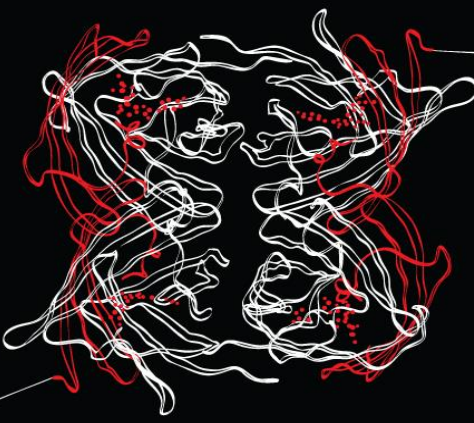




MASTER THESIS BY MIKE PADBERG

# Big Data and Business Intelligence: a data-driven strategy for e-commerce organizations in the hotel industry



UNIVERSITY OF TWENTE.



HotelSpecials



MASTER THESIS

# **Big Data and Business Intelligence: a data-driven strategy for e-commerce organizations in the hotel industry**

**Date** 03-09-2015

**Personal information**

**Author** Mike Padberg

**E-mail** m.d.padberg@student.utwente.nl

**Study program** Master in Business Administration  
Business Information Management  
University of Twente  
The Netherlands

**Graduation committee**

**First supervisor** DR. IR. A.A.M. (Ton) Spil

**Second supervisor** IR. B. (Bjorn) Kijl

**External supervisor** MSc M. (Merijn) de Ruijter van Steveninck

## **Acknowledgements**

This research is the final result of my Master Thesis project to obtain the Master of Science degree in Business Administration with a specialization in Business Information Management at the University of Twente. Information Technology (IT) related topics have always attracted my attention. This is why I chose the Business Information Management track, which gave me the opportunity to write my Master Thesis about an Information Technology related topic.

I would like to thank several people who supported me during this Master Thesis project. Firstly, I would like to thank my supervisors from the University of Twente, first supervisor Ton Spil and second supervisor Bjorn Kijl, for all the time and constructive criticism during this process. Secondly, I would like to thank all my colleagues from HotelSpecials and especially Merijn de Ruijter van Steveninck who offered me the opportunity to finish my graduation assignment at HotelSpecials and for all his support during my master thesis. Last but definitely not least, I would like to thank my parents and brother for their continuous support during my study.

Mike Padberg

Enschede, September 2015

## Summary

This thesis focused on creating a practical approach to become a more data driven organization. There are multiple ways an organization can become more data driven. One way an organization can become more data driven is by using Big Data technologies and by optimizing the Business Intelligence process.

Therefore, in this thesis the following research question will be answered: *How can an organization start with Big Data to get more value out of the available data and optimize the Business Intelligence processes in such a way that it will be more frequently used for decisions?*

To answer this research question, an experiment and multiple interviews were conducted. The interviews with Big Data and/or Business Intelligence experts were aimed at getting a better understanding of Big Data and Business Intelligence. The experiment took place at HotelSpecials and was conducted to create extra validation and to get a better understanding of the transition an organization has to make to become a more data driven organization.

An organization could start with Big Data, Business Intelligence, and Decision Making by: (i) selecting a test department with an open-minded and data friendly manager (ii) identifying and selecting opportunities that can be solved with Big Data, Business Intelligence, and Decision Making (iii) starting an innovation process with the following steps: experimentation, measurement, sharing, and replication (iv) train employees about the capabilities of Big Data (v) start with Big Data and learn about Big Data tools while implementing and using them (vi) make a list of all Big Data tools that meet the MAD requirements (vii) choose the right tool that fits the purpose of the organization (viii) and do not focus on developing smarter systems or smarter algorithms than the competitors.

The results of this thesis also indicate that: (i) In general, Big Data is still considered as a new subject and research area. For example, in the interviews nearly all respondents mentioned that they see Big Data as “buzzword” (ii) Within the Big Data community two streams can be distinguished – the first stream are people without an background in computer science or software engineering and argues that Big Data is related to Business Intelligence and Decision Making, while the second group with an background in computer science or software

engineering argue that Big Data is an enabler of artificial intelligence, self-learning software, and smarter algorithms (iii) If an organization starts with Big Data, Business Intelligence, and Decision Making the organization should not replace their current platforms (v) It is important to create an evidence based culture (vi) Making timely decisions and understanding your customers can create a huge performance increase (vii) It is important to maximize the visualization of the available data.

## Contents

Acknowledgements.....	i
Summary.....	ii
1. Introduction.....	1
1.1 Problem statement.....	2
1.2 Research goal and research question.....	2
2. Theoretical framework.....	4
2.1 Used method for literature reviewing.....	4
2.2 Big Data.....	6
2.3 Business intelligence.....	10
2.4 Decision making.....	12
2.5 Big Data, Business Intelligence, and Decision Making.....	14
3. Methodology.....	16
3.1 Research types.....	16
3.2 Units of analysis.....	16
3.3 Data collection tools and data analysis.....	17
3.4 Reliability and Validity.....	19
4. Interview results.....	21
4.1 Definition and basics of.....	21
4.1.1 Big Data.....	21
4.1.2 Business Intelligence and Decision Making.....	22
4.1.3 Big Data, Business Intelligence, and Decision Making.....	23
4.2 Examples and the use of.....	24
4.2.1 Big Data.....	24
4.2.2 Business Intelligence and Decision Making.....	26

---

5. Experiment results .....	28
5.1 Survey .....	28
5.2 Data collection .....	29
5.3 Selection process .....	30
5.4 Financial results.....	31
6. Explanatory analysis.....	33
7. Conclusion and recommendations .....	36
7.1 Conclusion.....	36
7.2 Contribution to literature .....	38
7.3 Limitations and future research .....	38
References.....	40
Appendix A: search results scopus .....	44
Appendix B: interview framework.....	48
Appendix C: survey design .....	50

---

## 1. Introduction

Modern organizations do not only want to know what happened and why it happened, but also want to know what is happening right now and what is likely to happen next (LaValle et al., 2011). Since organizations hunger for these insights and the adoption of the World Wide Web, the generation of data and collection speed has increased exponentially (Chen et al., 2012). The roughly biannual doubling of computing power and storage for the same price, also known as Moore's law, has also done remarkable things - for example in 1994 people paid \$1000 for a Gigabyte of storage, while in 2010 the costs of a Gigabyte of storage was only \$0,10.

The demand for all this information and all these rapid technological developments enabled organizations to capture, store, and analyze large amounts of data. Take for example Flickr, a public photo sharing website, which received in 2014 an average of 1,83 million photos each day (Flickr, 2014). Assuming that the storage size of each photo varies around 1,5 and 3 Megabytes it will result in a terrifying storage size of 3,9 Terabytes each day.

With the help of these rapid developments more organizations are shifting their focus to exploring and exploiting all this data. This phenomenon is called "Big Data" and is identified on the emerging technology hype cycle as one of the biggest IT trends of the last few years (Gartner, 2014). Since Big Data is still a trend, people use Big Data as catch-phrase to describe the massive amount of information that is too difficult to process by a traditional database or traditional software techniques.

In general, organizations see Big Data as an asset. Some organizations make the comparison with oil, because like oil this Big Data needs to be refined before it gets a value. However, here is where most organizations struggle. For example, when going to a Big Data conference some organizations argue that: "Big Data is like teenage sex; everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it" (Ariely, 2013). Therefore, the goal of many organizations today is to create a more practical approach to start with Big Data and try to go beyond the buzzword and catch-phrase.



## 1.1 Problem statement

An organization that recognizes the problems with and the added value of Big Data is HotelSpecials. HotelSpecials is an organization that provides various hotel deals through different websites in the Netherlands, Belgium, Germany, Sweden, Denmark, and Norway. Their organizational ambition is to; capture more, store more, and analyze more data and increase the organizational performance by making more decisions based on data. To realize this ambition HotelSpecials already created a department for Big Data issues and specific Big Data analytics. Furthermore, the organization is at a first stage to change their enterprise architecture and ready to deploy new Business Intelligence and Big Data software.

Right now, HotelSpecials is experiencing a few drawbacks to become more data driven. Firstly, many employees still see Big Data as a buzzword and catch-phrase. Furthermore, the culture within the organization is that they still rely on intuition, experience and gut feelings for day-to-day decisions. Therefore, the organization is struggling to become more data driven and to increase the organizational performance by make more decisions based on data. To conclude, HotelSpecials is searching for a good practical approach to start with Big Data and to optimizing the current Business Intelligence processes.

## 1.2 Research goal and research question

The main goal of this thesis is to create an approach to become a more data driven organization. Therefore, this thesis will examine how an organization could start with Big Data and how an organization could optimize the current Business Intelligence processes.

The research question that fits this goal is:

*How can an organization start with Big Data to get more value out of the available data and optimize the Business Intelligence processes in such a way that it will be more frequently used for decisions?*

The sub-research questions that help to answer this research question are:

- What is Big Data?
- What is Business Intelligence?
- What is a choice or a decision?
- What is the current status of Big Data and Business Intelligence at HotelSpecials?

- What do experts believe is the best way to start using Big Data and to optimize a Business Intelligence process?
- What are the steps and possible improvements for that specific organization to start with Big Data and to optimize their Business Intelligence process?

## 2. Theoretical framework

*The purpose of this theoretical framework is to provide an academic foundation for this thesis. In section 2.1 an overview will be given regarding the used method for literature reviewing. Section 2.2 describes the history, the definition, and the theoretical examples of Big Data. In section 2.3 an overview will be given of the history, the definition, and the theoretical examples of Business Intelligence. Section 2.4 will elaborate on the history, the definition, and the theoretical examples of Decision Making. Lastly, section 2.5 will provide a conceptual model and overview of these three strongly related research areas.*

### 2.1 Used method for literature reviewing

A vital step to create a proper foundation for any researcher is the task of completely reviewing a chunk of academic literature (Wolfswinkel et al., 2013). The paper of Wolfswinkel et al. (2013) offers a guidance, in order to create a systematic literature review, by using a grounded theory approach.

The first step is marking out the scope of the review and to define the criteria for inclusion or exclusion of resources. These criteria are:

- Firstly, only the top 15 cited articles or papers with a particular search term in their title are selected. This way the search engine will provide a list of really specific articles and papers, which also enables the use of forward and backward citations.
- Secondly, articles or papers acquired through forward or backward citation analysis must exceed 50 citations.

The second step is to identify and select the appropriate 'fields' of research. The subject areas for this thesis are: Business management, Computer science, Decisions science, Economics, Information systems, Psychology, Social sciences.

