Viljami Kaasalainen

BUSINESS INTELLIGENCE SYSTEM IMPLEMENTATION AND DESIGN FRAMEWORK: A PUBLIC SECTOR CASE STUDY

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2020

ABSTRACT

Kaasalainen, Viljami
Business intelligence system implementation and design framework: a public sector case study
Jyväskylä: University of Jyväskylä, 2020, 90 pp.
Information Systems, Master's Thesis
Supervisor(s): Pulkkinen, Mirja

Business intelligence could be considered a well-researched topic, but there is a limited amount of studies about BI systems and their implementation. Especially studies that consider BI systems outside of the technical point of view are rare. This research will do so by examining BI system implementation from the view of a project manager. Common technical decisions, design choices, and considerations of BI projects were defined and explained, but also guidance and recommendations based on previous studies were provided. This alone can be considered as a sufficient justification for a research topic, but an additional theme is present in this study. The relation and differences between BI in the public and private sectors were studied and the initial hypothesis expected clear differences. The main objective of the research was to create a framework, that was designed and developed by following the design science research methodology and by utilizing the DSRM process model. The framework includes information about different components and parts of BI systems, such as data sources, data processing methods, data warehouses, software, users, and training. The framework was continuously developed and tested in a real business environment, and the implementation project and the research study progressed in tandem. The process took months and it allowed a very in-depth analysis. The framework was evaluated as being sufficient enough to assist with common problematics and questions of BI implementation projects and considering a public sector case in the context of BI system implementation, the conclusion is that the differences compared to a private sector case are minor.

Keywords: business intelligence, implementation, system architecture, project management, data warehouse, training, ETL (extract-transform-loading)

TIIVISTELMÄ

Kaasalainen, Viljami
Viitekehys liiketoimintatiedon hallintajärjestelmien käyttöönottoon ja suunnitteluun
Jyväskylä: Jyväskylän yliopisto, 2020, 90 s.
Tietojärjestelmätiede, pro gradu -tutkielma
Ohjaaja(t): Pulkkinen, Mirja

Liiketoimintatiedon hallintaa on yleisesti ottaen tutkittu suhteellisen paljon, mutta siihen liittyvä järjestelmä ja sen käyttöönotto on saanut vähemmän huomiota. Erityisesti sellaisia tutkimuksia, jotka tarkastelevat BI-järjestelmiä teknisen näkökulman ulkopuolella, on harvassa. Tämä tutkimus paikkaa kyseistä aukkoa käsittelemällä BI-järjestelmien käyttöönottoa projektipäällikön näkökulmasta. Tutkimuksessa käsitellään erilaisia kysymyksiä, harkintoja ja ongelmia liittyen BI-projekteihin, sekä kyseisiin valintoihin tarjotaan opastusta ja suosituksia perustuen aikaisempiin tutkimuksiin. Tällainen lähestymistapa itsessään olisi jo riittävä peruste uudelle tutkimukselle, mutta sen lisäksi tutkimuksessa käsitellään julkisen ja yksityisen sektorin eroja liiketoimintatiedon hallinnassa. Alustava hypoteesi kysymykseen on, että sektorien välillä löytyisi selkeitä eroja. Tutkimuksen tärkeimpänä tavoitteena oli luoda viitekehys, joka suunniteltiin ja toteutettiin seuraamalla suunnittelututkimuksen periaatteita ja hyödyntämällä DSRM-prosessimallia. Viitekehyksen tarkoituksena on vastata käyttöönottoprojektien tärkeimpiin kysymyksiin, ja se sisältää tietoa BI-järjestelmien tietolähteistä, tiedonkäsittelystä, tietovarastoista, sovelluksista, käyttäjistä ja koulutuksesta. Sen kehittäminen ja testaaminen toteutettiin erään julkisen sektorin organisaation käyttöönottoprojektissa. Viitekehys ja projekti etenivät yhtäaikaisesti, ja kehittämisprosessi kesti kuukausia, mikä mahdollisti perusteellisen analysoinnin. Viitekehyksen arvioitiin olevan riittävän kattava ja hyödyllinen, jotta sen avulla on mahdollista toteuttaa BI-järjestelmän käyttöönotto ja vastata siihen liittyviin tärkeisiin kysymyksiin. Lisäksi sektorien välisten erojen liittyen BI-järjestelmiin todettiin olevan vähäisiä.

Asiasanat: liiketoimintatiedon hallinta, käyttöönotto, järjestelmäarkkitehtuuri, projektin hallinta, tietovarasto, koulutus, ETL-prosessi

FIGURES

TABLES

ABBREVATIONS

BI	Business intelligence
DBA	Data mart bus architecture
DBMS	Database management system
DSS	Decision support system
DW	Data warehouse
EDW	Enterprise data warehouse architecture
EIS	Executive information system
ETL	Extract Transform Load
FED	Federated architecture
G2B	Government to business
G2C	Government to citizen
G2G	Government to government
HUB	Hub and spoke architecture
IDM	Independent data marts
MIS	Management information system
OLAP	Online analytical processing
OLTP	Online transaction processing

TABLE OF CONTENTS

1 INTRODUCTION

This study examines business intelligence and BI systems from various angles to find insights on multiple less known areas. Business intelligence in general is a relatively studied topic, but the system implementation, architecture and design is a less examined area. Additionally, this study will have a more specific view for the topic, which is to inspect relations and differences between BI in private and public sectors. For organizations, there is a vast amount available information that can provide help in BI system implementation projects. However, operating sector of the organization is a factor that is rarely mentioned in studies. An initial hypothesis of this study is that the sector of an organization is a factor that will affect to the implementation and design of BI systems. This study has a public sector case organization that is currently acquiring and implementing a new BI system. The organization has multiple issues and questions related to the implementation process and it provides an excellent environment for testing and demonstrating of the knowledge and results of this study in a real business environment.

The scope of this study turned out to be ambitious and large. There are multiple goals and objectives, that the study aims to achieve. Firstly, it aims to build a comprehensive and theoretical knowledge base. Secondly, since this study follows guidelines of design science research methodology, it will aim to solve a real business problem, which in this case is to help an organization in various issues related to their BI system implementation project. Lastly, few specific research questions are answered. The research questions are following.

- What are the most important design choices and considerations of BI system implementation projects?
- How to approach these design choices and considerations?
- Does a generalized framework that has been built for common BI projects work for a public sector organization?

The objectives and the research questions were first approached by gathering a comprehensive knowledge base of prior studies. Theory building process was done carefully and by following scientific methods, and mostly the extant papers of the research field were selected as references. Because of the large scope of the study, a wide range of topics had to be included. Theoretical information was gathered about business intelligence, BI systems, architecture, design, implementation, users, software, and technology.

Most of the information for the theoretical knowledge base was searched by using various different keywords over Google scholar, but additionally IEEE Xplore -search engine and MIS quarterly journal were utilized in the searching process. Initially the theory exploration began with combining various keywords with the term "business intelligence". For example, common keywords were "business intelligence technologies", "business intelligence implementation", "business intelligence system" et cetera. After the big picture information about BI and BI systems was distinct enough, more detailed search terms were used. For example, when the architecture and common technologies of BI systems were known, more specific knowledge about them was searched. Said technologies and parts of BI systems were used in the searching process by combining them with keywords like "implementation" or "success factors". The more specific keywords were for example, "ETL implementation" or "Data warehouse success factors".

The primary factor used in the choosing of reference papers was the amount of references the studies had. To maintain the quality of the theoretical knowledge, this study attempted to refence only the most distinguished authors of the research field. A certain limit for minimum references was not used for the choosing process of studies, but papers with less than 50 references were considered carefully. One disadvantage of classic and renowned papers of one research area is that they are not usually very recent and therefore, there could be a lack of latest information. That is why some less referenced papers were used in this study to get more recent studies included.

A comprehensive knowledge base was required for development of a framework. It was done by utilizing a vast number of prior studies and design science research methodology process model (Peffers, Tuunanen, Rothenberger & Chatterjee, 2007). The framework was first designed and developed, and then demonstrated and evaluated by its utility in a real business environment. This study has a case organization, that started a BI system implementation project, and it was proven to be an exemplary subject for testing. This study and the framework were developed and evaluated at similar pace to the development of the case project, and therefore in-depth analysing was conducted over months. After the evaluation phase, the results and conclusions were presented and discussed.

The study proceeds in following fashion. Sections two, three, four and five are summarizing prior studies and knowledge about multiple topics. Section two examines business intelligence as a term and various features of it, such as benefits of its use. Section three attempts to find and explain the common parts and technologies of BI systems. Section four examines BI projects and general requirements and success factors for BI implementation. Fifth section will again examine BI system and its technologies, but from another angle than the third section. Now the focus is on the choices and considerations about implementation of different parts of BI systems.

The latter part of the study consists of four sections. Section six explains the research methodology of this study, and the case organization is introduced. The seventh section is the most essential part of this study. It includes the entire development and evaluation process of the framework. Lastly, in the eighth and ninth sections the conclusions, limitations and summary are presented and discussed.

2 BUSINESS INTELLIGENCE

As a term, business intelligence is a relatively recent one, but over the past decades, its role as an important information technology research area has been stabilized (Chen, Chiang & Storey, 2012). Despite the recentness of the term, the general idea of it has been around for a long time (Negash & Gray, 2008). The second section of the study will act as an introduction to business intelligence and it will go through the background, description, and benefits of it. Lastly, BI in the context of the public sector will be discussed.

2.1 Background

Business intelligence is a widely used and known term today, but the origin of it is from decades ago. In order to understand the current practices of BI, it is important to know the past of it (Watson, 2009). This subsection explains a few prior systems and ideas that lead to the emergence of BI.

In the field of knowledge management and decision support, there are few perceivable eras of systems that have led to the BI of today. Perhaps the most distinctive were management information systems (MIS) of the 1960's, decision support systems (DSS) of the 1970's and executive information systems (EIS) of the 1980's (Watson, Rainer, Koh, 1991).

The first systems, that distantly supported decision making, were already present in the 1960's (Watson, 2009). Back then, technology in general, was expensive and not very mature. This limited the implementation of systems that could support decisions efficiently. A need for something like that could execute similar tasks still existed. The implementation of such a system became practical in the 60's and it led to the initial emergence of decision support systems in the 1970's (Power, 2007). One of the first frameworks about decision support systems was presented in 1975 by Sprague and Watson (Watson, 2009). The framework states four criteria for DSS (Sprague & Watson, 1975);

1) The goal of the system is to support decision making

2) The system should be flexible enough to serve different levels and areas of an organization
3) The system should be dynamic in updates, in order to avoid ad-hoc revisitation.
4) The systems should be sophisticated. This means that it should utilize relevant and modern technology.

Turban, Sharda, and Delen (2014) consider DSS as an umbrella term, that can be used to describe any computer-based system that supports decision making. Power (2008) describes DSS more precisely as "an interactive computer-based system, or sub-system intended to help decision makers use communications technologies, data, documents, knowledge, and/or models to identify and solve problems, complete decision process tasks, and make decisions." The description is moderately broad, and it is very similar compared to the description of BI and other similar terms in the field. Additionally, the primary idea of BI is to support decision making, which can cause confusion between the terms BI and DSS. (Airinei & Homocianu, 2009).

In addition to DSS, executive information systems are part of the history of business intelligence. EIS emerged in the 1980's and it brought the computerized support for decision making to the top-level executives (Turban et al., 2014). Watson and others (1991) defined EIS as a computerized system that provides the organization's executives critical internal and external information to support decision making. Executive information systems provided increased visualization and measurements compared to DSS, and later on the concept of EIS transformed into what business intelligence is today (Turban et al., 2014).

2.2 Description

Business intelligence as a term was first introduced by Howard Dresner in 1989 (Power, 2008). Amongst the researchers, there is no clear consensus on the definition of business intelligence (Watson, 2009), hence there are various definitions of it. Because of the multiple definitions available, this subsection will go through a vast number of them and choose or form a description that suits the study in the most optimal manner without being biased.

Power (2008) defines the BI system as a data-driven DSS, whose main objective is to support the querying of an archived database and production of periodic summary reports. Power sees BI as a form of DSS and it is not an unusual view. At times, authors wonder if business intelligence should be considered as a completely new term for decision support systems or is it a replacement for data-driven decision support systems (Airinei & Homocianu, 2009). Slightly similarly to Power's vision, Chaudhuri sees business intelligence as a set of different DSS technologies that aims to enable knowledge to improve decision making (Chaudhuri, Dayal & Narasayya, 2011).

In an article by Negash and Gray (2008), business intelligence is defined in a different way compared to how Power described it. In the first place, Power saw BI as a type of DSS, but Negash treats it as a replacement for MIS, DSS, and EIS. Secondly, Negash provides more a comprehensive description: "BI systems combine data gathering, data storage, and knowledge management with analytical tools to present complex internal and competitive information to planners and decision makers (Negash & Gray, 2008)." The authors also emphasize that BI systems are meant to offer information at the correct time, location, and at correct form. They also recognize that business intelligence could refer to online decision making, which usually means keeping of the reaction time of decision making low as possible.

Wixon and Watson (2010) also state that BI does not have a universal definition. Their definition is following "Business intelligence is a broad category of technologies, applications, and processes for gathering, storing, accessing, and analysing data to help its users make better decisions." This definition is very similar and also almost identical to the previous definition by Negash and Gray. Wixon and Watson (2008) mention that BI has changed from using only historical data to utilizing real-time data to support operational decisions. Although the emphasis on the data used by the BI systems would be on recent or real-time data, the use of business intelligence is proactive (Negash & Gray, 2008). Historical data can be utilized for making forecasts about organizations' performance. Another change in BI is that it used to be something that only a few special people of an organization used. Making BI available to organizations' different user groups such as customers, suppliers, and operational personnel has been a recent change (Wixon & Gray, 2010).

