et al. (2013) also identify as the most critical issue in analytics the scalability of the analytic process when the data set grows by orders of magnitude. The growth rate in the amount of data collected is outpacing our ability to (1) design appropriate systems for the handling of data and (2) analyze the data to extract relevant meaning for decision making (Hashem et al., 2015).

Another problem with big data analytics pointed out by Labrinidis and Jagadish (2012) is the lack of coordination between the databases that host the data and provide SQL querying, and the analytics packages that perform data mining and statistical analyses. This leaves today's analysts with a tedious process of exporting data from the database using a SQL process, analyzing it and bringing the data back into the database.

Solutions to these problems have been found through a technology called cloud computing. According to Hashem et al. (2015), "Cloud computing is a model for allowing ubiquitous, convenient, and on-demand network access to a number of configured computing resources (e.g., networks, server, storage, application, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". Basically, this means that cloud service providers can offer on-demand platforms as a service to organizations that need for example computing power or applications to analyze their big data. Today, complex and massive-scale computing is performed through cloud computing. Cloud computing eliminates the need for expensive hardware, software and dedicated storage space. (Hashem et al., 2014)

Big data focused professionals have been called “data scientists” since the coining of the term by Jeff Hammerbacher and DJ Patil in 2008 (Patil, 2011) According to Manyika et al. (2011), the techniques developed to analyze and visualize big data draw from several fields including applied mathematics, statistics, computer science, and economics. The Venn diagram below demonstrates the interdisciplinary nature of the expertise required for a data scientist.

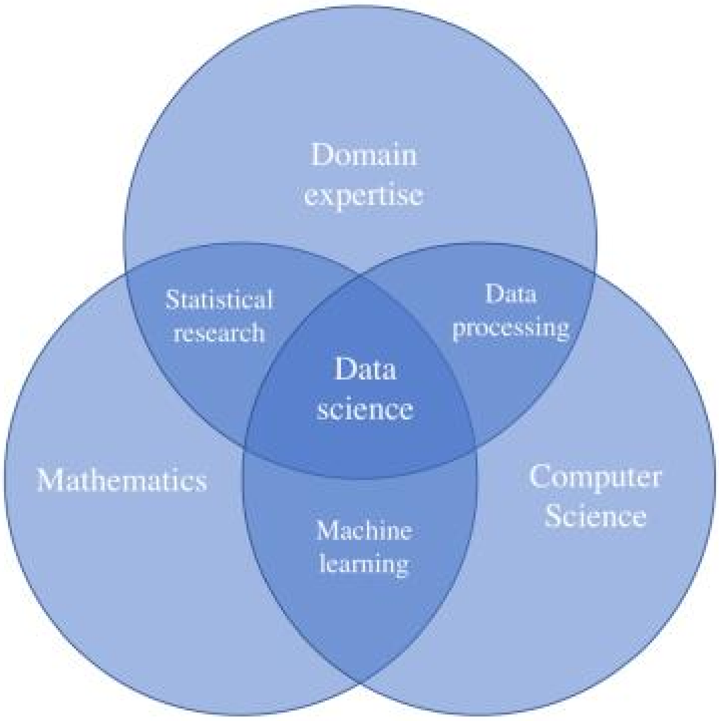

Figure 2. Interdisciplinary nature of data science (modified from Palmer, 2015).

Dhar (2013) also concludes that an integrated skill set of mathematics, statistics, machine learning, artificial intelligence, databases, optimization and an understanding of problem formulation for engineering effective solutions makes a data scientist. Dhar (2013) goes even deeper into the technical skills required, such as machine learning, XML and text mining, making it quite clear that becoming a big data professional is not something achieved over-night or through a weekend training.

2.3 From big data problems to big data value

There are two types of challenges posed by big data: managing data of unimaginable scale efficiently, an engineering challenge, and finding and combining data into meaningful information relevant to one's concerns, a semantics challenge (Bizer et al., 2011). In addition to these challenges, irresponsible use of big data may pose a threat to privacy in the future. It has been suggested that analysis of government records, phone logs, health records and social media interactions will create a new set of privacy incursions rather than better tools and services. (Kaisler et al., 2013)

In a survey shown by Russom in his 2011 report, when 325 professionals were asked if big data is considered mostly a problem or mostly an opportunity in their organization, only 30 % of the responses claimed that big data was mostly seen as a problem. 70 % of the respondents considered big data as an opportunity. This is telling of the potential that was seen in big data already in 2011, and the belief in organizations conquering these big data related challenges.