The third step is to determine the appropriate sources. In this step the researcher will select the appropriate databases. The databases that will be used for this thesis is: Scopus.

The fourth step is to precise formulate the possible search terms. The search terms in this thesis are not combined as one search term, because "Big Data" is a relatively new research topic in comparison with "Business Intelligence" and "Decision Making", while the goal for these search terms is to receive the top 15 cited articles. The search term, sorted on citations, are:

- TITLE(Big Data) AND ( LIMIT-TO(SUBJAREA,"COMP" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"DECI" ) OR LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"PSYC" ))
- TITLE(Business Intelligence) AND ( LIMIT-TO(SUBJAREA,"COMP" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"DECI" ) OR LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"PSYC" ))
- TITLE(Decision OR Choice) AND ( LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"DECI" ) OR LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"PSYC" ))

The final step is to make an analysis and present the papers found with the search terms, see appendix A. The author and scope of the articles found with the selected search criteria and used for this thesis are listed in Table 1.

**Table 1: Authors and the scope of the articles and papers used for this thesis**

<b>Author(s)</b>	<b>Scope of the articles and papers</b>
Chen et al. (2012)	An overview of the evolution, applications, and emerging research areas of Business Intelligence & Analytics.
Boyd & Crawford (2012)	Six critically questions/challenges about Big Data its assumptions and biases
Cohen et al. (2009)	A new analysis practices for Big Data
Jacobs (2009)	The issues that can arise when analyzing Big Data
McAfee & Brynjolfsson (2012)	The impact Big data has on a company and its decision-making culture
Herodotou et al. (2011)	A selftuning system for Big Data analytics
Madden (2012)	The impact of Big Data on various databases
Cuzzocrea et al. (2011)	An overview of the literature of Big Data
LaValle et al. (2011)	Big Data, analytics and the path from insights to value

Labrinidis & Jagadish (2012)	Challenges and opportunities with Big Data
Chen et al. (2012), Watson & Wixom (2007), Chaudhuri et al. (2011), Duan & Xu (2012)	An overview of the current state, evolution, applications, and emerging research areas of Business Intelligence & Analytics.
Cody et al. (2002)	The integration of Business Intelligence and knowledge management technologies
Elbashir et al. (2008), Lönqvist & Pirttimäki (2006)	Measuring the effects of business intelligence and business intelligence systems
Chung et al. (2005)	The visualization capabilities of Business Intelligence tools
Rivest et al. (2005)	The SOLAP technology
Lee & Park (2005)	Identifying profitable customers with Business Intelligence tools
Bellman & Zadeh (1970)	Decision-making in a fuzzy environment
Kahneman & Tversky (1984)	The cognitive and the psychophysical determinants of choice in risky and riskless contexts
Saaty (1990)	The analytic hierarchy process
Kahneman (2003)	A perspective on judgment and choice
Simon (1956)	Rational choice and the structure of the environment
Samuelson & Zeckhauser (1988)	Status quo bias in decision making
Tversky (1972)	A theory of choice
Edwards (1954)	Theory of decision making
Stanovich & West (2000)	Two distinct kinds of reasoning systems
Evans (1984), Evans (2006)	The heuristic-analytic theory of reasoning
James (1890)	The principles of psychology

## 2.2 Big Data

The understanding of customers increased dramatically once shopping moved online (McAfee & Brynjolfsson, 2012). Since the early 2000s, the World Wide Web began to offer unique data collection methods (Chen et al., 2012). For example, web-shops can not only track what customers bought, but also how the customers navigated through their web-shop, what else the

customer looked at, how much the customers were influenced by the page layout, and if the customer clicked on a promotion link (McAfee & Brynjolfsson, 2012). Web-shops are also able to conduct A/B testing – a test that takes the statistical difference on various metrics of the behavior of a customer with version A and version B (Cohen et al., 2009). Organizations that are born digital and that have the ability to create value from such data can achieve a competitive advantage on traditional organizations. Like McAfee & Brynjolfsson (2012) argue: “Traditional retailers simply couldn’t access this kind of information, let alone act on it in a timely manner”.

Organizations of today do not only want to know what happened and why it happened, but also want to know what is happening right now and what is likely to happen next (LaValle et al., 2011). Since organizations hunger for these insights and the adoption of the World Wide Web, the generation of data and collection speed have increased exponentially (Chen et al., 2012). For example, a born digital organization could have an immense databases from just a single source like clickstreams (Cohen et al., 2009). Organizations and employees see these “Big Data” databases as a real opportunity and use it for targeted advertising, for optimizing their offers, or even compare the rest of the market (Boyd & Crawford, 2012). In fact, the ability to make timely analytics on all this “Big Data” is a key ingredient for many successful organizations (Herodotou et al., 2011).

But what is “Big Data”? Within the literature there are a lot of definitions and opinions about Big Data. According to Cuzzocrea et al. (2011) Big Data refers to enormous amounts of unstructured data produced by high-performance applications. Boyd & Crawford (2012) define Big Data as a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and methodology. Madden (2012) argues that Big Data indicates that the data is too big, too fast, or too hard for existing tools to process. However, in this thesis Big Data refers to “the 3Vs” – Volume for the huge amount of data, Variety for the speed of data creation, and Velocity for the growing unstructured data (McAfee & Brynjolfsson, 2012).

Big Data has many opportunities and is now being recognized broadly. To illustrate, McAfee & Brynjolfsson (2012) provide a “getting started” guide for implementing Big Data in an organization:

1. The first job is to create a team and to pick a test department with an open-minded and data friendly manager. The team should contain one data scientist and should not contain more than five employees.
2. Next, begin a brainstorm session or meeting and identify and select no more than five opportunities that can be solved with Big Data within five weeks, with the team the organization selected at the previous step.
3. Start an innovation process with the following steps: experimentation, measurement, sharing, and replication.
4. Finally, an organization could, if possible, send out some analytic challenges on their Big Data to third parties.

Big Data has many advantages. For example, the recognition of Big Data has led to a growing enthusiasm for data driven decision making, also known as evidence based decision making (Labrinidis & Jagadish, 2012). In fact, the more an organization characterize themselves as a data driven organization, the better the organization performs on objective financial and operational key figures (McAfee & Brynjolfsson, 2012). In addition, an organization that identifies Big Data and analytics as an differentiation strategy are twice as likely to be a top performer in their market segment (LaValle et al., 2011). Big Data also has the potential to revolutionize management, because an important aspect of Big Data is the impact it has on how decisions are made and who makes these decisions (McAfee & Brynjolfsson, 2012). Thus, implementing Big Data and data driven decisions could already lead to a performance boost for that particular organization.

However, more data is not always better data. For instance, Boyd & Crawford (2012) argues that numbers do not speak for themselves, because there are many theories and disciplines that tell organizations and employees why customers show certain behavior or write certain things. In addition, a Big Data analysis is of limited value if the decision maker, within the organization, is unable to understand the Big Data analysis (Labrinidis & Jagadish, 2012). In fact, LaValle et al. (2011) argue that getting the data and getting the data analysis right is not the biggest obstacle in Big Data and analytics, but the adoption barriers are mostly related to managerial and cultural change. Furthermore, Big Data can also lead to finding false correlations, also called spurious correlations. An spurious correlation could be if an employee analyzes the mobile phone data of a customer and checks the time the customer spends with

another person to see who are important people in the customers life, but if a customer spends more time with his colleagues it does not necessary mean that his or her colleagues are more important than his or her family (Boyd & Crawford, 2012).

To help business analysts and data scientists finding good correlations and relationships an organization could buy or develop a Big Data technology. In fact, technology is often mentioned as one of the main areas to solve the Big Data problem (Madden, 2012). Organizations mostly rely on Business Intelligence tools – a tool that enables to turn raw data into valuable insights – that query on an Database Management Systems – the whole collection of software packages that enables someone to store and extract data in a database (Cohen et al., 2009). In fact, the solutions most standard organizations rely on can already deal with sizes up to multiple petabytes, big enough for billions of log records, clickstreams, or transaction data (Madden, 2012). But trouble comes to these standard solutions when an business analysts or data scientists want to take all these log records and analyze these within seconds or minutes (Jacobs, 2009). In addition, most open source systems – software with a publically available source code – like MySQL and Postgres are behind in terms of scalability compared to the commercial competition (Madden, 2012).

Therefore, Business analysts, data scientists and developers expect big data analytics systems to be “MAD” - Magnetic, Agile, and Deep (Cohen et al., 2009). A Magnetic system grabs all data regardless of the structure and the quality of the data, an Agile system is adaptable and data flexible, while a Deep system supports traditional Business Intelligence as well as machine learning and complex statistical analyses (Cohen et al., 2009). In the literature there are multiple examples of newly developed MAD systems for Big Data and analytics. For instance, a MAD system for Big Data that is becoming very popular is Hadoop (Herodotou et al., 2011). In a sense, one can argue that Hadoop is the next-generation of Database Management Systems (Cuzzocrea et al., 2011). According to Herodotou et al. (2011) the factors that contribute to Hadoop’s MADness are:

1. Hadoop is considered Magnetic, because the only step to get data in Hadoop is to copy the files into the distributed file system of Hadoop.
2. Hadoop is considered Agile, because it makes use of a so called “MapReduce” methodology. “Map” separates computational tasks into small and parallel tasks and assigns an appropriate <Key, Value> structure to the Big Data, while “Reduce”



obtains all the Big Data by combining all Values that share the same Key (Cuzzocrea et al., 2011).

3. Hadoop is considered Deep, because with Hadoop and third party Hadoop extensions a user can make computations in general programming languages like Java, Python, R, and SQL.

Thus, Hadoop can help organizations, business analysts, data scientists, and developers to find good correlations and relationships, become more MAD, and get more insights from their available Big Data sets.

### **2.3 Business intelligence**

“A critical component for the success of a modern organization is its ability to take advantage of all available information” - Cody et al. (2002). In fact, the ability to gather and timely transform all information in effective business information is not only essential to succeed, but also necessary to survive (Lönqvist & Pirttimäki, 2006). For example, a casino could gather information of a special event or the usage of a slot machine to track the preferences of a customer or the probability of various games and close unpopular, unprofitable, or unknown games quickly (Watson & Wixom, 2007). However, the challenge to transform all this information to effective business information becomes more difficult as the information keeps growing exponentially and the increasing amount of employees who need access to this information (Cody et al., 2002).