Lönnqvist and Pirttimäki (2006) view BI similarly compared to the previous authors, but they have few very clear and understandable definitions, that can be used to form a definition of BI that will be used further in this study. Firstly, business intelligence is introduced as a tool or even as a managerial philosophy, which is used to help decision making (Lönnqvist & Pirttimäki, 2006). Secondly, the authors offer two options that BI can refer to. The first one refers to BI as information or knowledge about the organization and its customers, market, environment, etc. Using the term BI almost as a layman term "intelligence about business" is uncommon in the context of BI. The second way to refer BI is as a process of acquiring, analysing, and disseminating information from various sources to improve decision making (Lönnqvist & Pirttimäki, 2006).

More definitions of BI could be examined in this study, but the marginal utility in doing so is decreasing. Lastly, about the definition, the term business intelligence and analytics (BI&A), will be introduced. BI&A is not as commonly used as BI and essentially, they are the same. The term emphasizes the analytical part of regular BI and it is related to the field of big data (Chen et al., 2012).

As can be perceived, even though there are many different definitions of BI, they are very similar in content. The basis of the term BI doesn't change much in the definitions, but many authors have included their own details to their versions. Generally, the authors agree on the key features of business intelligence. In summary, business intelligence is seen as a tool or a system, which is used to improve decision making. The activities of BI are broadly gathering data from the source, refining, combining, processing and storing of data, analysing of data, and finally producing reports from the analysed data. The net value from the activities is the earlier mentioned improved decision making.

2.3 Benefits

Why organizations need business intelligence? Based on the definition of BI, its main net benefit is to improve decision making, but it is not the only benefit. This subsection focuses on the benefits of BI and what can be achieved by using it.

Herschel and Jones (2005) state that BI systems are critical to the daily operations of organizations and the importance of BI is only increasing. Especially in the fast-paced modern world, there is a need to shorten the timeframe between data acquisition and decision making (Chaudhuri et al., 2011). It is most likely granted that BI can be useful to an organization, but the mandatory of it can be questioned. Lönnqivst and Pirttimäki (2006) go as far as stating that organizations do not need BI only for achieving success but to survive.

Benefits provided by business intelligence are mostly intangible (Negash & Gray, 2008). This is unfortunate for measuring of BI success because tangible benefits are easier to measure (Wixon & Watson, 2010). Intangible benefits are likely to affect across the organization. An example of an improvement in the organizational level could be increased productivity or revenue (Trieu, 2017). Achieving large scale intangible benefits can be costly, but it can lead to significant transformational impact (Wixon & Watson, 2010). Tangible benefits can be measured more accurately, and they are easier to anticipate than intangible benefits. While intangible benefits are happening on an organizational level, tangible benefits impact on a local level. An example of a tangible and local level benefit could be finding information about customer behaviour or the most profitable customers (Ranjan, 2009). Figure 1 portrays the vision of Wixon and Watson (2010) of BI benefits.

As the figure demonstrates, BI benefits vary by their level of impact and how easy they are to measure. Cost savings is the easiest benefit to measure and as a tangible benefit, it impacts locally. Wixon and Watson (2010) explain that cost savings aren’t unambiguously only impacting on the local level. The level of impact depends on the extent of BI used in the organization. For example, implementing only a few BI applications does have lower cost and more local impact in the organization than large-scale BI transformation and infrastructure implementation.

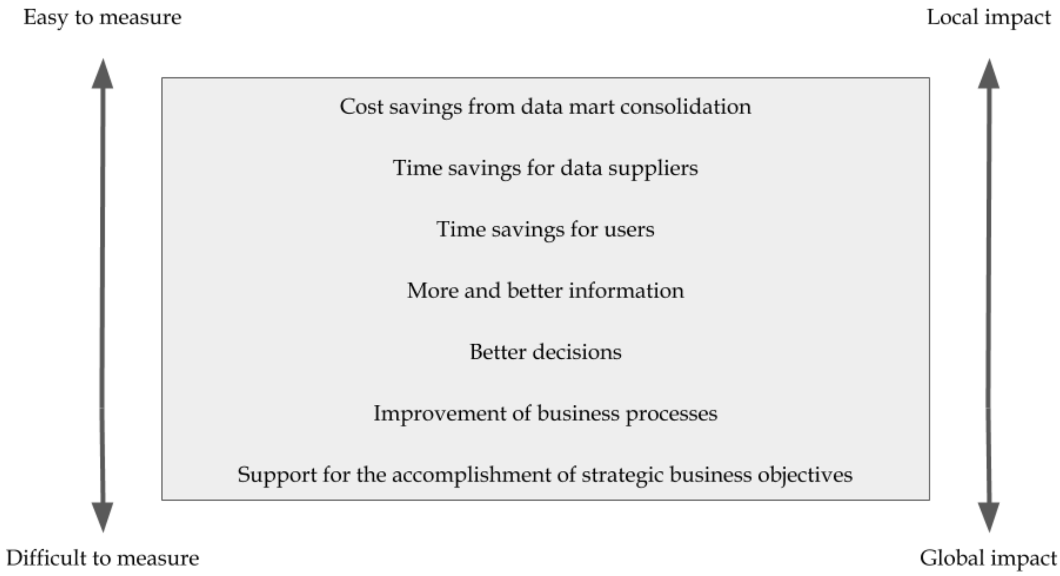

FIGURE 1 Business intelligence benefits (Wixon & Watson, 2010)

Time savings for both users and data suppliers were listed second in the figure 1 if it is read from the most tangible to the most intangible benefit. As an example, Ranjan (2009) lists various practical benefits of using BI. The ones that are directly related to time savings were the following; rapid detection of problems to reduce their impact, increased response rate from various methods of communication or campaigns, and lastly, reducing organizations' downtime by using predictive tools for timing maintenances.

In order to utilize business intelligence to its full potential, a certain chain of qualities should be fulfilled. Better information and decisions were listed in the middle of the figure by Wixon and Watson (2010). Watson (2009) states that the usefulness of BI is affected if the data used by the system is lacking in quality. As many of the earlier addressed definitions of BI mention, the main objective of BI is to produce quality information, which enables better decisions. Watson (2009) considers that BI itself is not a solution for quality data, but using BI encourages to focus on the data quality at its source. When the source data has acceptable quality, it enables quality information and better decisions (Watson, 2009). Additionally, the better quality of information is considered the most important benefit among the top Finnish companies in study made by Hannula and Pirttimäki (2003).

The most global benefits of the figure were improvement of business processes and support for strategic business objectives. Wixon and Watson (2010) see that these benefits have an organizational impact. Trieu (2017) views BI impacts and benefits in a different manner, but they do not exclude each other. According to Trieu, certain BI impacts are necessary conditions for improved organizational performance. For example, organizational performance can be improved by transforming business processes, improving intelligence quality, and by enhancing the products and services (Trieu, 2017). As we can see, these BI impacts are closely connected to the BI benefits of Wixon and Watson (2010). In order to fulfil a certain BI impact, for example, improvement of business processes, it is required to satisfy necessary BI assets. Assets can be for example, agile BI infrastructure, high level of skills, and effective BI applications (Trieu, 2017).

2.4 BI in public sector

The case organization of this study is a public sector organization located in Finland. Because of this, some business intelligence matters explicit to this context should be covered. This subsection will go through public sector specific factors related to the use of BI. A comprehensive description of the organization will be found in section 6.

Besides the private sector, business intelligence is also used in the public sector, but it usually is not exploited as efficiently. BI can certainly generate value and benefits for a public sector organization, but it should be noted that the environment for the BI system is different in the two sectors. BI outside the private sector has not been a very researched area, which is one reason for the earlier mentioned lack of exploitation (Boselli, Cesarini & Mezzanzanica, 2011).

2.4.1 Public sector vs private sector

The essential use of business intelligence is very similar no matter the sector, and with a successful implementation, it can deliver value in various types of organizations (Bodislav, 2015). This does not mean that the private and public sectors should be treated similarly. The root of the differences between the two sectors lies in their primary goals. While the private sector is driven by profits, the primary driver for a public sector organization usually is to provide quality services to the citizens (Boselli et al., 2011). Although maximizing profit is not the main driver for the public sector, the costs are not something to put aside. Requirements for the operation of a public sector organization are considered in various papers, and the importance of avoiding costs is almost always mentioned (Bodislav, 2015; Boselli et al., 2011; Coman, 2009; Nutt, 2005). Additionally, while operating with few resources and being cost-efficient, a public sector organization can still be required to improve and save money simultaneously (Boselli et al., 2011). Doing so is certainly not effortless or simple, but if we take a look at the BI benefits addressed in the previous subsection, the ambitious requirements for public sector organizations might be feasible. The main benefits of using business intelligence were in fact, cost savings and improved decision making.

With a more detailed look into the characteristics of the private and public sectors, additional differences can be found. Nutt (2005) found differences between the sectors for example, in their limitations, jurisdictions, and influences. The differences are listed in table 1, from where it can be perceived, many of the differences between the sectors arise from the environment of an organization.

TABLE 1 BI sectorial differences (Nutt, 2005)

Factors	Private organizations	Public organizations
Environmental market	Defined by the market	Defined by oversight bodies
Cooperation vs competition	Competition among similar organizations	Collaboration among similar organizations
Data availability	Performance and intelligence data available	Performance and intelligence data limited
Constraints	Autonomy and flexibility limited only by law and the need for internal consensus	Autonomy and flexibility limited by mandates and obligations
Political influence	Indirect and internal	Stems from authority network and from users
Transactional scrutiny	Can sequester the development of ideas	Cannot sequester the development of ideas
Ownership	Ownership vested in stockholders whose interests are interpreted using financial indicators	Citizens act as owners and impose their expectations about organization's activities and the conduct of these activities
Organizational process goals	Clear and agreed upon; efficiency dominant concern	Shifting, complex. Conflict-ridden and difficult to specify; equity dominant concern
Authority limits	Power vested in authority figures who have the authority to search	Stakeholders beyond the authority leaders' control influence the search for ideas

Multiple outside forces that can affect organizations depending on the sector, for example, economic trends, markets, or supervising bodies of government. Additionally, the stakeholders of organizations have influence on many factors. Public sector organizations generally have more individuals taking a part to the decision making, and in some situations, the citizens are involved as well. Partially because of this, the public sector has an increased need for consensus. Cooperation is also a factor that occurs more in the public sector because the organizations there are not completely driven by the market. Public sector organizations are not allowed to compete for customers, and they share more information amongst each other more than organizations in the private sector (Nutt, 2005).

2.4.2 E-Governance

The field of e-governance (Electronic Government or Governance) focuses on ICT capabilities of government or certain public sector organizations to improve the functioning of the organization at a broad level (Coman, 2009). According to Coman, E-governance and business intelligence are closely related terms and can be used in a cooperative manner. One reason why e-governance as a concept is presented in this study is that they have multiple common goals to improve public organizations, but the main reason is the major components or environments of e-governance. The large-scale goals of e-governance are to improve efficiency, services, and democratic processes of a government (Grönlund & Horan, 2005). Some more precise goals are, for example, providing information responsibly, being cost-efficient, and making government more accessible (Coman, 2009). Additionally, the case organization practices various forms of e-governance.

There seems to be a consensus about the different components or types of e-governance, but the terminology varies a lot. Backus (2001) refers to the components as target-groups, while terms such as solution, type, component, and environment are also being used amongst different researchers. The three components of e-governance are government to citizen (G2C), government to business (G2B), and government to government (G2G) (Backus, 2001; Coman, 2009; Palvia & Sharma, 2007). Additionally, there are few recognized but minor components, such as government to constituents (Palvia & Sharma, 2007) and citizen to government (Coman, 2009). The components are relevant to this study because they can be applied to business intelligence as well (Coman, 2009).

The major components can be sorted into external and internal solutions (Backus, 2001). Government to citizen is an external component of e-government, and it means the interaction between government and citizen on the web (Coman, 2009). G2C could be considered as one stop service provided to the citizen. The information related to this component varies from general contact information about the government to different laws and regulations (Palvia & Sharma, 2007). The second major component of e-governance is government to business, which is also considered to be external. In principle, G2B is very similar compared to the G2C, but it contains more of two-way interaction (Palvia & Sharma, 2007). The most important component of e-governance from business intelligence's point of view is government to government. G2G could be considered as an internal component, and its goal is to improve governments'' efficiency and create new partnerships (Coman, 2009). G2G uses both local area network and online databases (Palvia & Sharma, 2007). According to Coman, BI is best suited for G2C and G2G applications, because especially in those, the various technologies of BI are useful.

3 BI SYSTEM TECHNOLOGIES

As the background, description, and benefits of business intelligence have been inspected, it is time to move on to a more practical side of BI. On the one hand, this section views BI from afar to see the structure or architecture of a business intelligence system. On the other hand, a closer look at the different technologies and to their ways of working will be taken.

The definition of business intelligence was summarized in the previous section, it was concluded as a set of activities and technologies, that aims to improve decision making. In roughly chronological order, the different activities were collecting data from the source, cleaning, processing and storing of the data, and lastly analysing and acquiring knowledge from the information. BI is not just one software or simple solution that does all this, but a model-based approach to process a large amount of business data (Herschel & Jones, 2005). The next subsection aims to find the components or parts of generic BI systems.