Through solving the engineering and semantics challenges of big data, organizations have interesting opportunities for capturing value. The McKinsey Global Institute (2011) identified five broadly applicable ways to create value through big data:

1. Creating transparency
2. Enabling experimentation to discover needs, expose variability, and improve performance
3. Segmenting populations to customize actions
4. Replacing/supporting human decision making with automated algorithms
5. Innovating new business models, products, and services

These five methods of value-creation can be utilized across industries, which means that there is no industry unchanged by big data. Every sector of the global economy has now been reached by big data, with early examples of leveraging data-driven strategies to innovate and capture value. (Manyika et al., 2011) The McKinsey Global Institute identify such industries as retail, financial services and insurance as key benefiters in utilizing big data, although regardless of industry big data provides possibilities for development of new business models, products and services. The McKinsey Global Institute estimated that in the private sector a retailer could increase its operating margin by more than 60 percent by fully utilizing big data. The organization that can best adapt and leverage new technologies surrounding the explosion of data volumes will undoubtedly achieve a sizeable competitive advantage. (Manyika et al., 2011)

In addition to improving existing business models and improving efficiency in existing industries, The McKinsey Global Institute vision that big data will also create entirely new industries and categories of companies, for example companies that analyze and aggregate industry data. This translates to a sizeable amount of new jobs and demand for professionals of data analytics and data science.

2.4 Big data in Finland

2.4.1 Big data implementation in Finland

Rastas and Asp (2014) conclude that Finland has great qualifications to rise to the challenges of big data. According to international studies, Finland is one of the most digitally advanced countries, especially in terms of technical qualifications. Even in the areas in which Finland is behind the curve, big data analytics provide great benefits. (Rastas & Asp, 2014)

Still, in the study conducted by Alanko and Salo (2013), most of the data and analytics companies they interviewed had not executed a single commercial implementation of big data solutions, even though they might possess an excellent big data product portfolio. Clearly Finnish companies already saw the potential in big data despite the lack of concrete examples, as over half of those surveyed expected big data to rise as a central component of present or future operations or bring a meaningful rise in efficiency.

Rastas and Asp (2014) mention Konecranes and Outotec as Finnish forerunners of big data and industrial internet. They also bring up the high-class expertise Finland has in using big data to optimize user experience. Forerunners like Rovio and Supercell have grown extremely fast, and some of this is due to the collecting of big data and using real-time data analytics to keep users interested. (Rastas & Asp, 2014)

Antikainen et al. (2015) conducted a study for the Finnish prime minister's office that surveyed 1 189 Finnish companies on their use of big data. Then, in 2015 almost a quarter of the companies surveyed collected or used big data in their operations. Companies in the energy, finance and insurance sectors were the clear forerunners. In average 4 % of the workforce in companies utilizing big data were employed in big data related positions. It seems as if big data is making its way into the Finnish industry, with large and medium size companies being clearly the more probable users of big data according to the study. Antikainen et al. (2016) also concluded that a large share of Finnish companies still do not realize the potential of big data.

The most important part of big data collected in Finnish companies is the data collected from the companies' sales systems and web pages. This data is mostly used for decision making, market analysis and new product development. (Antikainen et al., 2016) According to the study by Antikainen et al. (2016), the most developed industries in utilizing big data are energy (67 %), financial services (43 %), and ICT and logistics (34 %).

2.4.2 Impeders of big data development in Finland

One surprising finding in the study by Antikainen et al. (2016) was that when asked of the barriers for utilizing big data, almost 38 % of the Finnish companies surveyed answered that they have no need for big data. Of the companies that did see utilizing big data as worthwhile, lack of skilled work-force was the most serious obstacle, followed by the high costs relative to potential gains. There is also a shortage of programmers with big data related skills in Finland (Rastas & Asp, 2014).