Organizations deploy, to support these data savvy employees, data warehouses and frontend applications that can access, analyze, summarize and visualize all available information (Rivest et al., 2005). For example, organizations create frontend applications with a visual dashboard that allow a decision maker to track key performance indicators of their operations (Chaudhuri et al., 2011). These frontend applications which organizations are creating and deploying are also known on the market as “Business Intelligence” applications (Rivest et al., 2005). Many organizations use these Business Intelligence applications to create a knowledge centric approach (Cody et al., 2002). In fact, Business Intelligence not only has the ability to improve the organizational knowledge, but also to decrease Information Technology costs by deleting duplicated data and eliminating unnecessary data (Watson & Wixom, 2007).

But what is “Business Intelligence”? Within the literature there are a lot of definitions and opinions about Business Intelligence. According to Duan & Xu (2012) Business Intelligence is the process of converting raw data into information that provide an organizations with new insights and benefits decisions making. Watson & Wixom (2007) define Business Intelligence as a process with two primary activities – the first activity is to get the data into a data warehouse and the second activity is to get the data out of the data warehouse and use it run a query, to perform an analysis, or use it for reporting. Chaudhuri (2011) argues that Business Intelligence is a collection of different technologies that enable an employee to make better and faster decisions. However, in this thesis Business Intelligence refers to the applications, methodologies, practices, systems, techniques, and technologies that analyze data to help an organization understand their operations and market and make timely decisions (Chen et al., 2012).

The landscape of Business Intelligence applications is growing and organizations are quickly adopting these applications (Chaudhuri et al., 2011). However, an essential question is what advantages are achieved by organizations that use Business Intelligence applications (Elbashir et al., 2008). For example, Business Intelligence applications enable organizations to identify profitable customers and build long term relationships with these profitable customers (Lee & Park, 2005). Furthermore, Business Intelligence applications could be used to systematically analyze the organizational external environment (Chung et al., 2005). For instance, a Business Intelligence application that runs on a weekly basis and helps to extract valuable market information of all competitors and identify new business opportunities (Chen et al., 2012). Business Intelligence applications could also be used for real-time data – a call center could use a few screens to display the performance or an airline can identify passengers who are at risk of missing their connecting flight (Watson & Wixom, 2007).

However, some organizations cannot directly see the opportunities of Business Intelligence applications, because the advantages of Business Intelligence applications are mostly nonfinancial and intangible (Lönqvist & Pirttimäki, 2006). Most Business Intelligence applications that claim to do analysis only provide a few different views of information (Chung et al., 2005). In addition, fifty percent of the costs and eighty percent of the time of a Business Intelligence application is due to poor data quality, legacy systems, and problems with data ownership (Watson & Wixom, 2007). For example, Business Objects – a Business Intelligence

application – requires a specific Information Technology infrastructure in order to function properly (Elbashir et al., 2008). Lastly, because of the limited visual capabilities and the new opportunities fueled by the web, organizations require new and smarter Business Intelligence applications (Chen et al., 2012).

According to Watson & Wixom (2007) organizations are more likely to have success with Business Intelligence when the following conditions exist:

1. Management of an organization should have a vision for Business Intelligence and believe in information-based decision making.
2. The use of Business Intelligence and analytics should be part of the organizational culture and counter decision making based on intuition or “gut feelings”.
3. Alignment between business strategies, business model, and Business Intelligence strategies enables an organization to create organizational change and new business opportunities.
4. An organization should have a strong and effective Business Intelligence governance and infrastructure, because it will address business alignment, funding, project prioritization, and data quality.
5. Lastly, an organization needs to provide users with appropriate Business Intelligence tools for their needs and give effective training and support to these users.

Thus, Business Intelligence applications can enable organizations to identify profitable customers, help an organization to analyze their external environment, and counter decision making based on intuition or “gut feelings” (Chung et al., 2005; Lee & Park, 2005; Watson & Wixom, 2007).

## **2.4 Decision making**

People make decisions, knowingly or unknowingly, all the time (Kahneman & Tversky, 1984). For example, should one on a rainy day take an umbrella and go by bike or should one go by car. A few people make these decisions with the same outcome day after day (Samuelson & Zeckhauser, 1988). However, most people experience uncertainty when they are faced with a choice among several alternatives (Tversky, 1972). In fact, many decision take place in an environment where the goals, constraints, and consequences are mostly unknown (Bellman & Zadeh, 1970). Furthermore, decision making is so complex that even if a person makes a

particular decisions, it does not mean that he or she will always makes that same decision under identical conditions (Tversky, 1972). For instance, a child standing in front of a candy store may decide one day to buy a candy bar and may decide the next day, under the same conditions, to save his money (Edwards, 1954).

Therefore, it is hardly surprising that many researchers, from economics and statistics to psychology and sociology, try to account for the behaviors and decisions of individuals (Kahneman & Tversky, 1984). The question of why individuals show certain behavior, why individuals have particular ideas, and why individuals make certain decisions has a long history in psychology theory (Kahneman, 2003). For example, researchers conduct tests and design experiments with two states, state A and state B, and where an individual needs to choose state A over state B or vice versa (Edwards, 1954). To illustrate such a decision experiment, consider the choice between receiving €800 for sure and a gamble with an 85% chance to win €1000 (Kahneman & Tversky, 1984). Interestingly most people decide to choose the sure thing, while the gamble is mathematically more interesting –  $0.85 \times €1000 + 0.15 \times €0 = €850$  (Kahneman & Tversky, 1984).

But what is a “decision”? According to Bellman & Zadeh (1970) a decision is a choice between certain alternatives. A decision can also be defined as a knowingly or unknowingly choice of an individual (Kahneman & Tversky, 1984). Saaty (1990) show that it is the identification of the available options and prioritized these options with the help of specific criteria. Kahneman (2003) argue that each choice problem can be considered as a separate decision. Simon (1956) defines a decision as a rational choice from the available options. Samuelson & Zeckhauser (1988) argue that a decision is an individual that selects one of a known set of alternative choices. Likewise, Tversky (1972) define it as a choice among several alternatives. However, in this thesis a decision refers to the choice between two or more options and where an individual prefers one options above the other available options (Edwards, 1954).

Now one knows what a decision is, one might argue that the most difficult task of a decision and the whole decision process is to choose the factors that influence the choice between the two or more options (Saaty, 1990). However, in psychology there are a few theories who deal with human behavior, human thinking, and people’s judgment that influence these factors. For example, many years ago psychologists already tried to assemble the theory of dual-process, a

theory who deals with cognitive processes such as human behavior, human thinking, and peoples judgment (Evans, 1984). While not all psychologists have the exact same vision about this theory, they considerable agree on a general view – two distinct systems, with two sets of characteristics, within the human brain that compete with each other and try to control human behavior and the judgment of a person (Evans, 2006). To sum up, it is important to understand this dual-process if one wants to understand why and how people make certain decisions.

The theory of two distinct kinds of reasoning systems within the human brain has a long history. Already in 1890 William James provided a foundation for dual-process thinking. James (1890) argued that one system within the brain is used for repeated tasks that are based on past experiences, also called associative thinking, while the other system within the brain is used for uncommon situations that need additional reasoning, also called true reasoning. After James (1890) a lot of different researchers used his theory and vision to develop new or additional dual-process theories. For example, Stanovich & West (2000) also argued that the human brain has two distinct kinds of reasoning systems, but they used the labels system 1 and system 2 and provided an extensive list of characteristics for these systems. System 1 can be described as an automatic, associative, heuristic, holistic and relatively fast system, while system 2 can be described as an analytical, controlled, rational, rule-based, and relatively slow system (Stanovich & West, 2000). Kahneman (2003) used the theory of Stanovich & West (2000) and considered system 1, the relatively fast and emotional part, as the intuition of a person and system 2, the relatively slow and non-emotional part, as the reasoning part of a person.

## **2.5 Big Data, Business Intelligence, and Decision Making**

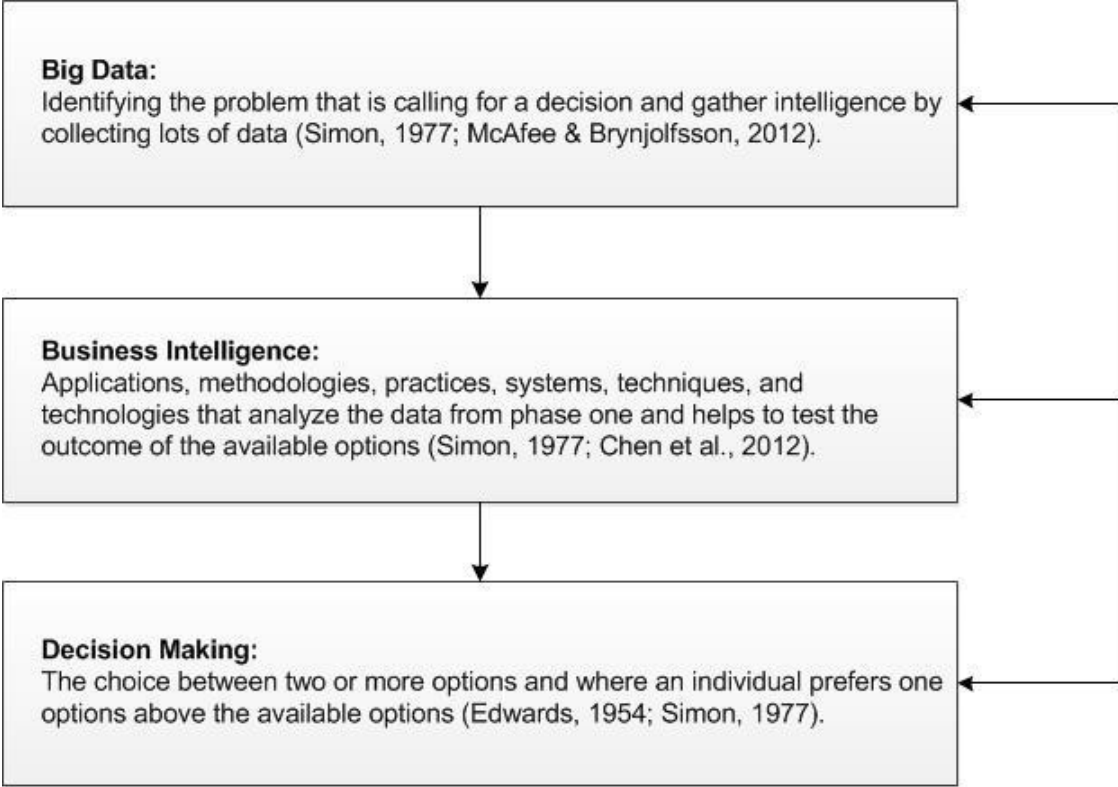
In the literature, Big Data, Business Intelligence, and Decision Making are considered as three strongly related research areas. For example, in 1977 Simon already introduced his normative model of decision making that provides a clear overview of the link between Big Data, Business Intelligence, and Decision Making. Simon's (1977) famous model of decision making contains three phases:

1. Intelligence gathering: the identification of the problem calling for a decision and the data collection of the problem.
2. Design: inventing, developing, and analyzing the available data to test the outcome of the available options.
3. Choice: select a particular option based on the selection criteria.