3.1 BI system architecture

The definition of business intelligence, the concept of the BI system, and the contents of it varies in different papers, but the main idea remains relatively alike. Negash and Gray (2008) arguments, that the activities of BI systems vary depending on the structure of the data, but the architecture should be designed in a way that the system can process both structured and semi-structured data. Data that will be stored in the databases can come from different sources, and it must be processed before it can be utilized as information (Wu, Barash, Bartolini, 2007). The back-end of BI systems is commonly the point of the focus (Wu et al., 2007), and the architecture is typically centred around the data warehouse of the system (Negash & Gray, 2008). Even if the emphasis in the operation of a BI system would be on the back-end technologies, it does not reduce the importance of the decision-making activities operated on the front-end.

One way to inspect and understand BI systems would be to view the systems in a big picture and divide it into parts. Chaudhuri and others (2011) separate a common BI system into five parts: (a) data sources, (b) data movement and streaming engines, (c) data warehouse servers, (d) mid-tier servers, and (f) front-end applications. The generic BI system architecture is presented in the following graphic (figure 2).

Data sources are usually operational databases of organizations that feed the system with data (Chaudhuri et al., 2011) In a BI system, there can be multiple data sources that are very different from each other. In order to populate the system with data, it must be processed and moved into the data warehouse. ETL-process is responsible for doing so. The process consists of three steps, which are data extraction, transformation, and loading (ONG, 999). When the data has been cleaned and loaded into the DW, various operations and data mining can be performed. Additionally, online analytic processing (OLAP) servers in data warehouses allow operations such as data aggregation, filtering, pivoting et cetera (Chaudhuri et al., 2011). The last part of the generic BI system is the front-end applications, which allow users to present data in various formats such as spreadsheets and other visualizations.

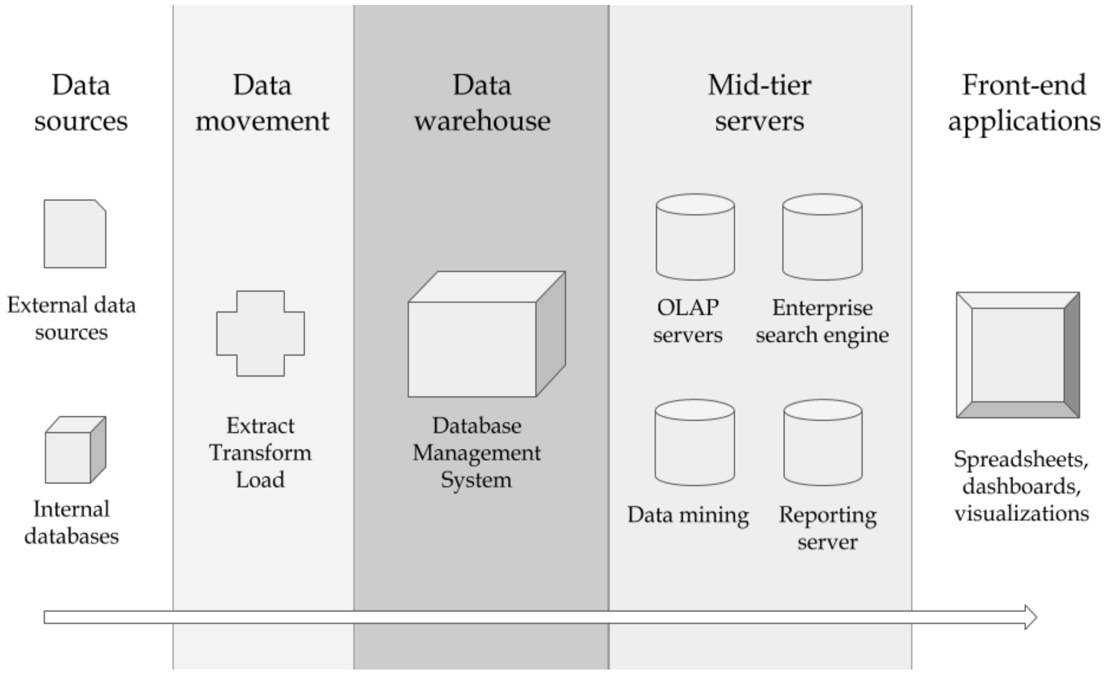

FIGURE 2 Generic BI system architecture (Chaudhuri et al., 2011)

A business intelligence system converts data from the sources into information and knowledge, by using both automated and manual analysis (Negash & Gray, 2008). Whether the work is done by hand or by an automated algorithm, the back-end heavy structure of the BI system is clearly visible in figure 2. Wu and others (2007) argue, that the emphasis of both ends in BI system architectures has lately become more balanced, but in a common BI system, the back-end is heavier.

The architecture of a common BI system could be constructed without much trouble since many authors had aligned views of the subject. Figure 3 was created to summarize the views of various authors. It contains an outlined generalization of how five different papers saw the structure of a generic BI system.

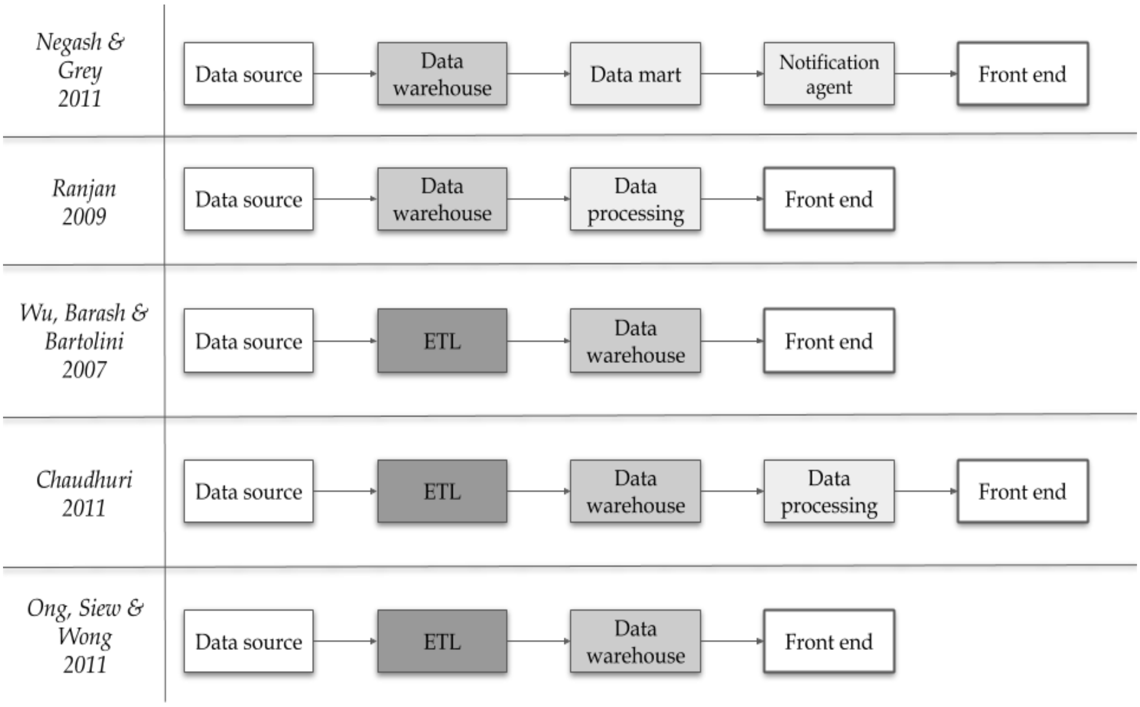

FIGURE 3 Simplified summary of different views on BI architecture

The previous figure was constructed by simplifying similar illustrations from the original papers and some shortcuts were made. Hence, it must be noted that they are not the perfect description of the authors' view. Additionally, in many cases, the differences between figures were a matter of semantics. By utilizing the figure, subjects for the next five subsections could be assigned. The rest of this section will provide a profound description of the different components of BI system architecture.

3.2 Data sources

All BI systems require source data to work, and the origin of it can vary a lot. According to Yeoh and Koronios (2010), source data is an important part of BI systems operation, because the quality of it can affect the reports and decision making that are derived from that data. They also emphasize, that business data can't be utilized to the full extent if the quality and integrity of source data are lacking. Choosing the right sources and ensuring the quality of the data are key to developing a scalable and highly functional BI system (Moss & Atre, 2003).

There are no clear rules for what type of data BI systems have to use. Ranjan (2009) explains, that the source data can exist on various platforms and the structure of it can vary from spreadsheets to images or text files. The source of the data can be for example internal database or external data exported from the Internet. Also, the structure of the source data can vary. Negash and Gray (2008) separate BI source data structures into unstructured, semi-structured, and structured data. Unstructured data is seldom used, and BI systems mostly deal with only structured data. Still, both structured and semi-structured data are needed and useful for BI systems (Negash & Gray, 2008; Rudin and Cressy, 2003). The reason for this is because it has been estimated that 85% of all business data are semi-structured (Blumberg & Atre, 2003), and leaving them out would decrease the available source data drastically. Additionally, Negash and Gray (2008) carefully claim, that semi-structured data is not necessarily inferior to structured data.

As mentioned earlier, the lacking quality of data can have an effect on the eventual decision making. This is why one of the BI systems' goals is to decrease the gap between the quantity and quality of the source data (Popovič, Coelho & Jaklič, 2009). A high amount of data is not very useful if it lacks quality and vice versa. According to Popovič and others (2009), this type of gap can result from many factors, for example, too rigorous management of data, organizations' use of invalid or unreliable data sources, lack of external data, or poor allocation of resources. The requirements for BI systems are becoming increasingly challenging over time (Dayal et al., 2009). For example, expectations for the latency of decision making are lowering. These kinds of challenges are concerning data sources as well, and they encourage investing in their improvement.

3.3 Data extraction, transforming and loading

In order to utilize source data of a BI system, it has to be moved from the source to the data warehouse. As the previous subsection stated, the source data can often be unstructured and unclean. Because of this, BI systems require tools for data extraction, cleansing, integrating, transforming, and loading (Dayal et al., 2009). ETL tools are the answer to this requirement, and they are something that exists in almost every BI system. Technical details of how the tools work are beyond the scope of this study, but their purpose and general functioning will be explained.

The first step of ETL is the extracting of data. Like every other part of ETL, extracting requires a lot of planning and researching before the implementation of it. The most important questions about the source data extraction are: "What kind of data we need?", "What data is available to us?", and "What form or structure the source data has?" (Jun et al., 2009). Extracting is the first phase of the normal ETL process and after the data has been collected, the transforming phase begins.

The data transform part includes cleansing and conversion of data, and the goal of it is to get the source data to match the target format. This can be done by completing various activities, for example, removing unwanted data, such as incomplete or duplicate data, processing of character strings, and sorting and grouping of data (Bansal & Kagemann, 2015; Jun et al., 2009). After the data has been processed, it can be loaded to data warehouses by updating the new data into the database table (Jun et al., 2009). This was the last phase of the ETL process.

3.4 Data warehouse

A data warehouse is an important component in almost every BI system (Ranjan, 2009). In the common BI architecture, the data warehouse locates in the very middle of everything; between the data source and front-end applications. Inmon (1992) defines a data warehouse as a database, which provides useful and consistent information for the organization to utilize. He uses the words subject-oriented, integrated, and time-variant, to describe data warehouses.

Data warehouses are usually located separately from the organizations' internal systems, from where the source data is extracted (Gray & Watson, 1998). As was mentioned earlier, the data sources of the BI system can be varied in many ways, for example internal, external or relational databases and the format of the data can vary a lot (Ranjan, 2009). Also, data warehouses can contain both aggregated and pure transactional data. This means that it can have new data, historical data, summarized data, and meta data.

Even though data warehouses are normally implemented using traditional database management systems (DBMS), they have a few notable differences compared to regular operational databases (Moody & Kortink, 2000). The first difference is how end users can access the database. With data warehouses, the users normally write queries to perform searches from the database, while operational databases are often accessed through front end applications. The second difference is that data warehouses are usually read-only. This is because the extract process of ETL tools is generally the only method to move data to the warehouse (Moody & Kortink, 2000).

3.4.1 Data warehouse architecture

The centre of BI systems, the data warehouse, has an architecture of its own. When different authors discuss data warehouses, they tend to use a generic data warehouse architecture in their text, but alternative architectures do not receive as much consideration. Ariyachandra and Watson (2006;2008;2010) makes an exception to this trend when they represent five major options for DW architecture. The differences of the architectures are described and illustrated (Figure 4) in this subsection. The different architectures are the following:

1. Independent data marts (IDM)
2. Data mart bus architecture with linked dimensional data marts (DBA)
3. Enterprise data warehouse architecture (EDW)
4. Hub and spoke architecture (HUB)
5. Federated architecture (FED)

Independent data marts were the earliest iteration of data warehouses. This architecture has a siloed structure, which means that the individual data marts were not connected together as they are for example in the DBA, which will be presented later (Ariyachandra & Watson, 2010). IDM architecture is only intended to support local needs and specific tasks, and the data is store in a model that works best for the data type (Ariyachandra & Watson, 2008). Additionally, IDM did not have consistent definitions of data or ability to analyse data in the organizational level. (Ariyachandra & Watson, 2010). Independent data marts are the simplest and least costly way of data warehousing (Turban et al., 2014).