Lack of know-how clearly is a critical impeder of utilizing big data in Finland, which is no different than what was found out in the global study by Manyika et al. (2011). Alanko and Salo (2013) point out that many data analytics focused programs were being founded in Finnish higher education, hopefully alleviating the problem. Rastas and Asp (2014) also emphasize the importance of fast and targeted educational measures in replying to the disruption caused by big data. It seems that there is a clear consensus that the lack of skilled big data professionals is the most serious impeder of big data implementation in Finland.

3 BIG DATA IN FINNISH FINANCIAL SERVICES

3.1 Conducting the study

Based on the literature, there was a clear need for deeper research into big data in the Finnish financial services industry. The following questions were not sufficiently answered by the literature review, and therefore were used as the research questions for the study:

1. What kind of big data solutions being implemented in the Finnish financial services sector?
2. Which factors impede faster implementation of big data solutions in the Finnish financial services sector?

The study was done as a qualitative content analysis of publicly available data, focused on finding both information describing the use of big data in the industry at the moment, and predictions of uses in the near future. The aim was to reach an understanding of what kind of big data opportunities have been capitalized on and what are seen as the next steps in the near future.

3.1.1 Choosing the data

Publicly available data of implemented big data solutions and big data trends in the Finnish financial services industry proved very scarce. There were a few channels through which some data on the subject was available:

1. The web sites of financial services companies
2. Web sites of technology consulting companies operating in the FS sector
3. The professional social media LinkedIn

The financial services companies (banks, insurance companies, pension funds) themselves understandably did not share a lot of information of their processes via their web site, and the LinkedIn pages of the companies were very hard to go through systematically. The web sites of technology consulting companies proved to be the most effective channel for interesting data on the subject. Through search engines and personal knowledge, the following technology consulting service providers that published information specific to the Finnish financial services sector on their web sites were identified: Accenture, Affecto, Avarea, BIGDATAPUMP, CGI, InsightsAtlas, PwC, and Tieto.

3.1.2 Analyzing the data

These web sites contained three different types of useful information: blog posts and ‘white papers’ on financial services specific subjects, and accounts of completed projects, often called ‘success stories’ or ‘customer cases’ by the companies. The blog posts and white papers were grouped together, because they were very similar in content and thus were analyzed in a similar fashion. The following types of accounts were searched for through the texts:

1. Big data solutions that have been implemented in the industry
2. Big data solutions that will probably be implemented in the near future
3. Factors impeding big data development.

The customer cases provided specific and useful mentions of solutions already implemented, on top of some more general statements on the same subject that were found in the blog posts and white papers. The most effective source of information on the big data implementations in near future were the blog posts and white papers, where the authors often described global frontrunner type companies and forecasted the implementation of similar methods in the Finnish financial services sector. Impeders of big data development were a bit harder to come by, but some information on these was also found in the blog posts and white papers.

The completed projects that the technology consulting companies described in their cases were categorized by the type of solution. Some of the cases were included in multiple categories, especially the broader scope projects completed. Table 1 summarizes these categorizations.

Table 1. Big data solutions types in the customer cases analyzed.

Solution type	# of cases in this subject	Case companies
Data visualization	3	Invesdor, KELA, Nordea
Internal cloud services	3	If, Etera, Varma
Privacy/Regulation compliance	3	If, Santander, Unnamed bank
Automation	2	If, Varma
Marketing/Social media analytics	2	Invesdor, Nordea
Advanced analytics	1	Varma
Data warehousing	1	Varma
Extract-transform-load system	1	KELA
Integrating multiple systems	1	If

The same methodology was used while going through the blog posts and white papers. Three types of accounts were looked for in the texts: descriptions of big data implementations already done, future targets for big data solutions, and descriptions of factors impeding developing big data solutions. Table 2 summarizes the categorization of big data implementations, table 3 summarizes the categorization of future targets for big data solutions, and table 4 summarizes the categorization of factors impeding big data implementation.

Table 2. Types of big data implementations found in the blog posts and white papers.