Based on this model of Simon (1977) and the literature of chapter 2.2, 2.3, and 2.4 a conceptual model was created. Figure 1 shows this conceptual model and the link between Big Data, Business Intelligence and Decision making. The different phases in this conceptual model are:

- 1. Big Data: identifying the problem that is calling for a decision and gather intelligence by collecting lots of data (Simon, 1977; McAfee & Brynjolfsson, 2012).
- 2. Business Intelligence: applications, methodologies, practices, systems, techniques, and technologies that analyze the data from phase one and helps to test the outcome of the available options (Simon, 1977; Chen et al., 2012).
- 3. Decision Making: the choice between two or more options and where an individual prefers one options above the available options (Edwards, 1954; Simon, 1977).

**Figure 1: conceptual model**



### **3. Methodology**

*The purpose of this methodology chapter is to describe the chosen methodology and it includes information about the steps and data gathering tools used for this thesis. In section 3.1 an short overview will be given of the different types of research. Section 3.2 will provide an short elaboration on the units of analysis. In section 3.3 the used data collection tools and data analysis method will be described. Lastly, section 3.4 will elaborate on the reliability and validity of this thesis.*

#### **3.1 Research types**

Research is the act of finding something out (Babbie, 2007). This “something” can be anything, therefore the researcher needs a plan. A plan, also known as a research design, could have many purposes. According to Babbie (2007) three of the most common purposes are: exploration to examine a for the researcher relatively new topic, description to describe a particular event or observation, and explanation to explain why a certain event or observation happens. Next, the researcher has to select the units of analysis. Furthermore, there are many tools that could help a research to gather information. For example, an experiment and an interview can help a research to answer how and why questions, while a survey can answer who, what, where, how many, and how much questions (Yin, 2009). All these tools can only produce qualitative data if all data in a category is non-numerical and will produce quantitative data if a category contains numerical data only (Babbie, 2007). Finally, it is important to describe all these steps as extensive as possible to make the results valid and reproducible.

#### **3.2 Units of analysis**

The term Big Data is relatively new and the meaning is also subjective and sometimes unclear. Therefore, this thesis has to distinguish and select multiple units of analysis. The first unit of analysis are organizations who are transforming to using Big Data to get more value out of the available data and to using Business Intelligence in such a way that it will be more frequently used for new insights and decisions. The second unit of analysis are people or organizations who already are experts in Big Data or Business Intelligence. For example, an university professor that is studying Big Data or an organization that already developed some sort of best practice technique like Booking.com.

This thesis will focus on the e-commerce market of hotel accommodations, because this information was available for me as a researcher. The first unit of analysis is HotelSpecials. HotelSpecials is an organization that already has access to lots of data and the organization wants to capture more, store more, and analyze more data to gather insights into the customer's behavior to gain a competitive advantage. However, HotelSpecials cannot be labeled as an expert in Big Data or Business Intelligence and is therefore the first unit of analysis. The second unit of analysis are the "experts". For example, a professor of the University of Twente who is studying Big Data or organizations like Booking that are leading the e-commerce market of hotel accommodations with the help of Big Data and Business Intelligence.

### **3.3 Data collection tools and data analysis**

#### **Interview**

The main research question of this thesis starts with a how question and, like mentioned before, in depth interviews are perfectly suited for how and why questions (Yin, 2009). Therefore, interviews with different unit of analysis are selected as the data collection tool of choice. In total 9 interviews will be conducted. Since the goal of this thesis is to provide an organization with a set of recommendations on how to extract more value from the available data with the help of Big Data and to make more decisions based on data with the help of Business Intelligence, the interviews were aimed at getting a better understanding of Business Intelligence and Big Data and going beyond the buzzword. For this research it is important to know how different employees within an organization define Big Data and Business Intelligence, why they want or do not want to use Big Data and Business Intelligence, and how Big Data and Business Intelligence can deliver additional value for different business units.

It is also important to define the appropriate selecting tool for gathering data, because the researcher needs to choose between multiple interview structures. For example, a researcher could choose between open ended or closed ended questions, a structured, unstructured or a semi-structured interview, and so on (Babbie, 2007). In this thesis only opened question were used, because the goal of this thesis is aimed at better understanding the Big Data and Business Intelligence industry. In addition, to only asking opened question this thesis used a semi-structured format. A semi-structured interview is a format where one starts with a more general topic or question. This way it provides sufficiently structure to study a specific topic and related phenomenon, while it allows the interviewer to focus more on the conversation, ask



questions that arise during the interview, and leaves space for the participants to offer new meanings to the study (Galletta & Cross, 2013). To sum up and illustrate foregoing, appendix B was created. Appendix B provides the interview framework that is based on the information above.

## **Experiment**

According to Yin (2009) an experiment is also perfectly suited for how and why questions. To gather even more data for this thesis, in addition to the interviews, an experiment was conducted.

The experiment was best suited for the first unit of analysis, which is an organization who is transforming to start with Big Data to get more value out of the available data and to optimizing Business Intelligence process in such a way that it will be more frequently used for new insights and decisions. The hypothesis that started this experiment was “during the day customers were searching on their mobile and would finish the order on their desktop in the evening”. Therefore, this experiment was conducted amongst the employees of HotelSpecials and aimed at creating awareness and explore how the total organization could extract more value from the available data with the help of Big Data and how this could stimulate employees from HotelSpecials to make more decisions based on data with the help of Business Intelligence. In fact, this experiment was presented as a game to make it more “fun” and to maximize the awareness and insights.

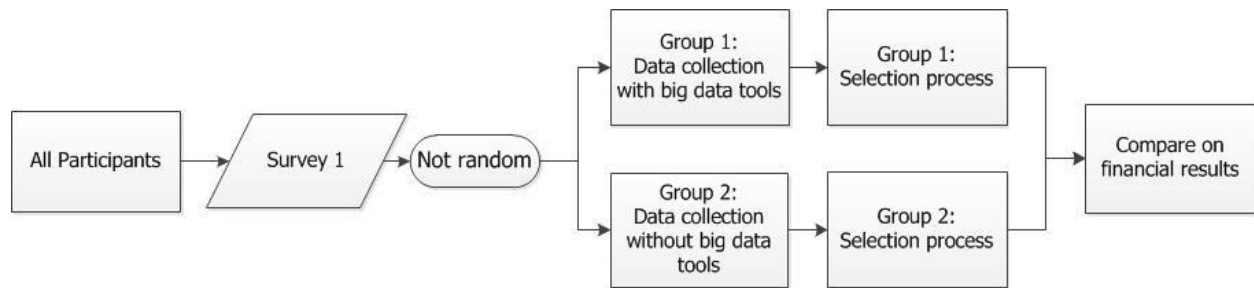
The experiment took place on a normal working day and the participants could sign up via an internal memo a week before the experiment took place. Furthermore, in this internal memo the researcher also clearly stated that participation was on a voluntary basis, all results would be anonymized, and that there was no intention to harm the participants (Babbie, 2007).

According to Babbie (2007) a research can choose between multiple designs for an experiment, although the most preferred experiment is a classic experiment. This type of experiment examines the effect of an independent variable on a dependent variable, where the independent variable is often a stimulus that is present in an experimental group and absent in a control group (Babbie, 2007). In the experiment the independent variable is the use of Big Data Analytic tools and the dependent variable is does this lead to an performance increase, or in terms of the

research question, does this lead to more value out of the available data. Therefore, the experimental group may use new Big Data Analytic tools that provide the user with more and newer data - like for example Kibana an Big Data visualization tool of Elastic - for new insights and decisions, while the control group may only use classic Business Intelligence - like for example Excel - for new insights and decisions, because this is the normal procedure.

Lastly, is the decision the participants had to make and how the success of this decision was measured. Both the experimental and control group had to make a choice which hotel, hostel, or bed & breakfast they will put on the outlet page - a webpage where HotelSpecials sums up 20 deals they think are attractive for their customers. The only requirement was that the hotel, hostel, and bed & breakfast should be available on the website of HotelSpecials. To sum up and illustrate foregoing, figure 2 was created. Figure 2 provides a clear overview of the design of this thesis experiment.

**Figure 2: Design of thesis experiment**



### 3.4 Reliability and Validity

- Big Data is relatively new and its meaning is also subjective and sometimes unclear, therefore it could lead to reliability and validity problems. This is the reason why it is very important to ask every participant of the interviews and experiments to define Big Data. This will increase the reliability and it will also make it easier to reproduce the results.
- This research is only limited to the e-commerce market of hotel accommodations, because this was the only information available for the researcher. Therefore, this could be a thread to the reliability and validity of the gathered data for this thesis.
- To receive reliable and valid data and get a better understanding regarding the points mentioned in the methodology section, the interview framework was conducted with the help of theory.

- To recognize and control the experiment for variables other than the dependent and independent variable, also called third variables, the participants also had to participate in a short survey, see Appendix C. In this survey the participants had to answer questions about their prior experience and their decision making style. Therefore, this experiment could recognize and control for any third variables.

## 4. Interview results

*This chapter will present the results of the interviews. In section 4.1 an overview will be given of the definition and the basics of Big Data, Business Intelligence, and Decision Making. Lastly, section 4.2 will provide examples and case analysis of Big Data, Business Intelligence, and Decision Making.*

### 4.1 Definition and basics of

#### 4.1.1 Big Data

First of all, nearly all respondents consider Big Data a “buzzword”. For example, one respondent stated that: “Everybody talks about Big Data, nobody really knows how to do Big Data, everybody thinks everyone else is doing Big Data, so everybody claims they are doing Big Data”. However, some respondents do not agree on that point of view. For instance, the respondents with an computer science or software engineering background do not see Big Data as a “buzzword”, but as a new movement that already has an impact on peoples life – like one respondent argued: “There are already enough practical examples of Big Data from big organizations like Google till start-ups that are using Big Data to make smart cities”.

All respondents agree with McAfee & Brynjolfsson (2012) that Big Data is about huge amounts of data. However, most respondents do not quantify what they think is “huge”. For example, one respondent stated: “Big Data is something of the last few years where datasets are growing and where datasets contain more information than a few years back”. This statement is of course true, but it is not quantified and less detailed than: “a few years back 1TB of data was considered huge, however today 1TB is not considered huge anymore. Therefore, you could see the volume of Big Data as a changing factor and where the organization will face problems with storing the data and loading the data into memory”.

