Evaluation of Big Data Requirements in the Insurance Sector

Masterarbeit

Zur Erlangung des akademischen Grades „Master of Science (M.Sc.)“ im Studiengang Wirtschaftswissenschaft der wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover

vorgelegt von

Name: Dohne Vorname: Denis Pascal

[redacted] [redacted] [redacted] [redacted]

Prüfer: Prof. Dr. M. H. Breitner

Hannover, den 02.10.2017

Table of content

1. Introduction

„Information is the oil of the 21st century, and analytics is the combustion engine." (Sondergaard, 2011). According to this reference, from the senior vice president of google, the impact of Big Data can be clarified. The amount and usage of information has never been as high as in the twenty first century. Despite the assumed potential of the digital era and its amount of information, Big Data has become the focus of academic and corporate investigations.

The relevance of Big Data has been pointed out as "management revolution" (McAfee et al., 2012, p. 62), "the fourth paradigm of science" (Strawn, 2012, p. 34) or "the next frontier for innovation, competition, and productivity" (Manyika et al., 2011). In summary, Big Data opens the field for new research methods and management implications to make strategic decisions more profitable. The study of Manyika et al. contributes of a lack in required skills until 2018 in managing the big amount of available information and making them effective (Manyika et al., 2011).

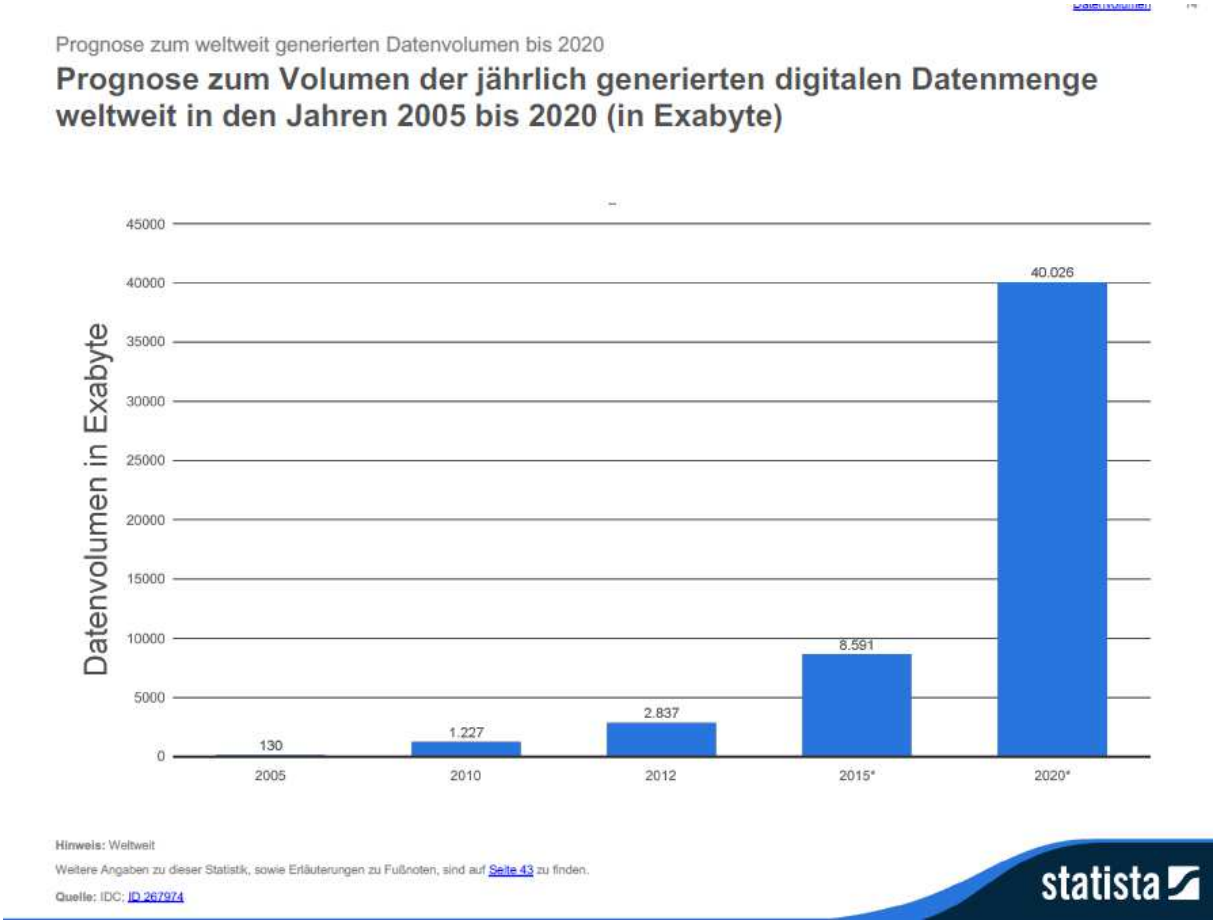

Figure 1: Prediction of generated Data Volume until 2020 (Statista, 2016)

On the one hand, the majority of literature and companies agreed on the relevance of information for future market competition, but on the other hand companies are still

struggling with the adequate implementation of Big Data solutions. Therefore, Big Data becomes the buzzword in IT management related to the digitalization of processes and business activities. As a result, it is even more fundamental to understand how Big Data can be used and which requirements are related to handle the ubiquity of data (Chen, Chiang & Storey, 2012, pp. 1168-1170).

This document provides an overview about the common concepts of Big Data and concentrates on the characteristics, which are important, while working with data. In context of the increasing amount of data sources and the different kinds of structure, the transfer into the company processes is getting more difficult. To explore this complexity, this work is focusing on the insurance sector, and providing an overview of general and particular requirements to facilitate the embedding of data. Based on a qualitative literature review and interviews with respondents from the insurance sector, the evaluation will lead to a requirement framework, which sums up certain aspects in relation to Big Data characteristics. Finally, the classification into the business landscape of an insurance company supports the understanding about certain requirements and which departments are affected in specific.

1.1 Motivation

Keeping in mind that the amount of sources of Big Data had grown over the last decades it is necessary to clarify the meaning when the term is used in literature. In general, many definitions can be summarized by the following way of the Gartner institute:

> *"Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."* (Garner Research, 2017).

While the origin of the term "Big Data" is not clarified, other definitions e.g. Dumhilll (2012) or IBMs (2012) declaration, in general point out the same aspects Gartner mentioned. The characteristics of Big Data changed and developed over the recent years from a three component model, so called 3 V- Model (Laney, 2001), over the typical 4 V- Model (Gattiker et al., 2012) into a five component model (Gantz & Reinsel, 2012, p. 9):

1. Volume
2. Velocity
3. Variety
4. Veracity
5. Value

According to this background, the motivation of this thesis is to define the requirements, companies are suggested to think about, while establishing new methods or techniques to facilitate real time decision making. Severely dependent on the topic of Big Data and data management are the phrases Business Intelligence (BI) and data warehouse (DWH), which are concentrating on data savings possibilities and its analytical methods to gain significant and new insights from the huge amount of data (Watson & Wixom, 2007, p. 96). In opposite to conventional solutions which are difficult in their implementation and furthermore their success is depending on different aspects like resource capacities, team skills and development technologies (Wixom & Watson, 2001, p. 33). New upcoming open source software is now entering the stage, promising more than ease of implementation and a new systematic approach to handle the massive data occurrence. The MapReduce algorithm brought a new way into the digitalized world. Order and structure was entering the business world, supporting companies to achieve a successful implementation of customer created and company-acquired data by centralizing control flows and the acceleration of data calculations (Vavilapall et al., 2013, p. 1). Recent researches in academic and industrial circles show that the capability of transforming decision-making processes through more transparency and improved control mechanisms can lead to “5% more productivity and 6% more profit” compared to competitors (McAfee et al., 2012, p. 64). Even more significant was the result of the McKinsey report, which summarizes the potential growth through Big Data usage, especially in the insurance sector, as one of the highest potentials (Manyika et al., 2011, p. 9). Which is not a surprising perception, keeping the many sources of data generation and the impact on daily life in mind and one of the main tasks companies expecting from BI technologies, to make the risk management even more precisely. According to this, many researcher are focusing on data generated by social networks or the behavior of customers. In opposite Alex Pentland remarks that even more data is collected from the location tracking function of smartphones, credit cards and most things people are not aware of doing daily (Pentland, 2012). Referred to this approach, Big Data generation is not only conscious, but describes the human behavior in a more detailed and fine granular way. In context of the insurance sector, the possibility of individual contracts based on consumer behavior, becomes even more attractive.