In the *data mart bus architecture with linked data marts approach*, the architecture consists of multiple data marts, that are designed for specific business processes. This is why the first step of the DBA approach is to identify business processes and technical requirements for the data marts (Ariyachandra & Watson, 2010). Unlike in the siloed IDM, in this approach, the data marts are connected through common dimensions. Because of this, DBA has a wider organizational impact than IDM, but it depends on the number of connected data marts (Ariyachandra & Watson, 2010). The DBA architecture is like an evolved version of the previously represented IDM. Turban and others (2014) even describe the DBA architecture as "a viable alternative to the IDM".

In the typical *enterprise data warehouse architecture*, there usually is a large and centralized data warehouse (Sen & Sinha, 2005). The centralized DW contains multiple data types and models, and it is an enterprise-wide repository (Ariyachandra & Watson, 2010). The key benefits and differences in EDW are its scale and the organizational impact it can provide. When it comes to these factors, the EDW architecture is pictured as a clear winner, but Ariyachandra and Watson (2010) argues that a large DBA can provide very similar benefits compared to EDW. A typical EDW architecture does not utilize any data marts, but if departmental and dependent data marts are used in supportive or enhancing manner, the architecture is no longer considered as the EDW, but as the *hub and spoke architecture* (Turban et al., 2014), which is also referred as centralized data warehouse with dependent data marts. The strength of hub and spoke architecture is when both a large data warehouse and a strong connection between different departmental units are needed (Ariyachandra & Watson, 2010).

Federated architecture differs from the previous approaches to warehouse architectures, because FED allows the utilization of previous systems unlike the other approaches (Turban et al., 2014). This architecture aims to integrate existing data marts and systems either by using common elements, for example, metadata or dimensions or by queries (Ariyachandra & Watson, 2010). FED provides a practical solution for mature organizations when the integration of existing and new data marts is required, and it is considered as a realistic and suboptimal solution (Turban et al., 2014).

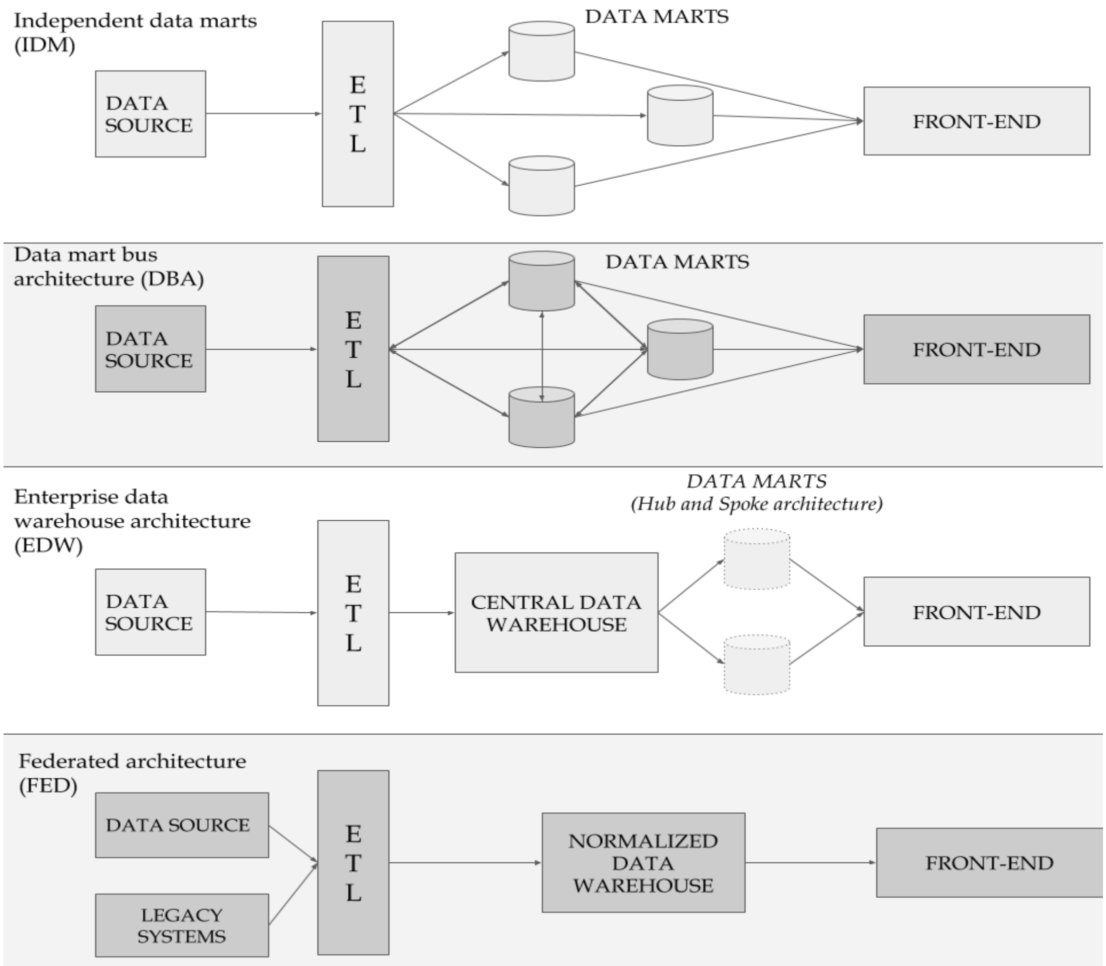

FIGURE 4 Data warehouse architectures (Ariyachandra & Watson, 2006; Sen & Sinha, 2005)

As can be perceived, the main processing between the ETL-layer and the front-end happens in either data marts, data warehouses, or even the both can be utilized in one solution in a similar fashion to the hub and spoke architecture.

3.4.2 Online analytical processing

Data warehouse could be described as a collection of different types of data, which can be utilized in decision making processes with the help of online analytical processing (OLAP) techniques (Hüsemann, Lechtenbörger & Vossen, 2000). In warehouses, data is usually stored multidimensionally, and they are built to support OLAP tools. The support for multidimensional data models does require a specific way of implementation, that is not usually provided by database management systems that are targeted at operational databases (Chaudhuri & Dayal, 1997).

In general, online analytic processing techniques are queries that are used to analyse and aggregate data to discover important factors and trends (Ranjan, 2009). They allow performing different activities to data, such as filtering, drill-down, aggregation, and pivoting (Chaudhuri & Dayal, 1997). The idea of OLAP technology is mainly about navigating through a vast amount of data and it lacks rigorous mathematical algorithms. In comparison, operational databases usually use online transaction processing (OLTP), which has more a rigorous mathematical framework (Lenz & Thalheim, 2009; Schewe & Thalheim, 1993). OLTP applications usually do repetitive and constant tasks to automate transactional data (Chaudhuri & Dayal, 1997).

3.4.3 Data marts

In some situations, a data mart could be an option to either support or replace a data warehouse (Moody & Kortink, 2000). Data marts are like low-cost and small-scale data warehouses, and they are targeted for a specific group of decision-making users or tasks. One reason for the implementation of data marts is to use the more affordable option. The econd one is to utilize both, by distributing data from the central warehouse to department-specific data marts to address sectional needs better (Gray & Watson, 1998; Inmon, 1999).

3.5 Front-end applications

"Business intelligence can be defined as the process of turning data into information and then into knowledge" (Ranjan, 2009). The last part of the phrase is done in the front-end applications. The front-end applications are software that the end users of the system operate. BI front-end tools are an easy way to access data warehouse without using any SQL, and a good way to understand the architecture and form of the BI data (Howson, 2007). The main objective of the front-end software is to improve decision making and to produce informative reports and to publish measures. Those objectives are the same that were defined as the goals of the entire BI system. While most of the work happens in the earlier layers of the BI system, the reaping of the fruits and the final work will happen in the front-end applications.

The main challenge with the front-end applications is how to efficiently represent the contents of a very large data base. Probably the most used method is to visualize data in spreadsheets (Ranjan, 2009) In addition to the spreadsheets, the front-end software utilizes other visualizations such as scorecards and performance measures to transform data into information (Howson, 2007). Besides representing, BI software is also used for analyzing of the data. Visual analytics is a term that includes the combination of visualization of data and predictive analysis (Turban et al., 2014). This means that the front-end applications are not only answering the questions "what is happening?", but also "what will happen?" and "why?". In BI software, common operators and functionalities for analysis are the ability to rollup or drill-down in the different layers of the OLAP cube, and to slice and dice through different dimensions (Ranjan, 2009). The analytics in BI software can be done in various ways. Turban and others (2014) divide the utilization of BI front-end applications into three categories, which are reporting, predictive, and prescriptive analytics (Figure 5).

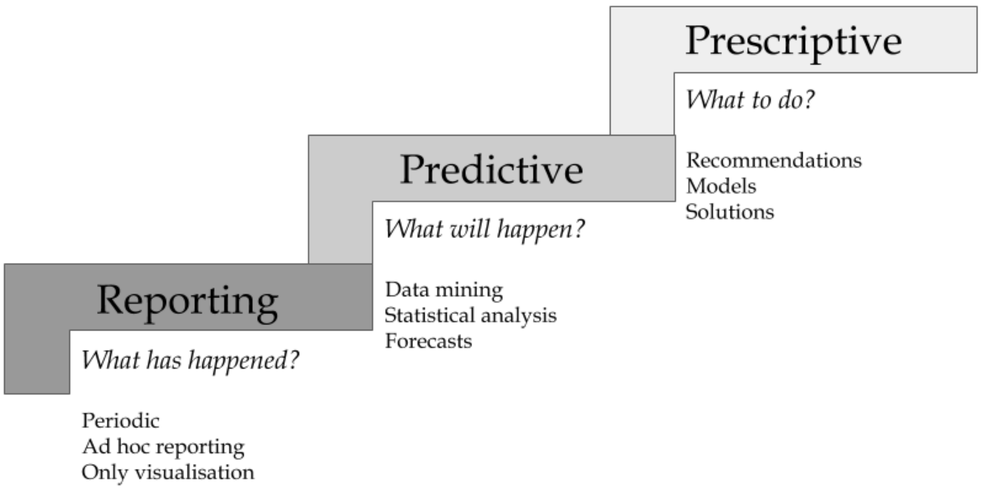

FIGURE 5 Different analytical methods of BI applications (Turban et al., 2014)

Using front-end applications for reporting only is an ad-hoc, low latency and usually an automatic way of making visualizations from the data. In contrast, predictive analytics requires more manual work. In predictive analytics, more profound trend analysis and various techniques are used to predict future development (Turban et al., 2014). Prescriptive analytics is the next level for predictive analytics. Now the goal is not only to understand what has happened or what will happen but in addition, the objective is to guide and give recommendations for what to do in the future.

4 BI IMPLEMENTATION PROJECT

The implementation and design of a business intelligence system are not trivial tasks. In this context, the implementation process is not just acquiring software and hardware, and it requires a lot of planning on how different parts of the system fit together as a whole (Yeoh & Koronios, 2010). This section aims to give a comprehensive view of BI system implementation and the topics remain in designing, defining, and preparing related activities. The section after this will go into more detail with the BI implementation in the context of different BI technologies.

4.1 Characteristics and activities

The goal of a BI implementation project is to develop a system that can provide business departments useful and informative applications for decision making (Reinschmidt & Francoise, 2000). BI projects may seem similar to other IT projects, but there are some differences. The first difference is that in BI projects, the presence of the business department is more important than in common IT projects. For example, selecting the right sources for data and developing informative knowledge from it requires a lot of business knowledge alongside IT skills (Reinschmidt & Francoise, 2000). Another difference is that in BI projects, the requirements can be harder to define precisely. When implementing a business intelligence system, accurate requirements can usually be defined only after working with the systems for a while (Inmon, 1992). Because of this, the implementation and design of BI systems should be done in an iterative manner to maintain flexibility in change management (Reinschmidt & Francoise, 2000).

BI implementation project is a large process and it consists of various activities. According to Chaudhuri and Dayal (1997), BI projects have multiple common activities, that start with defining the system in a big picture. This includes system architecture and database and storage-related defining. The next activity is to integrative the internal and external storage and the servers. When source storages are set up, data warehouse and its technologies including ETL should be designed and planned. The next activity is to data flow pipelines, which means data sources, ETL, and data warehouse. The remaining activities are designing and implementation of front-end applications (Chaudhuri & Dayal, 1997). It should be noted, that the implementation process is not as straightforward in practice, and the emphasis is on planning and designing instead of actual “doing”. The activities do not follow a chronological and simple path as earlier mentioned and the BI project should remain iterative and flexible, instead of fully completing each building block individually.

4.2 Design models

If the generic BI architecture (Figure 2) presented in the third section is viewed vertically, the data sources would be at the bottom, and the front-end applications would be at the top. Such a setting is a good way to understand the main differences between design models. There are two common approaches in the traditional BI and data warehouse implementation process, and the logic for them can be understood by taking the vertical perspective of the architecture.

The first one of the two design models is often referred to as top-down strategy (Bodislav, 2015), the Inmon model, or the enterprise data warehouse (EDW) approach (Turban et al., 2014). The first term will be used in this paper because of it’s the most descriptive. In the top-down strategy, the data warehouse is implemented first and later the individual data marts are implemented for specific business processes. The implementation and design will proceed from “top” to “bottom”, which in this case means from the data warehouse to data sources. The strength of top-down strategy is that it will first prioritize the organizational needs and later the local needs of different departments.