Type of big data implementation	# of mentions	Sources
Applications handling	2	Accenture, 2017b CGI, 2015
Asset allocation automation	1	Accenture, 2017a
Asset trading	1	Accenture, 2017b
Creating customer-specific services	1	Accenture, 2017a
Front-office operations	1	Accenture, 2017b

Table 3. Future big data solutions found in the blog posts and white papers.

Future big data solution	Source
Automating insurance price renegotiating	CGI, 2015
Automating black list checking	CGI, 2013
Blockchain utilization	PwC, 2017b
Fraud detection	CGI, 2013
Insurance risk calculation	CGI, 2013
Money laundering detection	CGI, 2013
Pay-as-you-drive –type insurance pricing	CGI, 2015
Risk calculations	CGI, 2013
Wearable computer utilization in insurance pricing	CGI, 2015

Table 4. Factors impeding big data implementation found in the blog posts and white papers.

Factor impeding big data implementation	# of mentions	Sources
Hesitation by top management	2	Accenture, 2017a PwC, 2017b
Shortage of talent	1	Accenture, 2017b
Weight of regulation compliance projects in development portfolios	1	PwC, 2017a

3.2 Findings

Accounts of completed big data projects were found through the study, resulting in automation, data visualization or advanced data analytics. Based on these accounts, at the moment big data is utilized by at least some of the larger organizations of the financial services industry, like insurance companies and pension funds. The use cases analyzed included things like applications handling automation and social media/marketing optimization. Most of the financial services industry customer cases that the technology consulting companies showed on their web sites were not directly related to big data. Instead, more often these companies had completed projects in regulations compliance, data warehousing, and systems integration/internal cloud services. This does not give strong evidence for a widespread, large-scale big data disruption happening in the Finnish financial services industry right now.

The future prospects for big data found through the study can be roughly categorized into two categories: automation of tasks formerly done manually, and creating completely new services, like big data based insurance pricing, or blockchain technologies. Out of the two categories, there were more mentions of automation possibilities than creation of new services, but this can be considered natural, as development of old processes is easier to forecast than the creation of new services.

Regarding impeders of big data development, there were three factors that were found in the study. Hesitation by the management of financial services industry companies occur twice. The second factor is a collection of privacy and data management related regulations, all having to be implemented within a short timeframe. The regulations mentioned in the customer cases are GDPR (BIGDATAPUMP, 2015), MifiD II (Accenture, 2017a), and Solvenssi II (Affecto, 2017a). Making the processes of these financial services industry companies compliant with the regulations is taking a lot of resources in the development portfolios of these companies. These regulation compliance projects were also seen in the customer cases previously analyzed in the study, further confirming their influence in the financial services industry. Lastly, the shortage of talent is something that is pointed to as the reason for difficulties in keeping up with the digital trends, with Accenture claiming 40 percent of employers reporting talent shortages (Accenture, 2017b).

4 CONCLUSIONS

In the literature review, first the concept of big data was defined through the three V's. Based on this definition, the big data landscape of the Finnish industry was studied. Then the scope was further focused on the Finnish financial services industry, by means of the empirical study. Overall, the results of the empirical study confirmed or strengthened many of the points previously laid out in the literature review, while also offering some additional in-depth information specific to the financial services industry.

In the study by Alanko & Salo (2013), most of the Finnish technology solutions providing companies that had big data products in their portfolio had not completed a single project. Now, in 2017, the empirical study showed that many such projects have been completed, in the financial services industry alone. If, in 2013, big data solutions were mostly produced in-house, now they are produced by technology consulting providers as well. The empirical study also strengthened the point made in the literature review, that in Finland the larger organizations are more probable users of big data than organizations of smaller scale. All but one of the 'customer cases' shown in the empirical study were done in very large institutions or companies. Antikainen et al. (2016) claimed that in Finland big data is mostly used for decision making, market analysis and new product development. With the findings of this study, one very important use case can be added to the list, automation. At least in the personal and pension insurance field big data has already been utilized for automating tasks formerly done manually. This is very concrete evidence of the industry transforming potential of big data solutions.