McAfee & Brynjolfsson (2012) also consider Velocity and Variety as two areas of Big Data. Nearly all respondents mentioned that the rise of the mobile phone increased the speed of data creation, while only a few mentioned the unstructured form of this data. For instance, one respondent argued: “Take a twitter message from a mobile phone as an example. A twitter message is much more than just 140 characters. If you analyze the data, it has a lot of meta data – e.g. the geographical location of the tweet, hashtag of the tweet, the username of this

account, the number of followers. All this metadata is mostly unstructured and gets created pretty fast”.

Finally, some respondents also added a fourth V and a fifth V to the three V's of McAfee & Brynjolfsson (2012). They argued that it is also important to look at the trustworthiness of the data, which they called “Veracity”. In fact, the respondent who talked about twitters metadata also mentioned that Veracity is important, because a twitter message can contain typing mistakes, it can contain wrong hashtags, or the GPS sensor could create a wrong geographical location. Furthermore, all respondents that consider Big Data as a “buzzword” mentioned that having data is great, however they argued that it is also important that the organization is using this data to create new business opportunities and to turn the data into “Value”.

#### **4.1.2 Business Intelligence and Decision Making**

Firstly, the respondents defined Business Intelligence in a lot of different ways, however there was a clear trend in these definitions. All respondents considered Business Intelligence as a process that turns raw data into effective business information and new knowledge, which will then be used to make a decision that will lead to innovations. For example, one respondent argued: “our team of business analysts are making daily reports of our visitors and customers, which we will then review and use to adjust our organization accordantly”. Furthermore, most respondents argued that Business Intelligence enabled the organization to identify the customer's needs and therefore they could create more added value for the customer.

Secondly, the definition and examples above illustrate that most respondents combine Business Intelligence and Decision Making. In fact, most respondents failed to notice that their definition of Business Intelligence also contained a part of Decision Making. For instance, the respondent that argued to review and use a report is using a major part of decision making, because the respondent has to analyze this report and then define the available options which will help the organization to adjust accordantly. In addition, the respondent also has to make a rational or irrational decision between the available options, which also involves a part of decision making.

Thirdly, all respondents identified Business Intelligence as a critical component of their organization. In fact, one respondent stated that: “currently there is a lot of competition in every e-commerce branch, because it is not hard to create e-commerce organization these days.

Therefore, it is even more important that an organization does have the ability to gather data and transform this data into effective business actions faster as its competitors”. However, doing this requires the right mix of Business Intelligence systems and data collection methods. It also requires numerous of recourses to provide the right people, with the right information, at the right time. Therefore, one respondent argued that an organization should not make it to complicated and create a few dashboards, with a few metrics that will be measured across the organization.

Finally, like mentioned above some respondents acknowledge that Business Intelligence is using a lot of resources. In fact, one respondent argued that the Business Intelligence technology is developing so fast that capturing the data is not a bottleneck, however the capability of an organization to process all this data and adjust to all these new technologies is. In addition, other respondents argued that because of all the new Business Intelligence technologies they are facing an information overload. In fact, this is causing problems for a lot of organizations. For example, one respondent mentioned that organizations nowadays are creating a lot of new Business Intelligence dashboards, because they think it can be useful for their employees. Yet most of these dashboards provide inadequate information, inconsistent information, or even misleading information. Therefore, some respondents got the feeling that they cannot trust the Business Intelligence tools or use the data with its full potential.

### **4.1.3 Big Data, Business Intelligence, and Decision Making**

Since the respondents got questions about Big Data, Business Intelligence, and Decision Making the next step was to combine all fields. This is why all respondents were asked to reflect on the whole process and all three fields together. For the final conclusions it could be very useful to see if there are any patterns in how the respondents see Big Data in combination with and in comparison with Business Intelligence and Decision Making.

- Based on the respondents one can distinguish two groups. The first group is a group without a hardcore IT background. This group sees Big Data as a new way of analytics and the search for new intelligence from a dataset. In fact, some respondents even called it "Big Data Analytics". The second group is a group with a hardcore IT background like computer science or software engineering. These respondents see Big Data as an enabler of artificial intelligence, self-learning software, and smarter algorithms. One respondent stated: “An organization does not need more data for better

analytics or reporting, however an organization needs more data to get better self-learning software and train a machine”.

- The respondents that see Big Data as a new way of analytics also consider Business Intelligence and Decision Making as two related field. Furthermore, this group of respondents also see Big Data, Business Intelligence, and Decision Making as a critical component for the organization. This group of respondents also see Big Data, Business Intelligence, and Decision Making as one process. Where the first step in this process is Big Data with lots of data. The second step in this process are the Business Intelligence tools that make sense of all this data and that provide the user with useful business information. The third step in this process is Decision Making where the user makes a decision based on the information of the Business Intelligence tools. To conclude, this group argues that this is a sequential process from Big Data as the first step to the information and final decision as the last step.
- Finally, the respondents that see Big Data as an enabler of artificial intelligence, self-learning software, and smarter algorithms do not consider Business Intelligence and Decision Making as two related fields. In fact, some respondents thought this question was unclear or vague, because they do see Big Data as one whole new field. This group also confirms the importance of Business Intelligence and Decision Making, while they also acknowledge the need of more data for Business Intelligence tools. However, they clearly stated that more data for Business Intelligence and analytics does not necessarily have to be called Big Data. To illustrate, one respondent argued that: "Big Data has to be about lots and lots of data, creating a self-learning software tool that can take and analyze all twitter messages. This is what I call Big Data, because it will find patterns and correlations automatically and not with the help of human interaction or a Business Intelligence tool". In addition, these respondents also argued that the more data an organization puts in these self-learning software or algorithm the better it will perform, while if an organization puts in more data in the Business Intelligence tool the output will not explicitly be better.

## **4.2 Examples and the use of**

### **4.2.1 Big Data**

The first step an organization should take to get more value out of the available data with the help of Big Data is to make one definition of Big Data. Since respondents mentioned two types

of definitions, it can consume a lot of time but can also create confusion for all departments. In addition, getting one definition of Big Data does not mean that an organization has to choose between using Big Data Analytics or Big Data as an enabler of artificial intelligence, self-learning software, and smarter algorithms. Like one respondent stated: "An organization could have the best of both worlds". However, the goal of this first step should be to create more clarity and get rid of the tag "buzzword".

Respondents also argued that the next step an organization should take to get more value out of the available data with the help of Big Data is to train employees about the capabilities of Big Data. As an illustration, one respondent mentioned that this process starts with educating the employees throughout the whole organization. For example, if the IT department is creating a new algorithm to optimize the organizations real-time advertising bidding system, the department will need the help of the marketing department. However, if both departments have a different education level of Big Data it could lead to frustration and also to an inefficient algorithm. Another respondent mentioned: "some people are afraid that for example an automatic algorithm based on Big Data can take over their job, however one could also argue that this specific job is just changing and the employee has to adapt and work together with the automatic algorithm. In fact, it is important to see it this way, because the competitors will change and then the employee has to battle versus an automatic algorithm of a competitor". Furthermore, some respondents also argued that people do not have to be afraid, because human interactions are very important and can help to create an algorithm in the first play, check the outcome, make corrections, and improve the algorithm.

Finally, all respondents agreed that the next step an organization should take is to just start with deploying Big Data tools.

- The first sub step an organization should do is make a list of all Big Data tools that meet the magnetic, agile, and deep requirements. For instance, an organization could deploy free Big Data tools like Hadoop to store and enable self-learning algorithms or Kibana from Elastic to do Big Data Analytics.
- Secondly, one must make a requirements list and take the Big Data tool that fits the purpose. For example, an organization wants to build a self-learning recommendation algorithm based on historical data. If this is the case, than the selected Big Data tool should be able to collect all this historical data, but also to detect data contamination -



e.g. if the organization implements this recommendation algorithm on the first of January 2015, the tool should also be able to detect if the behavior of the users after January 2015 has been changed based on implementing this algorithm. To illustrate, if one buys a pair of shoes before January 2015 and after January 2015 one buys a pair of jeans because the recommendation algorithm give one that advice, this algorithm should be able to recognize that the costumer was not per definition looking for a new pair of jeans.

- Thirdly, the organization should not focus on developing smarter Big Data algorithms or smarter Big Data systems than its competitors. Like one respondent mentioned: "Do not try to reinvent the wheel. Try to use the knowledge and tools which already are available and build on top of that basic knowledge. However, an organization should focus on collecting more data as its competitors, because more data will in the end beat a really smart algorithm".
- Finally, all respondents agree that Big Data is not about replacing the organizations current warehouse and Business Intelligence platforms. In fact, it is all about complementing those warehouse and Business Intelligence platforms. For example, Hadoop is working perfectly with and next to traditional warehouse systems.

#### **4.2.2 Business Intelligence and Decision Making**

In general all respondents believe that the first step an organization should take to get more value out of Business Intelligence and Decision Making is to create an evidence based culture. The respondents argued that in this evidence based culture every employee has the opportunity to come up with an idea or hunch. Next, the employee will formulate a problem statement and hypothesis for the organization. As an example, one respondent mentioned that: "currently our organization is formulating a problem statement and hypothesis, because we believe that a specific reassurance text before selecting a product could lead to a performance increase". After creating the problem statement and hypothesis, the employee has to collect data to confirm or refute the hypothesis. Most respondents agreed that an evidence based culture gives more power to the employees and could create new opportunities based on facts and not on hunch or based on experience.

Respondents also argued that the next step an organization should take to get more value out of Business Intelligence and Decision Making is to also focus on prescriptive analytics and predictive analytics. According to the respondents most organizations are focusing on "what

happened?" and "what is happening right now?", also known as descriptive analytics. However, if an organization wants to maximize the added value of Business Intelligence and Decision Making they also have to focus on "why did it happen?", prescriptive analytics, and focus on "what will happen?", predictive analytics. Therefore, the organization will gain more knowledge about the customer and the market segment, which help to gain a competitive advantage with the help of smart Business Intelligence and Decision Making.

Furthermore, all respondents agreed that visualization is the most important step in getting more value out of Business Intelligence and Decision Making. Most respondents mentioned that visualization of data is the most important step in transforming data into knowledge. As one respondent argued that people do not take the time to go through every tab or through every table, they demand an answer as soon as possible and an answer that is easy to understand. Therefore, an organization could deploy multiple tools that allow an organization to create interactive data visualization. Another remarkable statement by one of the respondents was "with the help of Tableau, an interactive data visualization tool, our organization can now deploy dashboards that contain multiple visualizations. These visualization are presented in an understandable and easy-to-grasp format so the user really understand the data".