Additionally, the automatic generation of individual behavior can be useful in another way. While conventional methods of data analysis are bias susceptible, the incomprehensible amount of available information and its evaluation can lead to error elimination through statistical significance in observations (Anderson, 2008, p. 2). Despite the development of such budding solutions, many companies are still struggling with the adjustment of transforming their IT into the new cloud based solutions. Related to afore-

said, this thesis examines the source and reason for the implementation issue, to understand why those obviously new technologies, which can improve the usage of all those available information, are not embedded in the daily work. This aspects will be core of the study and result in an overview about the most relevant requirements, which can be found in academic literature and be confirmed through interviews with representatives from the insurance sector.

The following visualization shows the analytic methods already used in companies today and their potential for future entry. Obviously, many techniques BI solutions are related with, are already in use, but connected to high efforts and limited in its transferability through the data structure.

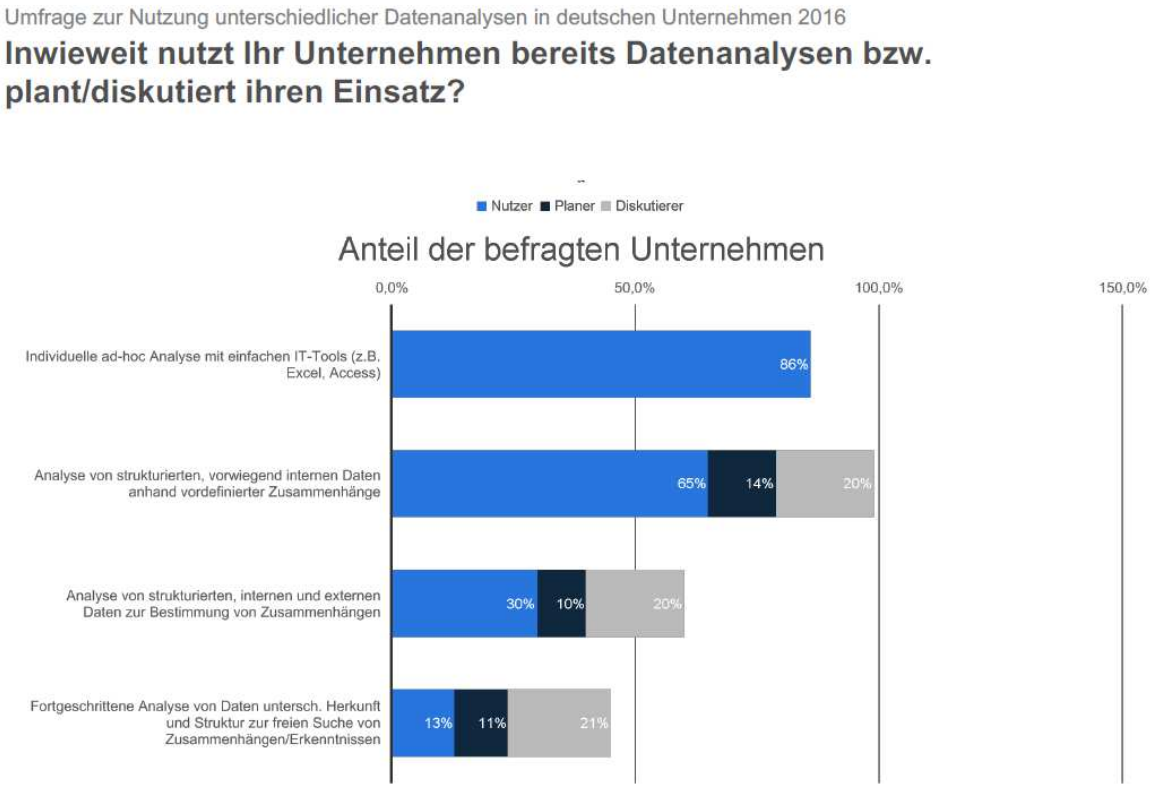

Figure 2: Survey over Technologies in Use (Statista, 2016)