Through the literature review shortage of talented work-force was established as the main impeder of big data development in Finland. This point was further confirmed by additional evidence through the empirical study. In addition to this, a financial services specific impeder was found: the new regulations coming into effect concerning privacy and data management. With huge projects aiming to make organizations compliant with these regulations, it seems sensible that these projects are somewhat reducing the resources available for developing big data solutions. In the study by Antikainen et al. (2016), high perceived capital costs were another impeding factor of big data development. The empirical study offered no further information on the subject, making the cost-effectivity of the big data solutions implemented an interesting research question for possible further interview-based research.

Overall, the research questions “what are the big data solutions being implemented in the Finnish financial services industry?” and “which factors impede faster implementation of big data solutions in the Finnish financial services industry?” did get some interesting answers through the research. At the moment, it seems that automation of tasks, data visualization, advanced analytics and social media/marketing optimization are the most common implementations of big data solutions in the Finnish financial industry. The most notable impeders of big data development were the shortage of competent work-force, and a surge of new financial services specific regulations requiring development resources. This research crystallized into a preliminary overview of the big data landscape of the Finnish financial services industry, and gives a foundation for more specific and in-depth research in to the subject. The reliability of these results is somewhat hindered by the type of source material. As the customer case stories and blog posts analyzed are primarily written as advertising material, there is a possibility that they show things in an overtly positive light. Additionally, finding negative experiences or accounts of any type of hardships faced is nearly impossible using this kind of source material. This makes further research in the form of anonymous interviews a very attractive possibility. Further research could explore such topics as the cost-effectivity of the big data solutions implemented, the relative popularity of different big data technologies, and how often are the solutions developed in-house as opposed to being bought from outside consultants.

REFERENCES

Accenture, 2017. Digital disruption in Nordic wealth management [online document]. Available: https://www.accenture.com/no-en/insight-digital-disruption-nordic-wealth-management/ [referenced 13.11.2017].

Accenture, 2017. Nordic Financial Services Workforce Revolution [online document]. Available: https://www.accenture.com/fi-en/insight-nordic-financial-services-players-late/ [referenced 13.11.2017].

Affecto, 2017. If:n aktuaarilaskenta uuteen kuosiin [online document]. Available: http://www.affecto.com/fi-fi/tarinat/asiakas/ifin-aktuaarilaskenta-uuteen-kuosiin/ [referenced 13.11.2017].

Affecto, 2017. Kela tehostaa tietovarastoprosesseja ETL-järjestelmällä [online document]. Available: http://www.affecto.com/fi-fi/tarinat/asiakas/kela-tehostaa-tietovarastoprosesseja-etl-jarjestelmalla/ [referenced 13.11.2017].

Affecto, 2017. Santander's powerful and gentle consumer privacy protection [online document]. Available: http://www.affecto.com/insights/customer-case/santander-data-masking-solution/ [referenced 13.11.2017].

Affecto, 2017. Tietovarastoratkaisu varmistaa katkeamattoman raportoinnin Varman eläkekäsittelyssä [online document]. Available: http://www.affecto.com/fi-fi/tarinat/asiakas/tietovarastoratkaisu-varmistaa-katkeamattoman-raportoinnin-varman-elakekasittelyssa/ [referenced 13.11.2017].

Alanko M. & Salo I. (2013) Big data Suomessa. Liikenne- ja viestintäministeriön julkaisuja 25/2013.

Antikainen J., Eskelinen J., Koski H., Niemi T., Pajarinen M., Pyykkönen S. & de Vries M. (2016): Massadatasta liiketoimintaa ja tehokkaita julkisia palveluja.

Avarea, 2017. Case Invesdor [online document]. Available: https://avarea.fi/wp-content/uploads/2017/04/Avarea_Case_Invesdor.pdf [referenced 13.11.2017].

BIGDATAPUMP, 2015. Applying new EU data protection (GDPR) legislation, analysis and action plan [online document]. Available: http://www.bigdatapump.com/case-studies/eu-data-protection-legislation [referenced 13.11.2017].

Bizer C., Boncz P., Brodie M.L. & Erling O. (2012) The meaningful use of big data: four perspectives--four challenges. ACM SIGMOD Record 40(4): 56-60.