## 5. Experiment results

*In this chapter the results of the experiment will be discussed and analyzed. In section 5.1 an overview will be given of the starting survey. Section 5.2 and 5.3 will elaborate on the two processes, the data collection process and selection process, of the experiment. Lastly, section 5.4 will provide an overview of the financial results of the experiment.*

### 5.1 Survey

The first step the participants had to do, was to fill out a survey. This survey was then used to make the teams as even as possible and to control for third variables. In this survey, see Appendix C, the participants had to answer questions about their prior experience and their decision making style. Next step was to create comparable groups based on the scores of this survey. In total 6 participants took part in this experiment, which is a participation rate of 100 percent - in other words these 6 participants are beside this experiment the ones who are already responsible for selecting the outlet deals on HotelSpecials. This is why both the control group and the experimental group had 3 participants.

**Table 2: Survey results of the control group and experimental group**

	Control group	Experimental group
Number of participants	3	3
Work experience	3	3
The use of analytics	3,67	3,67
Keen on analytics	3,67	3,33
Decision Making style	3,33	3,33
Opportunities for analytics	3	2,67
Company vision	2,67	2,67
Company usage of analytics	3	2,67
Company composition	2,67	3
Company culture	2,67	3
Company support	3,33	3,67

While analyzing the results, one can assume that the control group and the experimental group both had a similar level of work experience, both use analytics in the same way, and both have

a similar decision making style. In fact, this is not a coincidence but done on purpose because of the low number of participants and the need for equal groups. Therefore, these variables now show a comparable pattern and one can assume that the control group and the experimental group are as even as possible.

After making the groups and measuring the characteristics of both groups, the groups now had to participate in the experiment. The data was gathered by observing and interviewing the participants and by measuring the groups on financial results. The first step is elaborating on the observations and then measure the groups on the financial results. For example, the conversion rate of a product is an important financial metric in the e-commerce branch. Therefore, the conversion can help this experiment to measure the financial performance of both groups. The conversion rate of a product in this experiment is calculated by the number of bookings divided by the number of unique page views.

## 5.2 Data collection

The first group process in this experiment was the collection of the data they wanted to use for the selection process. Like mentioned in the methodology, group 1 can only use Business Intelligence data that already was available before the experiment, while group 2 may use new data and the Big Data Analytic tool. This resulted in the following distribution of data:

**Table 3: Collected data sources by group**

Group 1 (control group)	Group 2 (experimental group)
Data: Historical table of all the bookings HotelSpecials made including the customer information.	Data: Historical table of all the bookings HotelSpecials made including the customer information.
Goal: To gain knowledge about the good booking hotels and to gain customers about the target group.	Goal: To gain knowledge about the good booking hotels and to gain customers about the target group.
Data: Historical table of all performed searches in the search box on Hotelspecials.	Data: Historical table of all performed searches in the search box on Hotelspecials.
Goal: To gain knowledge about the demand in various regions and in specific seasons.	Goal: To gain knowledge about the demand in various regions and in specific seasons.
Data: A table of the availability of a hotel on	Data: A table of the availability of a hotel on

Hotelspecials. Goal: To gain knowledge and link the demand to the availability.	Hotelspecials. Goal: To gain knowledge and link the demand to the availability.  Data: data from Kibana, a Big Data Analytics tool for all real-time events on the HotelSpecials website. Goal: To gain additional information about real-time customer activities.
------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

As expected by the survey, the control group and the experimental group collected the same type of information. As a researcher it was very interesting to notice that both groups picked the same data sources, in a similar order, and with a similar style of reasoning.

### 5.3 Selection process

After the groups decided how they were going to get their information, they had to analyze and, interpret this data, and had to make a decision based on this data. Based on the survey one could argue that this process would be approximately the same, however based on an observation the opposite has been proven to be true. A short summary is given in table 4. While both groups had a different process, the outcome was the same. Group 1 just needed more time than group 2 for selecting the deals they thought were appropriated.

**Table 4: Observation of the analytical process**

Group 1 (control group)	Group 2 (experimental group)
Group 1 started by focusing on each data source separately. They were analyzing and examining the information by finding relationships, and searching for trends on all data sources individually. For example, group 1 started with the historical table and was trying to reveal specific trends for a specific hotel, while not looking at the search box or availability table. Therefore, it took group 1 a	Group 2 started by gathering all the information from all data sources and making an quick scan of this available information. They were then analyzing and examining the information, finding relationship, searching for trends on all data sources. Therefore, group 2 could analyze and examine the information much easier and faster. In addition, group 2 made good use of the new

lot of time to find the information they needed.

Big Data Analytic tool. Group 2 used it to analyze and gain knowledge about the real-time demand.

## 5.4 Financial results

When looking at survey, data collection process, and selection process of the control group and experimental group the differences between the two groups can be neglected. However, when looking at the financial results the difference between the two groups is huge.

The following data was measured by Google Analytics and collected about a 4 day period:

**Table 5: Financial results group 1 (control group):**

Hotel	Page views	Bookings	Conversion rate
Hotel 1	6498	0	0,00%
Hotel 2	349	0	0,00%
Hotel 3	250	2	0,80%
Hotel 4	198	1	0,51%
Hotel 5	99	1	1,01%
Total	7394	4	0,05%

**Table 6: Financial results group 2 (experimental group):**

Hotel	Page views	Bookings	Conversion rate
Hotel 1	6498	2	0,03%
Hotel 2	3046	5	0,16%
Hotel 3	870	5	0,57%
Hotel 4	250	3	1,20%
Hotel 5	91	0	0,00%
Total	10755	15	0,14%

*Disclaimer:*

*The results are anonymized, because else it could damage the hotels reputation. In addition, the hotels are not participating in this experiment only HotelSpecials did.*

As shown in table 5 and table 6, the difference between the control group and the experimental group is huge. Looking at the difference in page views, the number of times an unique visitor has visited the hotel page, between group 2 and group 1 is + 45,46% ( $(10755-7394) / 7394$ ). Furthermore, the bookings are also +275,00% higher for group 2. This is somewhat misleading, because group 2 generated more page views the booking measurement is not a comparable financial metric -e.g. a store that attracts 1000 visitors is more likely to sell something than a store that only attracts 10 visitors. Therefore, the conversion rate is calculated. The conversion rate for group 2 is 180,00% higher than group 1.

To conclude, the hypothesis that started this experiment was: “during the day customers were searching on their mobile and would finish the order on their desktop in the evening”. The general impression is that the hypothesis is true, however this experiment is only conducted once and is only conducted within HotelSpecials. In addition, the results of this experiment show that an organization could gain a huge performance increase with the help of Big Data Analytic tools, because it helps with receiving more and newer data to understand the customers and help to timely transform all available information in to effective business information.

## 6. Explanatory analysis

*The purpose of this explanatory analysis is to relate the findings from the interviews and experiments with the literature to extract relationships, similarities, and differences.*

In general, people still consider Big Data as a new subject and research area. For example, in the interviews nearly all respondents mentioned that they see Big Data as “buzzword”. However, in the literature there are enough examples and cases that go beyond the buzzword. For instance, a born digital organization could have an immense databases from just a single source like clickstreams and use them for targeted advertising or for optimizing their offers (Boyd & Crawford, 2012; Cohen et al., 2009). In addition, most researchers define Big Data as the 3V's of McAfee & Brynjolfsson (2012) – Volume for the huge amounts of data, Velocity for the speed of data creation, and Variety for the unstructured form of this data. Most respondents agree with this point of view, but also add Veracity for the trustworthiness of the data and Value for the ability to use this data and create new business opportunities. Therefore, it can be concluded that these respondents automatically or unintentionally call Big Data a buzzword, because their definition of Big Data match perfectly and in the theory there is also enough information available.

Furthermore, the respondents also see different capabilities of Big Data because they still consider Big Data as a new subject and research area. In the literature, Big Data, Business Intelligence, and Decision Making are considered as three strongly related research areas. For example, in 1977 Simon already introduced his normative model of decision making that provides a clear overview of the link between Big Data, Business Intelligence, and Decision Making. However, based on the respondents one could also argue that there are two streams within the Big Data community. The first stream are people without an IT background, like computer science or software engineering, argue that Big Data is related to Business Intelligence and Decision Making and the second stream are respondents with an IT background that see Big Data as an enabler of artificial intelligence, self-learning software, and smarter algorithms. Therefore, it can be concluded that this thesis cannot confirm the link between Big Data, Business Intelligence, and Decision Making that is visible in Simon (1977).



The respondents also mentioned how an organization could start with using Big Data. The most important steps are: (i) train employees about the capabilities of Big Data (ii) just start with Big Data and learn about Big Data tools while implementing and using them (iii) make a list of all Big Data tools that meet the MAD requirements (Cohen et al., 2009) (iv) choose the right tool that fits the purpose of the organization (v) and do not focus on developing smarter Big Data systems or smarter Big Data algorithms than the competitors. In addition, the “starting guide” of McAfee & Brynjolfsson (2012) can help an organization with phase two by not just starting with Big Data and learn about Big Data tools while implementing and using them, but to systematically: (i) create a team and to pick a test department with an open-minded and data friendly manager (ii) begin a brainstorm session or meeting and identify and select no more than five opportunities that can be solved with Big Data within five weeks (iii) start an innovation process with the following steps: experimentation, measurement, sharing, and replication (iv) and if possible, send out some analytic challenges on their Big Data to third parties (McAfee & Brynjolfsson, 2012). Therefore, it can be concluded that the literature confirmed and provided additional information to the answers of the respondents.

Building further upon this, the results described in this thesis also indicates that if an organization starts with Big Data, Business Intelligence, and Decision Making the organization should not replace their current warehouse and platforms. These findings are in line with Jacobs (2009) and Madden (2012). Like Madden (2012) mentioned the standard solutions most organizations rely on can already deal with sizes up to multiple petabytes, big enough for billions of log records, clickstreams, or transaction data. Jacobs (2009) added that trouble when an business analysts or data scientists want to take all these log records and analyze these within seconds or minutes. Therefore, an organization should not replace the current warehouse and platforms, but use it as an addition for analytical questions that need a lot of data within seconds or minutes.