The second approach to BI system implementation is the bottom-up strategy (Bodislav, 2015), also referred to as the Kimball model or the data mart approach (Turban et al., 2014). In this approach, the designing and implementation will proceed from “bottom” to “up”, so the data sources and departmental factors are considered first before. The reasoning for the bottom-up strategy is that it provides more focus on the specific departments and it doesn’t require building a large central data warehouse, unlike top-down strategy does (Turban et al., 2014). A more detailed view of the differences in the design models is presented in the following vision (Table 2)

TABLE 2 Design model characteristics (Turban et al., 2014)

	Inmon	**Kimball**
Overall approach	Top-down	Bottom-up
Architectural structure	Enterprise data warehouse is the centre of the system	Data marts for single business processes, enterprise consistency achieved through data bus
Complexity of the method	Complex	Simple
Discussion of physical design	Thorough	Limited
Data orientation	Subject- or data-drive	Process oriented
Data modeling	Traditional, such as entity relationship diagram	Dimemensional modeling
End-user accessibility	Low	High
Primary audience	IT professional	End users
Place in the organization	Technology oriented solution that provides an enterprise wiede single version of the truth	Solution that allows access to operational data for end-users

The two approaches differ almost in every area of data storage and processing (Breslin, 2004). Also, one of the primary differences is the implementation style. The top-down or Inmon model, is more focused on fulfilling technical specifications and the execution of the strategy generally requires IT-skills (Turban et al., 2014) The implementation style of the bottom-up or Kimball model, requires fewer IT-skills and it is more accessible to the users (Breslin, 2004). This also means that the bottom-up model is less complex, which can be a benefit or disadvantage depending on the situation.

4.3 Resources

Implementation of a business intelligence system is not only time consuming, but it also requires capital and a broad set of skills. The required resources that are examined in this subsection are the costs and skills. It is acknowledged that for example, time could be considered an important resource. However, that aspect of the implementation will not be as thoroughly analysed. The estimated implementation times of a BI system can vary between different projects a lot. One estimate is that the implementation time of only a data warehouse can vary from months to years (Azvine, Cui & Nauck, 2005).

4.3.1 Costs

Negash and Gray (2008) divide the costs of a BI system implementation into four categories, which are hardware, software, implementation, and personnel. The hardware costs can vary a lot between different organizations, because some of them could already have some of the required components, for example, an intranet.

The software costs depend on the pricing model, but Negash and Gray (2008) estimate that a common cost for BI software packages is approximately $60,000. However, annual software maintenance and possible additional software related costs are not included in this estimate. After the building blocks have been bought and acquired, the implementation costs can be considered. The implementation costs are mainly a one-time investment, but there still can be some recurring payments (Negash & Gray, 2008). After the implementation and training, the system is ready to use, but it must be noted that organizations, employees, and technology change. This is why the system requires maintenance and training after the initial implementation.

4.3.2 Skills

In terms of skills and knowledge, the use and implementation of BI can require a broad set of skills. Likely the most intelligible categorization of BI skills can be done by dividing them into two: IT- and business-related skills (Debortoli, Müller & vom Brocke, 2014). Multiple researchers consider aligning the technical implementation of BI systems and the business objectives of the organization is a critical success factor for a BI system (Arnot, 2005; Yeoh & Koronios, 2010). In their research, Debortoli and others (2014) provide a list of competencies that are required when working with business intelligence. Most of the required competencies are listed below.

- Sales and business development
- BI platforms (Microsoft, SAP and SAS)
- Digital marketing

- Database administration
- Software engineering
- BI architecture
- Project management
- Business analysis

The findings of the previous study suggest that business knowledge is considered as important as technical skills in the context of business intelligence (Debortoli et al., 2014). Whether the statement is true or false, it gives an idea of a wide spectrum of different skills are required to work successfully with BI. An additional way of viewing required BI competencies is to divide them into three categories: (1) analytical skills, (2) IT knowledge and skills, and (3) business communication skills (Chiang, Goes & Stohr, 2012). IT skills include management of databases and warehouses, knowledge of ETL methods, and source data management. IT skills are mainly used in the back-end of a BI system. IT skills are also needed in the front-end because the creation of visualizations and dashboards must be done. Analytical skills concerns techniques for the processing of the data. For example, data mining, statistical analysis, and econometrics are listed as analytical skills. These competencies are used to provide knowledge from the data that is managed with IT skills (Chiang et al., 2012). Business and communication skills are used to make decisions and find meaningful realizations from the information. A business-oriented person working with BI must understand the organization, accounting, financing, and marketing for example (Chiang et al., 2012).

To give a practical example of required BI skills, the topic can be inspected by real life examples. Reinschmidt and Francoise (2000) have made a directive list of job titles, that are needed in the BI development group. The titles are the following; technical project manager, BI solution architect, business subject area specialist, database administrator, platform specialist, tool specialist, and extract programmer.

4.4 Defining requirements

The main objective of BI implementation is to acquire the best possible system, which naturally varies in every organization. To implement a good BI system, the requirements for features and qualities should be defined and understood. In this subsection, the matter is approached by looking into different ways to define the desired qualities of BI systems.

Every organization has its own business and its own data. That is why each organization has its needs and requirements for BI systems. In their paper about BI design and implementation, Gangadharan and Swami (2004) present a set of questions that can help organizations decide which type of qualities and features are needed. Interpretations of the main ideas of the questions are listed below:

- What are the goals of organizations for using information?
- How the goals and objectives are prioritized?
- What are the user groups and how their requirements vary?
- Is it possible to use information as a strategic asset?
- What are the goals for the BI strategy?
- Who makes the decisions and how they are made?
- How the implementation of BI adds value to the organization?

These questions are useful for a corporate level requirement definition, but they do not provide means to define precise or technical requirements. One way to divide requirements is to divide them into two categories, which are micro and macro level requirements (Elena, 2011). Micro level describes needs for one specific part of the system or one department. Macro level requirements are concerning the needs and priorities of the whole organization or system. Requirements can be also divided by topics or themes. Ranjan (2009) divides BI requirements into categories, that are needed to be addressed in order to achieve an efficient BI system. The said requirement categories are security and user access, data volume, data storage, and performance-related requirements.

Even though planning and foreseeing are important, there are parts of the BI system that cannot be assigned requirements before they see some use. Inmon (1992) argues that for example, a data warehouse is a part of a BI system, that can't be treated as a requirements-driven system. Unlike with many other parts of BI systems, the data warehouse should at least be in partial use, so that the true requirements can be defined. Another important factor about the requirements by Inmon (1992), is that different level requirements fit better and easier to set for different BI technologies. This means for example that the macro level requirements can be used for data warehouse design because the data warehouse should be optimized on the organizational level. In comparison, the micro level requirements are better for data mart design, because data marts usually process departmental information (Inmon, 1992).

4.5 Critical success factors and challenges

Because the main objective of a project is to implement a successful BI system, it is helpful to look into critical success factors and challenges of the process. Critical success factors (CSF) are factors, features, and things that need to go well in order to achieve success (Boynton & Zmud, 1984). In other words, they are the key areas, that are most vital for the system or the organization. In the context of BI, the critical success factors have already been studied in several different papers, so the previous knowledge can and should be utilized in future projects and papers like this.

Understanding the critical success factors of the project is not useful only for avoiding failure but for knowing where to prioritize time and money, because it can help in budgeting and scheduling as well (Reinschmidt & Francoise, 2000). Every project is different, so the precise critical success factors for each system can vary. To understand the CSFs of BI projects at a general level, it is helpful to inspect the most common and often mentioned CSFs. In order to do so, research by Olszak and Ziemba (2012) was selected as a reference paper for the factors, because it collected CSFs from a wide variety of sources. The ten most common CSFs of the said research are listed below.

- Support from senior management
- Clear and realistic goals
- Strong plan and vision, that is updated
- Good communication in the organization
- Clients and users are involved
- The project team has a proper set of skills
- Effective change management
- Capable team or project leader and good leadership in general
- Working business plan
- Adequate resources

The previous CSFs seem to be very universal because they could fit for almost any project. In order to inspect CFSs in a more detailed and BI specific level, the original list requires factors that are more specific for the context. To complete and to refine the previous list, two additional studies were inspected (Arnott, 2008; Yeoh & Koronios, 2010). The additional papers had many common factors with the previous list, but they weren't identical. The most unique factors for BI systems were selected as supplements.

- Effective data management
- Good data quality and integrity
- Appropriate hardware and software
- The system requirements are accurately defined
- A balanced team between IT and business department
- Business-driven, scalable and flexible technical framework

Another way of analysing the means of success and how to avoid failure can be done by taking a look into possible ways of how the project can fail (Reinschmidt & Francoise, 2000). Reinschmidt and Francoise encourage to think about the measures of failure, which in principle, are the opposite of critical success factors. For example, the measures could be low performance, users being unhappy with the data quality or integrity, or that the users are not involved in the project or in the use of the system.

Additionally, to the CSFs, it is important to understand possible challenges and risks of the project. The previous measures are indicators of failing projects, and necessarily couldn't be considered as challenges or risks. For example, the challenges that BI implementation projects could face are related following issues (Gangadharan and Swami, 2004): meeting performance goals, creating a platform to support multiple applications, enforcing security, and working with limited resources.

5 IMPLEMENTATION CHOICES FOR BI SYSTEMS

The previous section examined the characteristics of BI implementation projects from a general and directive perspective. This section will take a look at the technical side of the project. Going into very specific technical details is beyond the scope of this study, and the matter is viewed in a big picture. This section is divided into three subsections. The first will briefly address data sources and then move onto ETL, the second one will examine the implementation of data warehouse and the last subsection is for front-end applications and end-users.

5.1 ETL implementation

As was mentioned earlier, data sources of organizations can vary from internal to external, and from unstructured to structured data. The data sources itself do not require much of an implementation, because they are most likely already existing in the organization. Negash and Gray (2008) explained, that the structure of the data can impact the important decisions concerning the BI systems. For example, different ETL methods are used for different structured data (Negash & Gray, 2008).

The implementation of ETL is a labour heavy process and it requires answers for various important decisions (Dayal et al., 2009). Jun and others (2009) argue that ETL does not often receive the attention it requires, and it might be because the task can seem more elementary than it actually is. To tackle the question about the implementation of ETL, this study will approach the matter by viewing various decisions or choices that are ought to be made. The choices that were selected are (1) existing tools vs in-house code (Amanpartap & Khaira, 2013; Jun et al., 2009), (2) ETL vs ELT vs ETLT (Dayal et al., 2009; Vassiliadis & Simitsis, 2008), and (3) options for data transmission scheduling (Dayal et al., 2009; Vassiliadis & Simitsis, 2008).

Before debating the implementation choices, it is helpful to understand what the desired qualities of ETL are. Dayal and others (2009) have provided a comprehensive list of such qualities and the said list is quoted and explained next.

- *Reliability and availability*: ETL should complete tasks that are planned for it in an appropriate timeframe. This should be an accurate and continuous process.

- *Maintainability and affordability*: Use of ETL should maintain at intended and affordable cost while functioning at a high level.

- *Freshness*: Data and information should move and update with low latency.

- *Recoverability*: The previous working version of the ETL system should be possible to be restored.

- *Flexibility, scalability, and robustness*: ETL should be able to maintain functionality even when encountering unusual conditions and without causing major damage. Such conditions could be for example anomalies in data or sudden requirement to process a high volume of data.

- *Consistency*: Integrity and consistency of data loaded into the data warehouse should remain at a high level.

- *Traceability*: Timeline and possible changes of the data should be able to be tracked.

- *Auditability*: Security and privacy of data should be protected.

5.1.1 Commercial ETL tools vs in-house code

ETL is normally implemented by producing in-house code or using existing ETL tools that are available on the market (Dayal et al., 2009). The implementation of ETL can be a difficult task and one of the most crucial questions concerning the implementation is whether to buy it or make it.

Jun and others (2009) compare the differences between using ETL tools and using home-grown code. The differences may seem self-evident, but it does not reduce the importance of the matter. The categories that were used in the comparison were flexibility, complexity, efficiency, development cycle, workload, and price. The authors came to the expected conclusion, that bought ETL tools require less work and they are quicker and easier to implement. Making your own ETL tool requires knowledge in multiple programming languages and the code itself can be considered as lengthy (Amanpartap & Khaira, 2013). Additionally, readymade ETL tools often have a graphic user interface, so that person without a high level of programming skills can operate them.

The downsides of bought tools were lower flexibility and higher price (Jun et al., 2009). However, manual coding from scratch can be a long procedure, but in general, it is considered as the cheaper and more flexible option, especially if the organization already possesses capable programmers (Jun et al., 2009). The advantage of self-coded tools is flexibility, which provides benefits, especially when setting up repositories and or when doing unit testing (Amanpartap & Khaira, 2013). The flipside of increased flexibility is that when everything is built from scratch, all changes and updates have to be done manually. The bought ETL tools conceivably have support and updates (Jun et al., 2009).

It should be noted that the subject is not as unambiguous as it seems. For example, choosing to develop the ETL tools in the organization can potentially be the cheaper option of the two, but without any possessing any existing expertise it could become the more expensive option. The second example of the complexity in this matter is that even though the commercial tools are considered less flexible in principle, they aren't necessarily inflexible (Jun et al., 2009), so this factor might not cause any problems for an organization.