CGI, 2013. Erinomaista liiketoimintaa big datan avulla [online document]. Available: https://www.cgi.fi/sites/default/files/files_fi/white-papers/white_paper_erinomaista_liiketoimintaa_big_datan_avulla.pdf [referenced 13.11.2017]. 14p.

CGI, 2015. Pay as you drive & live – Näin analytiikka muuttaa vakuutusalaa [online document]. Available: https://www.cgi.fi/blogi/pay-as-you-drive-live-nain-analytiikka-muuttaa-vakuutusalaa/ [referenced 13.11.2017].

Chen H., Chiang R.H. & Storey V.C. (2012) Business intelligence and analytics: From big data to big impact. MIS quarterly 36(4).

Dhar V. (2013) Data science and prediction. Commun ACM 56(12): 64-73.

Gandomi A. & Haider M. (2015) Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage 35(2): 137-144.

Hashem I.A.T., Yaqoob I., Anuar N.B., Mokhtar S., Gani A. & Khan S.U. (2015a) The rise of "big data" on cloud computing: Review and open research issues. Inf Syst 47: 98-115.

InsightsAtlas, 2017. Case: Nordea [online document]. Available: https://insightsatlas.com/case-nordea/ [referenced 13.11.2017].

Kaisler S., Armour F., Espinosa J.A. & Money W. (2013) Big Data: Issues and Challenges Moving Forward. System Sciences (HICSS), 2013 46th Hawaii International Conference on: IEEE: 995-1004.

Kambatla K., Kollias G., Kumar V. & Grama A. (2014) Trends in big data analytics. Journal of Parallel and Distributed Computing 74(7): 2561-2573.

Labrinidis A. & Jagadish H.V. (2012) Challenges and opportunities with big data. Proceedings of the VLDB Endowment 5(12): 2032-2033.

Laney D. (2001) 3D data management: Controlling data volume, velocity and variety. META Group Research Note 6: 70.

Manyika, J., Chui, M., Brown, B. & Hung Byers, A. (2011) Big data: The Next frontier for innovation, competition and productivity. McKinsey Global Institute.

Mayring, P. (2000). Qualitative Content Analysis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [on-Line Journal], 1(2), 28 paragraphs.

McAfee A., Brynjolfsson E. & Davenport T.H. (2012) Big data: the management revolution. Harvard Business Review 90(10): 60-68.

Palmer, Shelly. Data science for the C-Suite. New York: Digital Living Press, 2015. Available: http://www.huffingtonpost.com/shelly-palmer/are-you-ready-for-data-sc_b_6844032.html

Patil D.J. (2011) Building Data Science Teams. O'Reilly Media, Inc.

PwC, 2017. Harva pankki valmis PSD2-direktiiviin ja avoimeen pankkitoimintaan [online document]. Available: https://uutishuone.pwc.fi/harva-pankki-valmis-psd2-direktiiviin-ja-avoimeen-pankkitoimintaan/ [referenced 13.11.2017].

PwC, 2017. Suomen finanssiala vaarassa jäädä globaalin kilpailun jalkoihin [online document]. Available: https://uutishuone.pwc.fi/suomen-finanssiala-vaarassa-jaada-globaalin-kilpailun-jalkoihin/ [referenced 13.11.2017].

Rastas T. & Asp E. (2014) Big datan hyödyntäminen. Liikenne- ja viestintäministeriön julkaisuja 20/2014.

Russom P. (2011) Big data analytics. TDWI best practices report, fourth quarter. 1-35.

Tieto, 2017. Eteran siirtyminen pilvipalveluihin alentaa toimintakustannuksia. [online document]. Available: https://www.tieto.fi/menestystarinat/eteran-siirtyminen-pilvipalveluihin-alentaa-toimintakustannuksia [referenced 13.11.2017].

Tieto, 2017. Varma uudistaa järjestelmiä ja siirtyy Tiedon yksityiseen pilveen [online document]. Available:https://www.tieto.fi/menestystarinat/varma-siirtyy-tiedon-yksityiseen-pilveen [referenced 13.11.2017].

Ward J.S. & Barker A. (2013) Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.