The results of the interviews also indicated that it is important to create a vision, create an alignment, and evidence based culture. As indicated by Watson & Wixom (2007) organizations are more likely to have success with Big Data, Business Intelligence, and Decision Making when the management of an organizations creates a vision and alignment between business strategies, business model, and Business Intelligence tools. In addition, the recognition of Big Data, Business Intelligence, and Decision Making has led to a growing enthusiasm for evidence

based decision making (Labrinidis & Jagadish, 2012). In this evidence based culture employees are stimulated to define the problem, create a hypothesis, and collect data to confirm or refute this hypothesis. This is also in line with the findings of LaValle et al. (2011) that organization that identify Big Data, Business Intelligence, and Decision Making as a differentiation strategy are twice as likely to be a top performer in their market segment.

The understanding of customers increased dramatically once shopping moved online (McAfee & Brynjolfsson, 2012). In fact, the results of this thesis experiment indicated that making timely decisions and understanding your customers can create a huge performance increase. For example, one experiment in this thesis was based on the available data and the hypothesis that: “during the day customers were searching on their mobile and would finish the order on their desktop in the evening”. The experiment indicated that this hypothesis was true and that an organization could gain a huge performance increase when they start to understand their customers. This is also confirmed by Lönnqvist & Pirttimäki (2006) who argued that it is essential to understand the customer and timely transform all information in effective business information. Therefore, an organization should not only ask question like "what happened?" and "what is happening right now?", but also focus on "why did it happen?" – prescriptive analytics – and "what will happen?" – predictive analytics (LaValle et al., 2011).

Lastly, the results of the interviews indicated that it is important to maximize the visualization of data. For example, organizations can create frontend applications with a visual dashboard that allow a decision maker to track key performance indicators of their operations (Chaudhuri et al., 2011). In fact, the challenge to transform all this information to effective business information becomes more difficult as the information keeps growing exponentially and the increasing amount of employees who need access to this information (Cody et al., 2002). Therefore, all data that can be visualized should be visualized and presented in an understandable and easy-to-grasp format.

## 7. Conclusion and recommendations

*Chapter 7 will give an overview of the conclusions and recommendations. Section 7.1 will provide an clear overview of the conclusions of this thesis. Section 7.2 will describe the contributions to the literature. Lastly, section 7.3 will elaborate on the limitations of this thesis and the possibilities for future research.*

### 7.1 Conclusion

In this thesis, research was conducted about various components of Big Data, Business Intelligence, and Decision Making. The main purpose of this research was to create a set of recommendations that will help an organization to start with Big Data to get more value out of the available data and use Business Intelligence in such a way that it will be more frequently used for new insights and decisions.

The research question that fits this goal was:

*How can an organization start with Big Data to get more value out of the available data and optimize the Business Intelligence processes in such a way that it will be more frequently used for decisions?*

The results regarding Big Data, Business Intelligence, and Decision Making are:

- In general, Big Data is still consider as a new subject and research area. For example, in the interviews nearly all respondents mentioned that they see Big Data as “buzzword”. To counter this, an organization should set a clear definition of Big Data. In the literature, most researchers define Big Data as the 3V's of McAfee & Brynjolfsson (2012) – Volume for the huge amounts of data, Velocity for the speed of data creation, and Variety for the unstructured form of this data. Most respondents agree with this point of view, but also add Veracity for the trustworthiness of the data and Value for the ability to use this data and create new business opportunities.
- Furthermore, the respondents also see different capabilities of Big Data because they still consider Big Data as a new subject and research area. In the literature, Big Data, Business Intelligence, and Decision Making are considered as three strongly related research areas (Simon, 1977). However, based on the respondents two streams can be distinguished. The first stream are people without an IT background, like computer

science or software engineering, argue that Big Data is related to Business Intelligence and Decision Making and the second stream are respondents with an IT background that see Big Data as an enabler of artificial intelligence, self-learning software, and smarter algorithms.

- An organization could start with Big Data, Business Intelligence, and Decision Making by: (i) selecting a test department with an open-minded and data friendly manager (ii) identifying and selecting no more than five opportunities that can be solved with Big Data within five weeks (iii) starting an innovation process with the following steps: experimentation, measurement, sharing, and replication (iv) and if possible, send out some analytic challenges on their Big Data to third parties (McAfee & Brynjolfsson, 2012). In addition, the results indicate that an organization has to: (v) train employees about the capabilities of Big Data (vi) start with Big Data and learn about Big Data tools while implementing and using them (vii) make a list of all Big Data tools that meet the MAD requirements (Cohen et al., 2009) (viii) choose the right tool that fits the purpose of the organization (ix) and do not focus on developing smarter Big Data systems or smarter Big Data algorithms than the competitors.
- Building further upon this, the results also indicate that if an organization starts with Big Data, Business Intelligence, and Decision Making the organization should not replace their current platforms. Organizations should use it as an addition for analytical questions that need a lot of data within seconds or minutes.
- The results of the interviews indicate that it is important to create an evidence based culture. In this evidence based culture employees are stimulated to define the problem, create a hypothesis, and collect data to confirm or refute this hypothesis. Organizations are also more likely to have success with Big Data, Business Intelligence, and Decision Making when the management of an organizations creates a vision and alignment between business strategies, business model, and Business Intelligence tools.
- The results of this thesis experiment indicated that making timely decisions and understanding your customers can create a huge performance increase. For example, the experiment in this thesis was based on the available data and the hypothesis that: “during the day customers were searching on their mobile and would finish the order on their desktop in the evening”. The experiment indicated that this hypothesis was true and that an organization could gain a huge performance increase when they start to

understand their customers. Therefore, an organization should not only ask question like "what happened?" and "what is happening right now?", but also focus on "why did it happen?" – prescriptive analytics – and "what will happen?" – predictive analytics.

- Lastly, the results indicate that it is important to maximize the visualization of the available data. These visualization should be presented in an understandable and easy-to-grasp format.

## **7.2 Contribution to literature**

This thesis makes several contributions to the literature.

The first contribution is the performed literature review. From a scientific perspective the performed literature review on Big Data, Business Intelligence, and Decision Making is a contribution to the literature. It provides other researchers with a clear overview of the available articles, papers, and other scientific material.

The second contribution are the results of the performed interviews and the results of the performed experiment. Big Data is still a relatively new research topic and therefore the results of this thesis can contribute to the Big Data literature. The conducted interviews can provide researchers some empirical evidence on how people think and feel about Big Data. In addition, these interviews can also help to confirm or deny interesting patterns between Big Data, Business Intelligence, and Decision Making.

The last and most important contribution of this thesis is the development of a set of recommendations for other organizations. The aim of these recommendations is to help other organizations to start with Big Data in a more systematic manner. These recommendations are based on literature, the performed interviews, and an experiment. The main difference with other studies is that this thesis goes beyond the "buzzword" and has a more practical approach to Big Data.

## **7.3 Limitations and future research**

This thesis has some limitations, because of the limitations in time, resources, and in the used methods.

The first limitation is the case selection and the sample size. An important aspect of the level of generalization of a thesis is the case selection and the sample size. Since this thesis only made use of two groups of cases, there is a chance that these two groups do not cover all the variety of cases in the whole population. In addition, all cases were from a single country and within a single branch. Therefore, future research should be conducted with more cases, a bigger sample size, and cases from multiple countries and branches.

The second limitation of this thesis is the design of the experiment and the design of the interviews. The research design of this thesis experiment is threatening the internal validity, because the control group and the experimental group were not assigned randomly to one of these groups. However, to control for this limitation this thesis made use of a survey to control for most of the third variables and to make the teams as even as possible. The research design of the interviews is also threatening the internal validity, because nearly all interviews were conducted in Dutch. Translations had to be made during and after the interviews, to use these interviews for this thesis. During these translation stages a change in content or context could have occurred. Therefore, future research should be conducted with an experiment where all the participants get randomly assigned the control group and the experimental group or conduct an interview where the researcher does not have to translate the interviews.

The last limitation is that the generalization of these results are somewhat limited. Due to the nature of this study and the points mentioned above, the results are not generalizable for all organizations. Though, most organizations could tweak the set of recommendations of this thesis and will therefore still benefit from this thesis. Therefore, future research should focus on improving these recommendations with the help of a bigger sample and more cases from different countries and branches.

## References

- Ariely, D. (2013). Big data is like teenage sex. Retrieved from <https://twitter.com/danariely/status/287952257926971392>
- Babbie, E. R. (2007). *The practice of social research*. Belmont, CA.: Thomson Wadsworth.
- Bellman, R. E., & Zadeh, L. A. (1970). Decision-Making in a Fuzzy Environment. *Management Science*, 17(4), B–141. <http://doi.org/10.1287/mnsc.17.4.B141>
- boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. <http://doi.org/10.1080/1369118X.2012.678878>
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology. *Commun. ACM*, 54(8), 88–98. <http://doi.org/10.1145/1978542.1978562>
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q.*, 36(4), 1165–1188.
- Chung, W., Chen, H., & Nunamaker, J. F., Jr. (2005). A Visual Framework for Knowledge Discovery on the Web: An Empirical Study of Business Intelligence Exploration. *J. Manage. Inf. Syst.*, 21(4), 57–84.
- Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). The Integration of Business Intelligence and Knowledge Management. *IBM Syst. J.*, 41(4), 697–713. <http://doi.org/10.1147/sj.414.0697>
- Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. (2009). MAD Skills: New Analysis Practices for Big Data. *Proc. VLDB Endow.*, 2(2), 1481–1492. <http://doi.org/10.14778/1687553.1687576>
- Cuzzocrea, A., Song, I.-Y., & Davis, K. C. (2011). Analytics over Large-scale Multidimensional Data: The Big Data Revolution! In *Proceedings of the ACM 14th International Workshop*

- on Data Warehousing and OLAP* (pp. 101–104). New York, NY, USA: ACM.  
<http://doi.org/10.1145/2064676.2064695>
- Duan, L., & Xu, L. D. (2012). Business Intelligence for Enterprise Systems: A Survey. *IEEE Transactions on Industrial Informatics*, *8*(3), 679–687.  
<http://doi.org/10.1109/TII.2012.2188804>
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417.  
<http://doi.org/10.1037/h0053870>
- Elbashir, M. Z., Collier, P. A., & Davern, M. J. (2008). Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. *International Journal of Accounting Information Systems*, *9*(3), 135–153.  
<http://doi.org/10.1016/j.accinf.2008.03.001>
- Evans, J. S. B. T. (1984). Heuristic and analytic processes in reasoning\*. *British Journal of Psychology*, *75*(4), 451–468. <http://doi.org/10.1111/j.2044-8295.1984.tb01915.x>
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395. <http://doi.org/10.3758/BF03193858>
- Flickr. (2014). How many public photos are uploaded to Flickr every day, month, year?  
Retrieved August 16, 2015, from  
<https://www.flickr.com/photos/franckmichel/6855169886/>
- Galletta, A., & Cross, W. E. (2013). *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication*. New York: NYU Press.
- Gartner. (2014). Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business. Retrieved August 16, 2015, from  
<http://www.gartner.com/newsroom/id/2819918>
- Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B., & Babu, S. (2011). Starfish: A Self-tuning System for Big Data Analytics. In *In CIDR* (pp. 261–272).