5.1.2 ETL vs ELT vs ETLT

Traditional ETL works as earlier described: data extraction, data transforming, and finally loading data into the data warehouse. However, there are options for different ETL implementation styles, which are ELT and ETLT (Dayal et al., 2009). The difference between ETL and ELT is that transforming of data is done in a different time and place. In regular ETL, the data is transformed in the staging area, before it is loaded to the data warehouse. In ELT, the unclean data is loaded to the data warehouse, and the transforming is done there (Dayal et al., 2009). The main reason for using ELT instead of ETL is possible limitations in organizations' hardware or tools. If the engine of the data warehouse is much more powerful than the ones that are reserved for ETL, the transforming can be done more efficiently in the data warehouse (Vassiliadis & Simitsis, 2008). Additionally, ELT could be a better option when scalability is emphasized because in general, the data warehouse is more reasonable to scale up than the machinery and tools for ETL.

ETLT could be considered as a middle ground between ETL and ELT because the both styles are used in it (Vassiliadis & Simitsis, 2008). The principle of ETLT is that both ETL and ELT has their own benefits and ideal use cases, and that is why the transformations are divided into two groups (Dayal et al., 2009). One group is for the data that is ought to be transformed and loaded quickly or close to real-time because the data has to be usable fast. The use of the data from the other group is not so time-dependent, and that data can be transformed in a more definite manner (Vassiliadis & Simitsis, 2008).

5.1.3 Data transmission schedule

The data transmission from the source to ETL tools and eventually to the data warehouse can be scheduled differently in BI systems. The most common option for the transmission schedule is to use periodic data flow. This means that the data is extracted from the source in scheduled batches (Dayal et al., 2009). Two additional options for the scheduling are push- and pull-based scheduling policies (Vassiliadis & Simitsis, 2008). In the push design, the data moves to the warehouse as soon as it is ready, without a specific schedule or a request. The pull design is the opposite. When in push-based "scheduling" the data transmission happens at the earliest convenience, in pull-based design the data extraction and transmission happen when a request is given (Vassiliadis & Simitsis, 2008).

When an organization is choosing its transmission schedule, it should consider the strengths and weaknesses of each option. The periodic transmission could be considered as a safe choice because it is most commonly used and it is very straightforward (Dayal et al., 2009). However, if fresh and up to date data is highly valued in the organization, the speed of push design could be the optimal choice. When the freshness of the data is not the most critical feature, or if the source data updates irregularly, the flexibility of pull design could turn out useful (Vassiliadis & Simitsis, 2008).

5.2 Data warehouse implementation

The basic characteristics of data warehouses were examined in the third section and now it is time to take a look into different choices regarding the design of data warehouses. The implementation of a data warehouse is a large-scale project, and it is impossible for one person to master every aspect of it (Turban et al., 2014). Because of the extent of this subject, this study aims to find only the most important and relevant factors about it.

Chaudhuri and Dayal (2011) consider DW implementation as a very long and complex project that could take years to do right. In the big picture, the general implementation project proceeds more or less in the following manner (Chaudhuri & Dayal, 2011). First up is the planning and defining of the requirements, which is an important step that helps to choose the right architecture. When the basic design of the data warehouse has been selected, the servers, storage, and possibly the client-side tools can be integrated together. When the most important technologies are acquired, common practicalities should be decided, such as how to access the DW or where to store external data. Then it is time to connect the data sources and design and set up the ETL tools. When data flows are connected and working, the warehouse can be populated with data (Chaudhuri & Dayal, 2011).

To gain a general view of how to implement a data warehouse, the success factors and problems of it are examined. A literature review by Xu and Hwang (2007) lists the most common DW critical success factors from various known papers about the topic, and the CSFs seem to be very similar to business intelligence CSFs that were already listed in the previous section. For example, the first listed factors were well-defined business needs, top management support, and user involvement. While the CSFs of data warehouses did not provide any additional insight, the general pitfalls of the implementation and use of DW were more helpful. Turban and others (2014) list common implementation issues with data warehouses. The issues were, for example, setting unrealistic expectations, using information that is easily available but not useful, being unnecessarily technology-oriented instead of user-oriented, and discontinuing the development process when the first running version of DW is ready. Overall, the authors seem to emphasize good project management and information quality. To back up the previous points of emphasis, the general issues of the implementation phase seem to be in line with the most common problems of the use of data warehouses. The problems are, for example, system design, accessibility, and data quality issues (Shin, 2003).

5.2.1 Design model choice

In the implementation of a data warehouse, an organization should choose a design model, because it provides guidance for choosing the right architecture and in the implementation process (Breslin, 2004). There are two generally accepted design models for data warehouse design and implementation. The design models are the top-down and the bottom-up strategies that were first presented in the fourth section. Generally, the choosing happens between these two options, but it should be noted that the choice could be something in the middle (Breslin, 2004).

The top-down strategy is more commonly used of the two, but each strategy has its own strengths and weaknesses (Chenoweth, Corral, Demirkan, 2006). In the design model choice, there is no clear better or worse option (Turban et al., 2014). That is why each organization should define its own requirements in order to decide their preferable design model. Choosing the right design model depends on many different factors, such as previous experience and resources, and it also affects many elements of the BI system, such as the architecture and the implementation (Breslin, 2004; Turban et al., 2014).

The characteristics of the two design models were examined in the previous section, but to rehearse, the top-down strategy has more organizational impact and is generally larger as a project (Breslin, 2004). The top-down strategy requires more IT-skills and previous knowledge and it provides more a comprehensive view of an organization (Chenoweth et al., 2006). The bottom-up strategy provides more local impact and it chooses to operate on the departmental level. The methods of the bottom-up strategy are targeted for the end users and it is easier accessible (Breslin, 2004). Additionally, the data mart design is relatively fast to roll out (Chaudhuri & Dayal, 2011). However, the need for fewer resources and previous experience comes with a cost of having fewer data elements and it is less ambiguous in general (Breslin, 2004; Chenoweth et al., 2006).

Even though the choice of the design model is very important, it is possible to expand and evolve the data warehouse in the future. For example, if an organization would prefer using the top-down method and building a central data warehouse, it would be possible to start with the bottom-up approach. Using data marts is a convenient way to learn about BI systems without prior experience and the smaller scope data marts can still provide similar benefits than the EDW (Breslin, 2004). Even if an organization would start with the data mart design, it is possible to evolve it into an enterprise data warehouse strategy later. This could be a reasonable option if the organization wouldn't prefer or possibly couldn't commit to a large-scale warehouse project at the time (Breslin, 2004). On the other hand, if the organization would start with the large and central data warehouse design and the information needs of the local departments weren't fulfilled, the EDW design is also possible to expand later on (Turban et al., 2014). After the initial implementation, the enterprise data warehouse with a large organizational impact could be supported by building departmental data marts, and the local impact for the organization would improve.

5.2.2 Data warehouse architecture choice

The characteristics of the common data warehouse architectures were examined in the third section, and this subsection focuses on providing guidance for selecting the right architecture. The selection of DW architecture is an important choice, and it is connected with the choice of design model (Turban et al., 2014). The topic of choosing a data warehouse architecture lacks rigorous research and empirical evidence (Ariyachandra & Watson, 2008). Hence, only a limited number of referenced authors could be found for this topic.

Five common iterations of data warehouse architectures presented in the third section were independent data marts (IDM), data mart bus architecture (DBA), enterprise data warehouse (EDW), and federated architecture (FED). An additional option for the EDW approach, the hub and spoke architecture (HUB) was also briefly discussed. The problem of how to choose the right architecture will be approached by looking into the weaknesses and strengths of each iteration. It should be noted, that the data warehouse architectures are not only limited to the previous examples and there can be various combinations and hybrid versions of these designs (Ariyachandra & Watson, 2006).

The characteristics of each DW architecture were already discussed, and their performance-related factors are examined next. As one of the few quality studies about the topic, Ariyachandra and Watson (2006) have studied the performance and success of different DW architectures (Table 3). The data was collected by surveying organizations and it was used to rank DW architectures by different categories. The categories were information quality, system quality, individual impacts, and organizational impacts. The scores were given on a seven-point scale, and the higher number indicated being more successful.

Table 3 BI architecture scores (Ariyachandra & Watson, 2006)

	Independent data marts	Data mart bus architecture	Hub and spoke architecture	Enterprise data warehouse architecture	Federated architecture
Information quality	4.42	5.16	5.35	5.23	4.73
System quality	4.59	5.60	5.56	5.41	4.69
Individual impacts	5.08	5.80	5.62	5.64	5.15
Organizational impacts	4.16	5.34	5.24	5.30	4.77

The two lowest-scoring architectures were independent data marts and federated architecture. The IDM scored the lowest of five in all categories, and this result was not a surprise, since it is the simplest and least costly architecture (Turban et al., 2014). The FED scores a little better than the IDM but was still a clear second to last. The federated architecture could be considered as a compromise solution because it uses old infrastructure alongside new, and this limits possibilities of the data warehousing and thwarts implementation of an optimal system (Ariyachandra & Watson, 2008; Turban et al., 2014).

The best three architectures scored very similarly, and the differences were so marginal that one architecture could not be claimed to be the best (Turban et al., 2014). The similarity of the DBA, hub and spoke and EDW is most likely results from maturing and evolvement of the architectures (Ariyachandra & Watson, 2008). Over time the architectures have become more similar to tackle their weaknesses. Even though the scores varied more or less among the architectures, Ariyachandra and Watson (2008) emphasize that any of the major architectures can work and deliver good quality.

To state the obvious, even though there isn't a clear winner among the architectures, they definitely are not the same or equal in every case. Factors like development time, costs, and required planning are different for each architecture, and it should be noted that in practice, the amount of effort needed should also affect to the architecture choice (Ariyachandra & Watson, 2008), For example, it is clear that IDM scores lower than DBA in every category (Table 3), and by only inspecting the grades, there aren't any reasons to ever use the IDM. However, the conditions of an organization in reality and practice aren't always ideal, and there are many reasons that could potentially lead to the use of IDM, for example, limited tangible or intangible resources such as lack of time or money (Turban et al., 2014).

Because the best architecture for every situation couldn't be declared, Ariaychandra and Watson (2008) encourage to base the choosing on multiple important factors. The factors are for example time, resources, compatibility with previously used systems and technologies. Additionally, a list of ten factors that could have an effect on the architecture choice has been made (Ariyachandra & Watson, 2005; Turban et al., 2014). The list is the following:

1. How information interdependent the organization's departments are?
2. What are the upper management's information needs?
3. How quickly the data warehouse is needed?
4. What are the end-user tasks?
5. What are the possible limitations of resources?
6. What was the strategic view for data before the implementation?
7. Are the existing systems compatible with the new changes?
8. What are the abilities of IT staff?
9. Are there possible technical issues?
10. Are social or political factors limiting the system?

The choice of data warehouse architecture is clearly in line with the choice of the design model, and the same limitations and factors. As we can see from the previous list, some of the factors tie together with the key features of the design model. For example, the perceived ability of IT staff is something that can affect the choice of the design model, because the top-down strategy requires a lot more IT skills compared to the bottom-up strategy (Table 2).

5.3 Front-end

In order to use BI efficiently, the users of the system must possess sufficient skills and BI tools should fit the needs of the organization (Azvine et al., 2005). There can be multiple ways of how to proceed on the front-end side of an implementation process. Chenoweth and others (2006) have presented a simplified visualization of this process (Figure 6), and it provides general guidance into the topic.

Overall, this subsection will examine the users and the front-end BI tools of BI systems, and it aims to find the most important and common questions about the topic and provide guidance for it.

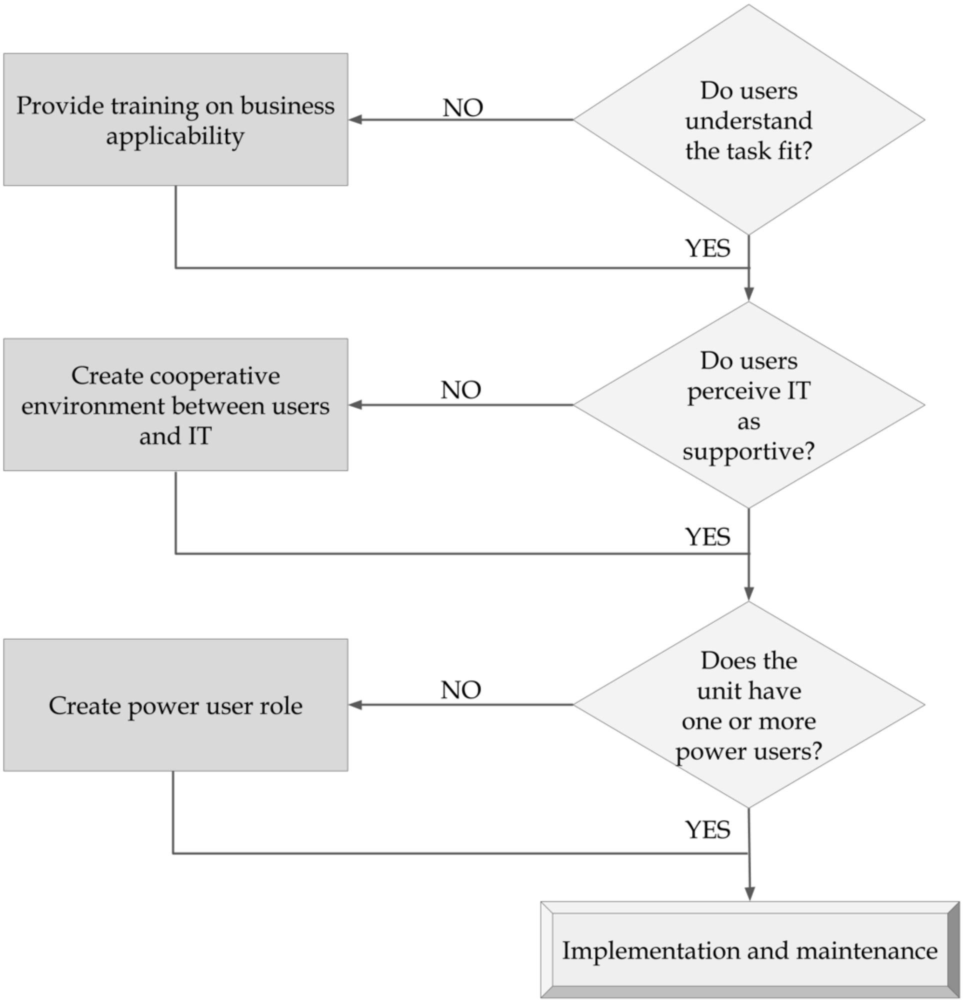

FIGURE 6 BI front-end approach (Chenoweth et al., 2006)

5.3.1 Choosing front-end software

The selecting of the right end-user tools for a BI system can be a hard task, because of the wide availability of different technologies. Additionally, the front-end BI tools can vary a lot in their complexity (Olszak & Ziemba, 2007). Organizations should choose the front-end software carefully because it is an important decision. For example, the functionality of the software can be a critical factor for productivity and effectiveness in the organizations (Amara, Solberg Søilen & Vriens, 2012), and it can also affect the information quality and the cost of decision making (Tutunea & Rus, 2012).