This paper presents common problem sources, like the structure of data based on the source the data is gathered from. Especially in times of smartphone users, social networks, 3D- simulations, finance and stock records, the amount of structured data decreases, while unstructured data reaches a new top (Ammu & Irfanuddin, 2013, p. 613). The practical significance will be presented later in the interview execution. In the academic literature exist a lot of theoretical frameworks and accepted approaches to understand the complexities companies are dealing with regarding the digital era, but unfortunately experiences from employees, who are working on these issues, are rare.

To close this gap of information and to make it more feasible, this thesis presents and classify the major requirements arising by the implementation of Big Data technologies.

1.2 Problem Statement and Thesis Outline

Motivated by the lack of knowledge about the implementation of Big Data technologies in a way that gains sustainable profits to a company, this thesis concentrates on the pre-definition of requirements before starting to implement a foreign technology into business processes (Unit, 2012, p. 4). The struggle with the effective use of unstructured data and its turn into capital returns is one of the most observed topics under investigation in time of digitalization.

Although many companies are arguing that Big Data is one of the most important issues to be involved into the strategic orientation, organizations are still not sure, if and how Big Data should be integrated into the business units (Unit, 2012, pp. 6-10). According to this, a possible research question could focus on the generation of Big Data – where data stems from and if the indicators of Big Data are fulfilled. While this could be an important topic for future analysis, this elaboration will not produce penetrative insights into this area. Assuming that the use of Big Data technologies currently is more available in technic and IT oriented companies, organizations in general limit their activities on brighter analysis, but not in terms of the Big Data definition mentioned in chapter 1.1. To reply to this assumption it will be part of the interviews and meetings with respondents of the insurance sector to get a feeling for the status quo in economic oriented digitalization. One outcome could be whether companies are still growing into this new subject, rather than having the huge variety of data already implemented in daily businesses.

As the matter of Big Data is quite big and nontransparent, the scope of this work needs to be narrowed on a more feasible level. Therefore, the technological aspects in relationship with Big Data solutions will not be part of the chapter about basic concepts. Due to the identification of requirements it is not excluded that single parts are exploring or overcutting with technical aspects, what evolves through the close relation between hard- and software aspects. Nevertheless, this thesis will not provide a hardware construction guideline to build an effective Big Data technology. The focus lay on decision support, through the representation where demand and needs in a company can be located. Moreover, the detected requirements can serve as a foundation for a framework to build an application and specify its functionalities. Data structure will be included when talking about the classification into BI context, but certain recommendations regarding the appropriate solution for a company won´t be taken into account.

To summarize the structure of this thesis the following chapters will be described in shortly.

Chapter two starts with the basic principles in context of Big Data. This means the evaluation of the common definitions and which aspects need to be considered, when an academic analysis about this theme is done. As already mentioned, the characteristics of Big Data will be explained and its relevance outlined, to be served subsequently as classification category of the requirement collection. Equally important is the introduction to the general business model of insurance companies, to ease the transfer of requirements into the subject landmark.

Chapter three is subdivided into two main steps. Firstly, a qualitative literature review will be taken into account. Next to the introduction of the research process, its outcome will provide an overview about demand and requirements. Similarly, the interviews in the second step are concentrating on the confirmation of those factors, which were already identified, as well as the reception of new information. In contrast to the literature review, it is assumed that the interview part will give a more specific view on the matter of BI.

Chapter four is the essential part of this document and mainly based on the pre-work in chapter three. On the one hand, the pure requirement list will be transferred into index cards, containing basic descriptions, dependencies and located areas. On the other hand, those cards will be classified into the exemplary landmark of an insurance company.

The outcome of this analysis shows the impact of the growing volume of data, referred to the located areas in a company and serves for future research about the implementation methodology of new applications into existing organizational architectures.