- Jacobs, A. (2009). The Pathologies of Big Data. *Commun. ACM*, 52(8), 36–44.  
<http://doi.org/10.1145/1536616.1536632>
- James, W. (1890). *The principles of psychology*. New York: H. Holt and Company.
- Kahneman, D. (2003). A Perspective on Judgment and Choice: Mapping Bounded Rationality. *The American Psychologist*, 58(9), 697–720. <http://doi.org/10.1037/0003-066X.58.9.697>
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341–350. <http://doi.org/10.1037/0003-066X.39.4.341>
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and Opportunities with Big Data. *Proc. VLDB Endow.*, 5(12), 2032–2033. <http://doi.org/10.14778/2367502.2367572>
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MITSloan Management Review*, 52(2).
- Lee, J. H., & Park, S. C. (2005). Intelligent Profitable Customers Segmentation System Based on Business Intelligence Tools. *Expert Syst. Appl.*, 29(1), 145–152.  
<http://doi.org/10.1016/j.eswa.2005.01.013>
- Lönnqvist, A., & Pirttimäki, V. (2006). The Measurement of Business Intelligence. *Information Systems Management*, 23(1), 32–40.  
<http://doi.org/10.1201/1078.10580530/45769.23.1.20061201/91770.4>
- Madden, S. (2012). From Databases to Big Data. *IEEE Internet Computing*, 16(3), 4–6.  
<http://doi.org/10.1109/MIC.2012.50>
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*.
- Rivest, S., Bédard, Y., Proulx, M.-J., Nadeau, M., Hubert, F., & Pastor, J. (2005). SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(1), 17–33. <http://doi.org/10.1016/j.isprsjprs.2005.10.002>

- Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, 48(1), 9–26. [http://doi.org/10.1016/0377-2217\(90\)90057-1](http://doi.org/10.1016/0377-2217(90)90057-1)
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7–59. <http://doi.org/10.1007/BF00055564>
- Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138. <http://doi.org/10.1037/h0042769>
- Simon, H. A. (1977). *The New Science of Management Decision*. Upper Saddle River, NJ, USA: Prentice Hall PTR.
- Stanovich, K. E., & West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral and Brain Sciences*, 23(5), 645–665.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281–299. <http://doi.org/10.1037/h0032955>
- Watson, H. J., & Wixom, B. H. (2007). The Current State of Business Intelligence. *Computer*, 40(9), 96–99. <http://doi.org/10.1109/MC.2007.331>
- Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22(1), 45–55. <http://doi.org/10.1057/ejis.2011.51>
- Yin, R. K. (2009). *Case study research: design and methods*. Los Angeles [etc.]: Sage.

## Appendix A: search results scopus

### Search term 1:

TITLE(Big Data) AND ( LIMIT-TO(SUBJAREA,"COMP" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"DECI" ) OR LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"PSYC" ) )

Author(s)	Title	Used?
Chen et al. (2012)	Business intelligence and analytics: from big data to big impact	Yes
Boyd and Crawford (2012)	Critical questions for big data	Yes
Cohen et al. (2009)	Mad skills: new analysis practices for big data	Yes
Schulenberg et al. (2005)	Trajectories of marijuana use during the transition to adulthood: the big picture based on national panel data	No
Jacobs (2009)	The pathologies of big data	Yes
Goldberg (2001)	Analyses of digman's child-personality data: derivation of big-five factor scores from each of six samples	No
McAfee and Brynjolfsson (2012)	Big data: the management revolution	Yes
Herodotou et al. (2011)	Starfish: a selftuning system for big data analytics	Yes
McAfee and Brynjolfsson (2012)	Big data: the management revolution	No
Madden (2012)	From databases to big data	Yes
Xiaofeng and Xiang (2013)	Big data management: concepts, techniques and challenges	No
Chen et al. (2012)	Interactive analytical processing in big data systems: a crossindustry study of mapreduce workloads	No
Cuzzocrea et al. (2011)	Analytics over large-scale multidimensional data: the big data revolution!	Yes
Lavalle et al. (2011)	Big data, analytics and the path from insights to value	Yes

Labrinidis and Jagadish (2012)	Challenges and opportunities with big data	Yes
--------------------------------	--------------------------------------------	-----

### Search term 2

TITLE(Business Intelligence) AND ( LIMIT-TO(SUBJAREA,"COMP" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"DECI" ) OR LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"PSYC" )

Author(s)	Title	Used?
Grigori et al. (2003)	Business process intelligence	No
Chen et al. (2012)	Business intelligence and analytics: from big data to big impact	Yes
Watson and Wixom (2007)	The current state of business intelligence	Yes
Bonabeau and Meyer (2001)	Swarm intelligence: a whole new way to think about business	No
Cody et al. (2002)	The integration of business intelligence and knowledge management	Yes
Chaudhuri et al. (2011)	An overview of business intelligence technology	Yes
Elbashir et al. (2008)	Measuring the effects of business intelligence systems: the relationship between business process and organizational performance	Yes
Chung et al. (2005)	A visual framework for knowledge discovery on the web: an empirical study of business intelligence exploration	Yes
Rivest et al. (2005)	Solap technology: merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data	Yes
Jourdan et al. (2008)	Business intelligence: an analysis of the literature	No
Shi et al. (2005)	Classifying credit card accounts for business intelligence and decision making: a multiple-criteria quadratic programming approach	No
Baars and Kemper (2008)	Management support with structured and unstructured data—an integrated business	No

	intelligence framework	
Lönnqvist and Pirttimäki (2006)	The measurement of business intelligence	Yes
Lee and Park (2005)	Intelligent profitable customers segmentation system based on business intelligence tools	Yes
Duan and Xu (2012)	Business intelligence for enterprise systems: a survey	Yes

### Search term 3

TITLE(Decision OR Choice) AND ( LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"DECI" ) OR LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"PSYC" )

Author(s)	Title	Used?
Thompson et al. (1994)	Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice	No
Charnet et al. (1978)	Measuring the efficiency of decision making units	No
Myers and Majluf (1984)	Corporate financing and investment decisions when firms have information that investors do not have	No
Bellman and Zadeh (1970)	Decision-making in a fuzzy environment	Yes
Kahneman and Tversky (1984)	Choices, values, and frames	Yes
Saaty (1990)	How to make a decision: the analytic hierarchy process	Yes
Lent et al. (1994)	Toward a unifying social cognitive theory of career and academic interest, choice, and performance	No
Thaler (1980)	Toward a positive theory of consumer choice	No
Kahneman (2003)	A perspective on judgment and choice	Yes
Charles et al. (1997)	Shared decision-making in the medical encounter: what does it mean? (or it takes at	No

---

	least two to tango)	
Simon (1956)	Rational choice and the structure of the environment	Yes
Nicholls (1984)	Achievement motivation: conceptions of ability, subjective experience, task choice, and performance	No
Samuelson and Zeckhauser (1988)	Status quo bias in decision making	Yes
Chen (2000)	Extensions of the topics for group decision-making under fuzzy environment	No
Tversky (1972)	Elimination by aspects: a theory of choice	Yes

---

## Appendix B: interview framework

	Big Data	Business Intelligence	Decision making
<b>Definition (1)</b>	How do you define Big Data?	How do you define Business intelligence?	How do you define an choice or decision?
<b>Definition (2)</b>	Do you see Big Data, Business Intelligence, and the final decision as three separate processes or do you see this as one process from Big Data with the help of Business Intelligence to the final decision? And why?		
<b>Current state</b>	<p>How do you currently use the available data?            What are the available tools?            Which tools and functionalities are according to you missing in the current Business Intelligence processes? And in which way does this affect you?            Did use Big Data or Business Intelligence this month? And in which way did this help you to create a final decision?</p>		
<b>Problem identification</b>	<p>Technology is often mentioned as one of the main areas to solve the Big Data problem (Madden, 2012)</p> <p>Do you agree or disagree with this view? And why?            And in which way can an organization use this knowledge for Big Data and Business Intelligence processes?</p>	<p>Organizations are more likely to have success with Business Intelligence when the following conditions exist:</p> <ol style="list-style-type: none"> <li>1. Management of an organizations should have a vision for Business Intelligence and believe in information-based decision making.</li> <li>2. The use of Business Intelligence and analytics should be part of the organizational culture and counter decision making based on intuition or “gut feelings”.</li> <li>3. Alignment between business strategies, business model, and Business</li> </ol>	<p>There are two reasoning systems within the human brain. System 1 a fast and emotional system also known as intuition and system 2 a slow and non-emotional system also known as rationality (Kahneman, 2003)</p> <p>In which way can an organization use this knowledge for Big Data and Business Intelligence processes?</p>

---

Intelligence strategies enables an organization to create organizational change and new business opportunities.

4. An organization should have a strong and effective Business Intelligence governance and infrastructure, because it will address business alignment, funding, project prioritization, and data quality.

5. Lastly, an organization needs to provide users with appropriate Business Intelligence tools for their needs and give effective training and support to these users.

(Watson & Wixom, 2007)

Do you agree or disagree with this view? And why?

And in which way can an organization use this knowledge for Big Data and Business Intelligence processes?

---



## Appendix C: survey design

### General information

Name	.....
Gender (Male/Female)	.....
Date of birth	.....
Department	.....

### Personal

Little work experience	1	2	3	4	5	Lot of work experience
I do not use analytics	1	2	3	4	5	I use analytics
I do not like analytics	1	2	3	4	5	I like analytics
I use experience for my decisions	1	2	3	4	5	I use data for my decisions
I see no new opportunities for analytics	1	2	3	4	5	I see new opportunities for analytics

### Company

Management has no vision for information-based decision making	1	2	3	4	5	Management has a vision for information-based decision making
The use of analytics is no part of the organization	1	2	3	4	5	The use of analytics is part of the organization
There is no link between analytics, business strategy, and new business opportunities	1	2	3	4	5	There is a link between analytics, business strategy, and new business opportunities
Company has an experience driven culture	1	2	3	4	5	Company has a data driven culture
I get little analytical support	1	2	3	4	5	I get a lot of analytical support