There is not a clear winner solution in the BI software market and declaring one as the most competitive option is impossible. Every tool has its own strengths and weaknesses, so it is up to the organizations to evaluate and choose the best one for their system (Amara et al., 2012). In general, the criteria used for the evaluation process are the functionality, the complexity, and the combability of the solution (Olszak & Ziemba, 2007). More specifically, the available BI software can be tested by utilizing free trials, demos, presentations, and third-party analysis. The more specific factors that should be considered in these tests are for example following (Amara et al., 2012):

- Ability to sort and display information using various rules
- Variety of different viewing models and visualizations
- Ability to view and demonstrate correlations and relationships between variables
- Having both standard and customizable templates for reports
- Ability to import and export data in various standard formats
- Ability publish reports in different formats, for example in print or digital form

The purchasing of a BI tool can also be dependent on other factors in addition to features and functionalities. One important factor in the selection of BI software is the provider of it. If the organization has already purchased any tools from one provider, choosing the same provider can have benefits, for example, bundle prices and improved combability between the tools (Olszak & Ziemba, 2007). On the other hand, buying multiple tools from the same provider can lead to not acquiring the most competitive option. Simplistically, organizations are choosing from two options: using one provider for multiple purchases to increase combability and possibly to lower the costs or acquiring the best tool for each purpose to guarantee the individual competitivity and performance of each tool (Olszak & Ziemba, 2007).

The last examined factor in the BI tool selection is the costs. BI systems should be built for scalability, and that also concerns the software (Olszak & Ziemba, 2007). It should be noted that the initial cost from acquiring the software is most likely not the last expense from it, and the possible expansions of the system should be kept in mind. Additional costs can also be for example updates or maintenance costs. One option for cutting down the costs is to utilize open-source software (Tutunea & Rus, 2012). Open-source tools can be considered as a serious option in the BI tool selection for multiple reasons. First of all, they are the less expensive option, because the license costs are low or nonexistent. Additionally, they can be tested and customized to specifics needs of the organization. The liberty and openness of open-source tools also lead to the possible downsides of them, which is the need for additional attention in certain areas. This means that extra work is required to achieve high security and data integrity (Tutunea & Rus, 2012).

5.3.2 User access

Fast and easy access to information is a good goal for every BI system (Ranjan, 2009). However, overseeing security and enforcing access are also great challenges of BI systems (Gangadharan & Swami, 2014). In general, user access should be fluent, but the different roles of users are something to consider. Every system and organization are different, but the users and their need for information access are different as well (Isik, Jones & Sidorova, 2011). Not every user of the system needs and even should have access to every feature or type of information. This issue can be solved by creating multiple different user roles for the software.

The need for different user roles can originate from different reasons. One reason is that there are user groups, that only need a few specific features of the software. For example, the users who manage the simple day to day business might need to access certain data with low latency, while the managers could only be interested in the big picture and strategic objectives (Isik, Jones & Sidorova, 2013). One benefit from this type of role division could be cost savings, for example, if the managers would only require view-only licences for the software. Another reason for user roles would be setting limits and restrictions (Isik et al., 2011). For example, an organization might possess sensitive data, that only certain users should be able to view.

Another possible way of setting up user roles for software would be dividing users to the basic and power users (Chenoweth et al., 2006). This means that the basic users would have access to the main features of the software, while the power users' access would be much less limited. The power users should have a very good understanding of both business and IT side of the organization. The users of this type are familiar with BI technologies and they are experienced with BI tools. Power users are usually granted access to every inch of the system (Chenoweth et al., 2006).

An organization could face a situation, where different user groups would have so dissimilar needs for the BI tools and their features that assigning various user roles might not be enough to address them. In this type of extreme cases, the organization might need to consider multiple BI tools for different user groups (Howson, 2006; Isik et al., 2011). Also, the situation could be that an organization might not feel the need for using any type of role. It should be noted, that the limited user access is not a mandatory thing, and some organizations prefer to provide full access for every user (Hostmann, Herschel & Rayner, 2007; Isik et al., 2013).

5.3.3 Training

One of the keys to BI success according to Watson and Wixon (2007) is to have users, who "have the necessary tools, training, and support to be successful". BI tools can be difficult to use and mastering them requires a lot of training and support. Additionally, understanding the basic features of BI tools is not enough, but on top of that, the users should understand the data they are using and what to do with it (Watson, 2009).

Giving enough attention to the users and their skills is certainly important, but Gangadharan and Swami (2014) go as far as suggesting that the success of a BI system is most dependent on the quality of training support for the users. It should be remembered that training is not only a one-time thing for the initial implementation, but it should be thought of as an ongoing process (Negash & Gray, 2008). The users can benefit from improved training when they have become more used to the tools, and new users may join the organization at any time. Continuous training is especially important in knowledge management. Training is not only needed for improving the competitiveness of an organization but to maintain it as well (Turban et al., 2014).

Training can be done in many different ways. The training methods, the trainers and the training environment can all vary. To succeed in the training of the users, every organization should design and carry out a training plan (Boyer et al., 2010). Not one organization is the same and this is why each organization should find out what are the optimal training methods for their context. Boyer and others (2010) have listed common training types that organizations could include in the training plan. The list of said training types and their strengths is provided below.

- Public classroom training – High engagement, little distractions, practical
- Onsite training – High engagement, detailed, practical
- Conferences – Comprehensive, possibility to network
- Online training – Cost-efficient, flexible
- One-on-one training – Very effective, good for total beginners

It should be noted that people respond to training methods differently and the efficiency of methods can vary in different environments. This is why each organization should find out what are the optimal and most natural training methods for their environment. (Boyer et al., 2010).

6 RESEARCH METHODOLOGY

This section goes through various details about the study and how it was conducted. Firstly, the research questions and objectives are examined, and after that, the case organization itself and its project will be discussed. Lastly, the research methods, methodologies, and the details of the framework are examined. This section tries to give an answer to the questions “What the study aims to achieve?”, “What is the case organization like and what is the role of it in this study?”, “Which research methods are used in this study?”, and “How the study attempts to solve the research problems and questions?”.

6.1 Research objectives

The topic of the study is vaguely business intelligence systems in general, but the more specific approach is to examine the design and implementation of BI systems. The study has various goals, problems, and objectives of different types. The approach that was taken to answer these challenges was to gather a large theoretical knowledge base that covers most important topics, that are business intelligence, BI technologies, and BI implementation projects. The theory base was utilized in the building of the framework, which could be thought of as the final product and result of this study.

Returning to the objectives of the study. There are various different objectives that the study attempts to achieve, and they are following.

- Building a comprehensive theoretical knowledge base of valid information
- Finding comprehensive answers to the research questions
- Finding answers and providing help to the problem of the case organization
- Building a framework that presents the answers in an informative form
- Demonstrating and evaluating the framework in a credible, valid and reliable manner

- Communicating the results following guidelines for scientific publications

Different factors that affected the research questions were the practical problems of the case organization and the research topic in general, but also the environment of the study had an effect to questions. Because of the chosen research methodology, which is design science, the issues and considerations of the case project were the main influencers, but in order to achieve generalizability, the practical side of the study had to be aligned with more common elements. The environment of the study and the context where the testing was conducted is public sector. This does not necessarily limit the study by a large amount, but it still is a factor that should not be ignored. The final research questions are following.

- What are the most important design choices and considerations of BI system implementation projects?
- How to approach these design choices and considerations?
- Does a generalized framework that has been built for common BI projects work for a public sector organization?

The research questions turned out to be relatively vague, which affects the scope and length of the study by expanding them. The downside of this would be that the study will be less compact and even cluttered to some extent, but on the other hand, the expanded scope will make the study more comprehensive. The next examined topic in this section will be the case organization of the study.

6.2 Case organization

The case organization of this study is the city of Mikkeli, located in eastern Finland. Mikkeli has about 54 thousand residents and it is the 18th largest city of Finland. As a public sector organization, the city of Mikkeli employs people from various industries and manages the infrastructure and services of the area.

6.2.1 The business problem of the organization

The situation that the case organization is facing, could be considered as a great environment to conduct such implementation research. Currently, most of the knowledge management of the organization is done manually, which practically means that it is slow and prone to human errors. The system that is now on use is managed by two entities, the case organization, and its IT-provider. The IT-provider manages the back-end, while the case organization is responsible for the front-end and the actual use. The current system doesn't necessarily have any major flaws, but it is used in an ad-hoc manner and not at its full capabilities.

The existing BI system of the organization has a very common architecture. It has the source data, ETL tools, data warehouse, OLAP cube, and few front-end software. As said, the organization is capable of producing ad-hoc reports from the source data, which consists of economic data like revenue, expenses, receipts, salaries, etc. The data is connected through common dimensions, such as ID numbers of different departments. Now the organization has a need to start using additional source data, which must be used and combined with the existing source data. The new source data does have a somewhat similar structure to the old data, but the problem is that the new data does not come from any system. The new source data consists of performance and use related values of various units around the city and it is stored in the local units and systems. The data is, for example, the number of visitors to the library, child count of schools, and operating hours of swimming halls. The reason for the implementation project can be slightly scattered and difficult to perceive, but overall, the practical problem of the case organization is to answer various design and technology-related questions about the architecture and implementation of a BI system.

Additionally, to the previous reasons for the project, the current front-end and user side of the BI system is not working to its full extent. The case organization is not only trying to solve issues with extracting and combining of the new source data, but another goal is to improve and automatize the reporting. To summarize the complex matter, the various issues that the organization has to consider are listed below:

- How to extract and load the new data source to the existing system?
- Does the new data type need its own data warehouse, data mart, or OLAP cube?
- At which point in the system architecture the two data sources should be combined?
- Is the currently used front-end software capable for the purposes of the organization?
- Are the users of the software capable to produce more than ad-hoc reports?
- How to train users to use the software to its full extent?

This study and the case implementation project will be conducted from the point of view of the central government of the organization, more precisely from the financial management department. Considering the nature of the task, the project could be more fitting for the IT department, but in this organization, the IT department manages mostly day-to-day tasks. As said, the project will be done in close cooperation with the IT provider, because the case organization does not possess enough know-how and equipment to design and execute most of the aspects of the project.

6.2.2 Role of the case organization

Firstly, to maintain the credibility of the study, it should be noted that the author of the study is working for the case organization. The study is made as objectively as possible and any biases are avoided to the full extent. The case organization does not have any reason for not being transparent, and the more precisely and truthfully the situation is described, the more value it provides to the organization. Additionally, the nature of public sector organizations allows and even encourages to be transparent and objective, because public sector organizations are not competing in the market like the private sector organizations are (Nutt, 2005).

The study was initially requested by the case organization, but the specific topic and approach of this study took its shape later on. In the beginning, the expectation of this study was to evaluate and improve the system that was being implemented. Because of the lack of commitment from the organization, the research topic steered from being very organization-specific, to more generalizable and at the same time more useful for the BI research field.

The case organization is valuable for this study mainly for two major reasons. Firstly, this kind of implementation project is a great environment to test and consider various approaches, because it has many open ends. Secondly, it is a great opportunity to be able to utilize a practical environment to improve the study, without having pressure with confidentiality or outcome, as opposed to private sector organizations.

As mentioned earlier, the main goal of the study is to create a framework that can be utilized in future BI implementation projects. The role of the case organization is to act as a test environment for the framework. Having very close cooperation between the study and the organization provides major benefits. For example, from the thought process to the progress, every part or stage of the project can be reviewed and used to improve the framework. If the case project has any considerations that are important and not too context-specific, the considerations will be added to the framework.

6.3 Design science research methodology

The chosen research method for this study is design science. This subsection will go through three things: the main ideas behind design science, why it was selected for this study, and how it will be applied.

Design science was initially known from the field of engineering, but it also has its place in IS research. In design science, knowledge is generated in a different way compared to "traditional" sciences and the methodology could be considered as scientific problem solving (Hevner, March, Park & Ram, 2004). "Whereas natural science tries to understand reality, design science attempts to create things that serve human purposes" (March & Smith, 1995).