5 Conclusion

5.1 Summary

In this master thesis I provided an overview about common concepts in BI context as well as the specification of Big Data requirements in the insurance sector. The 5V-model serves as base for an academic literature review to identify requirements and classify them based on characteristics. Besides dependencies and potential conflicts between each other were considered. The findings from literature review were derived in a questionnaire to discuss them with representatives from insurance sector and evaluate the significance of Big Data for practical reasons. Next to this evaluation the impact on an exemplary business landscape of an insurance could be discussed and highlights departments with lot of potential for BI solutions. Finally, the findings from literature review were compared to the conversations with experts and resulted in an explanatory whether academic requirements are even important in the reality.

Overall, Big Data was recognized as a sustainable opportunity to stay competitive in the insurance sector. Next to different projects, which are already set up to harmonize the application landscape or data warehouses in international use, organizations are still thinking about the right way of BI implementation. In general the interests in new source acquisition is very high, but the way to make use out of those potential information is still unclear. During the interviews one major finding was that IT related employees expect the business departments to describe use cases and its transfer would be mainly executed by the IT. The IT units understand themselves as strategic preparer to facilitate business colleagues in analytic tasks. According to that, the differences in the evaluation can be explained through the focus of divisions. For example ranked the business units strategic analysis and scenario building as major aspect concerning Big Data technologies while IT divisions prefer the explanatory approach. Nevertheless, all participants agreed on the potential from sources which are not connected today like specific climatic information or sensors in the industries. Observing the production line from industrial clients enable insurances to make risk calculations accurately as possible. In summary, the major potentials through Big Data were seen in customer marketing, relationships and product design. As micro marketing is an overall trend in different sectors, the insurances can profit from individual contract design in specific. Furthermore the probability to detect possible frauds increases with the use of more data and statistical conspicuousness.

Another outcome of the study results from the comparison of practical and academic findings. Taking a look on the literature distribution which deals with Big Data requirements, it is clear that the identification of certain requirements is still very difficult and hence exist much articles in form of conferences and discussion papers. In reality, the

requirement identification is part of the specification analysis and strongly depends on the purpose of a project. Indeed certain requirements were discussed and mentioned as relevant in context of Big Data, but as long as use cases are rare, the results can be understood as expert estimations. The limited transferability on other Big Data projects could serve as foundation for future investigations in this direction. Some articles described requirements on a very detailed level with specific declarations regarding data volumes, time for certain processes and algorithm to use for efficient structures. However even in those cases the transfer was limited for this work, due to the aspect that no concrete project was attended.

All in all, this work identified expectations about Big Data technologies and the reservations regarding the implementation of those unrehearsed solutions. The approach to combine academic requirements with practical impressions in context of BI is still infrequently. Therefore this work expands the field of use cases and gives an overview about the significance of certain characteristics for Big Data methodologies. Furthermore, index cards were produced and leverages readers to get in common with different requirements. This work confirms that investigations in the field of Big Data are still in demand and successfully connects theoretical requirements and practical experiences.

5.2 Limitations and Future Research

Although a broad overview about requirements and its significance could be presented, some limitations exist. As already mentioned this work predefines requirements and then proves its relevancy for reality. Therefore, it can be questioned if the specifications can be transferred to other projects or use cases or if certain points are missing. A possible approach for future work could go along with a BI implementation and verify which aspects should be content of a requirement analysis. Another limitation lies in the scope of this work. Obviously, the technological aspects were rather neglected and not treated in detailed way. The research emphasizes on general concepts to understand basics of Big Data like ETL processes and data warehouse structure.

Additionally, more qualitative interviews could lead to a double-check on the requirements and enable to make adjustments in the index cards. Having more time during the analytic phase would have given more insights through the practical experiences and thereby increasing the overall validity.

Possible extensions of this work can focus on the connection between requirements and the implementation into a data architecture. To understand which requirements exist and how the implementation into concrete steps work, would bring deep insights into the narrow field of Big Data. One possibility is to search for existing architecture

frameworks and select that one which can be used as reference model. Benchmarking different types also could lead to the identification of even more requirements to be part of experiments with comparable hardware components.

A last point to be mentioned deals with considerations about current software solutions. While there already exist many different open-source technologies to perform Big Data analysis, this can be used as starting point to test which solution counters the majority of requirements. Moreover, this could generate a confrontation in form of lists with requirements, categories or other classifications.