Hevner and others (2004) have created seven guidelines for DS in information systems. The guidelines demonstrate the characteristic of DS research, and they can be used as a framework for conducting research. The guidelines of Hevner and others (2004) are listed below (Table 4).

TABLE 4 Design science guidelines (Hevner et al., 2004)

Guideline	Description
1. Design as an artifact	DS research must produce an articaft, and it can be in the form of a construct, a model, a method or an instantiation.
2. Problem relevance	DS research must seek to develop a technology-based solution to important and relevant business problem.
3. Design evaluation	The design artifact must be evaluated rigorously to testify the utility, quality and efficacy of it.
4. Research contributions	DS research must provide a verifiable contribution to the research field.
5. Research rigor	DS research must applicate rigorous research methods.
6. Design as a search process	Available means must be used in the search for an artifact to reach desired goals.
7. Communication of research	DS research must be presented effectively to different audiences.

Choosing design science as a research method is a good fit for this study considering the guidelines and the fundamentals of DS. The case organization has a problem, that could be considered as both IT and business-related. Additionally, it is possible to construct an artifact to help with the problem. In this study, the framework for BI system implementation will act as an artifact.

Naturally, the previous guidelines and existing knowledge base of DS theory will be utilized in this study. More specifically, the study will use the design science research methodology (DSRM) process model as a framework for conducting the study and building of the framework (Figure 8). The process model contains six steps, which were adopted from other studies and the guidelines of Hevner and others (2004). The six steps or phases of the DSRM process model are (1) problem identification and motivation, (2) definition of the objectives for a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication (Peffers et al., 2007). The process model is presented in figure 7.

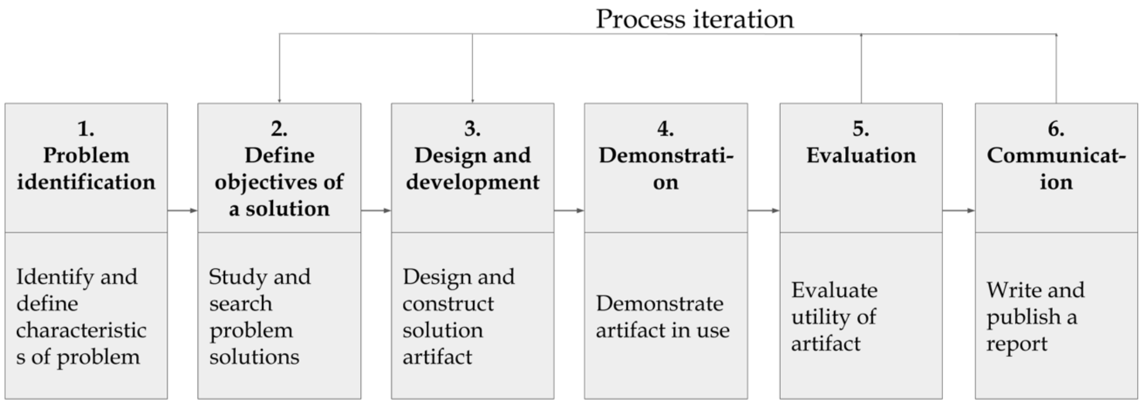

FIGURE 7 DSRM process model (Peffers et al., 2007)

6.4 Framework

The main goal of this study was to create a framework to guide organizations with the implementation of a BI system. The idea of the framework is to contain a comprehensive package of information but to keep each topic of it compact. As mentioned in the previous subsection, the framework will be evaluated, tested, and improved in a continuous cycle.

The case organization of the study is operating in the public sector. However, a claim could be made that in this study, the sector of the case organization does not limit the framework that much. As it was found in the second section, there are many differences between the two sectors, but they do not necessarily affect the architecture and technologies of the BI system. The major differences were that the public sector was found to be more transparent and less competitive (Nutt, 2005). Additionally, the public sector is considered to be less flexible and it doesn't have as clear goals as the private sector organizations usually have. These definitely are factors that should be considered in the framework, but at the same time, they are not limiting it too much. As said, the framework could be marginally more useful for public sector organizations, but it is still usable for the private sector as well. However, this consideration will be inspected more closely further in the study.

6.4.1 Development

The construction of the framework will begin with the first step of the DSRM process model, which was identifying of the problem and motivation. The identifying will be done to the problem of the case organization and it will be presented as a research question. In practice, this means that the problem must be first identified, then defined, and then it is possible to understand what main objectives of the study are. Roughly, the main problem of the case organization was known in the very early stage of the study, because the right direction was needed to build an appropriate theory base. Conveniently, the next step of the process model is to define the objectives. This step requires knowledge of the current problems, current solutions, and the efficacy of the solutions (Peffers et al., 2007).

After the problems and the objectives are defined, it is possible to move on to the third step, which is the design and development of a solution. The artifact, which in this case will be the framework, will be first constructed based on the theoretical knowledge base of the study. When the first version is ready, it can be tested and evaluated.

6.4.2 Demonstration and evaluation

Without any evaluation, the designed artifact will only act as a hypothesis that it could possibly solve some kind of a problem (Venable, Pries-Heje & Baskerville, 2012). Even though the demonstration and the evaluation are two different phases with different purposes, both are conducted by using similar logic to find out if the artifact fulfils the purpose of it. Venable and others (2012) have found five different purposes for evaluation in design science research. The purposes are the following.

1. Evaluate the artifact for its utility and how well it achieves its purpose
2. Evaluate the formalized knowledge and design methodologies that were used to build the artifact
3. Evaluate the artifact by comparing it to similar artifacts
4. Evaluate the artifact by the possible negative or positive side effects of it
5. Evaluate the artifact to identify its weaknesses or areas of improvement

Fortunately for the quality of the study, there is an opportunity to demonstrate the utility in the case organization, which works as a very viable testing ground. There are various possible ways to test a theory or framework (Peffers et al., 2007), but considering the research topic, its scope, and the project itself, there are not many viable methods. Few examples of the problems will be presented. If the quality of the framework would be evaluated by the quality of the system that has been built with the help of the framework, there would be few problems. Firstly, the quality of the BI systems can be affected by unfortunate conditions which have nothing to do with the framework itself, for example, a lack of resources. Secondly, if the performance of the final system would be evaluated, it would prolong the study by a great time. Considering the measure of the project and the nature of the context environment, it takes a long time for the project to be complete. The evaluation method of this study will be following the first and the fifth purpose of Venable and others (2012). Naturally, in an optimal situation, all the five evaluation purposes would be used, but it is not likely that all of them would be required for justifying the artifact. The objective or purpose of the framework is to provide guidance for organizations and to be comprehensive enough to cover the most important areas of the research topic. By evaluating how many issues of the case organization the framework can solve or help with, could be considered as a sufficient evaluation method to draw fairly solid conclusions. To avoid evaluating the quantity of information over quality, the framework has to truly offer help with considerations, so that exact topic can be accepted as being covered or sufficient enough in the framework.

Hevner and others (2004) have categorized various evaluation methods for different types of DS studies (Table 5). There are 5 different evaluation types, which each contain a few different evaluation methods. This study fits best with the observational category because the case study is the closest description of how the study will be conducted. However, it could be argued that the evaluation method is not purely only a case study. For example, the method that will be used for this study also meets the description of the dynamic analysis.

The methodology of testing will be following the fourth and fifth steps of the DSRM process model, which are demonstration and evaluation. The demonstration step means that the artifact will be used for example in a simulation or in a case study, and in the evaluation step, the fulfilment of the objectives defined in the second step is studied (Peffers et al., 2007). Naturally in this study, the demonstration step will be conducted in the case organization. Considering the possible ways to evaluate an artifact, the previously listed context-specific problematics, and the nature of the implementation project, the evaluation of the framework will be made based on how comprehensive and helpful it is.

TABLE 5 Design science evaluation methods (Hevner et al., 2004)

1. Observational	Case study: Study an artifact in a business environment
	Field study: Monitor the use of an artifact in projects
2. Analytical	Static analysis: Evaluate different static qualities of an artifact
	Architecture analysis: Evaluate fit of an artifact in IS architecture
	Optimization: Demonstrate optimal properties or behaviour of an artifact
	Dynamic analysis: Study the use of an artifact to evaluate the dynamic qualities of it
3. Experimental	Controlled experiment: Examine an artifact in a controlled environment for its qualities
	Simulation: The use of an artifact with artificial data
4. Testing	Functional testing: Execute artifact interfaces to discover failures and identify defects
	Structural testing: Execute coverage testing of some metric in the artifact implementation
5. Descriptive	Informed argument: Use information from the previous knowledge to convince the utility of an artifact
	Scenarios: Use detailed scenarios to convince the utility of an artifact

As mentioned earlier, the first version of the framework will be constructed by utilizing only the previous theoretical knowledge. However, if any shortages or flaws are found in the demonstration and evaluation phase, changes will be made. In this case, the demonstration and evaluation phase will be repeated until the framework satisfies the problems of the business environment.

7 FRAMEWORK FOR BI SYSTEM IMPLEMENTATION

The main objective of the study was to create a framework that provides help with the design and implementation of BI systems. The previous sections have been the groundwork for the framework. This section aims to process and merge the most important information of the study into the framework and later evaluate and develop it accordingly.

The structure of this section will follow the steps of DSRM process model, which proceeds in this order: (1) problem identification, (2) definition of the objectives of a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication (Peffers et al., 2007). Excluding communication, every part of the DSRM process model will be addressed in this section. Naturally, this study as a whole will act as the communication phase, so it does not require a section of its own.

7.1 Problem identification

The problem identification phase began very early in the research process because in this type of research, the problems of the case organization determine the perspective of the research questions. Naturally, the identification of the problems is important to the artifact in a general way, but additionally, it is needed in order to guide the theory-building process of the study. The main objective of the case organization was to find a way to automate reporting, which required implementation, refinement, storage, and visualization of the new source data type. The problem was to figure out how to design and execute the previously listed tasks. The more precise details of the problem were examined more comprehensively in the previous section, but to rehearse, the primary research questions were "What are the most important design choices and considerations of BI system implementation projects?", "How to make decisions and generally approach the design choices and considerations?", and "Does a generalized framework that has been built for common BI projects work for a public sector organization?".

A task regarding the practical problem was to shape the research questions accordingly. The case problem itself covers a relatively large research area and there are multiple matters that could be used for a research topic alone. However, the scope of this study was selected to be ambitious and large, and it views the BI system and its implementation in a big picture. When a single research covers various technologies and factors regarding them, it is not convenient to examine each topic exceedingly thoroughly. This leads to another issue with the case problem, which is how to select and present a proper amount of information for each topic or technology.

7.2 Definition of the objective of a solution

The problem of the case organization was the design and implementation of a BI system, so the main objective of the artifact must be something that provides help to the problem. This matter was approached by considering ways to guide the organization with the problem. Of course, the theoretical knowledge of this study does provide help to some extent, but the information is spread over multiple sections and it lacks a practical touch to the matter. Therefore, the framework should remain compact, informative, and have some practical side to it.

Examining the design and implementation of BI systems is a broad topic, so it is natural to divide it into parts. In the theoretical section of the study, the BI system was divided into different sections by the architecture of it. The division was roughly the following: data sources, ETL, data warehouse, and users. Each of these parts was covered twice. Firstly, the characteristics were examined in the third section. Secondly, the considerations concerning implementation were discussed in the fourth section. Such an approach appeared to be working for the theory building, and it may suit the framework as well. However, going over the characteristics and fundamentals of each technology is not something that is necessary to present in the framework. The framework itself has to cover a broad variety of topics and details, so anything that does not add crucial value should be left out. Selecting the major objectives of the framework was not particularly problematic. Deciding what kind of information and how much information on each objective should be fitted into the framework was a major consideration.

In a summary, the main objective of the artifact is to provide guidance for the design and implementation of BI systems, and that objective can be divided into parts by inspecting different components of BI system architecture. The next subsection will explain how each objective and component will be examined in the framework.

7.3 Design

The framework will be presented as a visualization, but for it be to comprehensive and helpful enough, it naturally includes a text section as well. To achieve the goals and objectives of the study, the framework has to be informative but compact. As mentioned earlier, there are multiple different components of a BI architecture, which are presented in the framework. To rehearse, the five components are data sources, ETL, data warehouse, front-end software, and users. The components are the main parts of a BI system, and it is natural to build the framework using a similar division.

Listing the BI components and representing the architectural design of a BI system can be helpful for understanding BI systems in general but examining only the technical aspects and choices of BI does not offer extensive enough help with the implementation side of BI projects. This is why in addition to the architectural components of BI, the critical success factors and implementation considerations should be fitted into the framework to provide a more comprehensive view of the subject.

The visualization of the framework (Figure 8) can only fit limited information, but it was possible to include most of the major aspects of a BI project in it. To some extent, the design of the visualization is adopted from process models, but it lacks many elements and practicality of them. For example, process models usually present specific activities for certain roles, and they are chronological. This type of information is difficult to define in this type of research because it aims to achieve generalizability. Additionally, BI systems and organizations vary, and the development process of a BI system is usually more of a cycle instead of a linear process (Gangadharan & Swami, 2004).

There are three different types of information given about each of the five parts of the framework. Consequently, the implementation considerations, the design choices and activities, and the critical success factors of each BI system architecture component will be presented in the visualization. These three themes were selected for framework, because they are something that was already examined in the study and as mentioned because the framework required information of both BI system design and BI implementation. The study has examined various CSFs and implementation considerations. Some of them are clearly affiliated with certain parts of a BI system, but many of them are very general and could be linked to any part of it. Therefore, the CSFs and additional information will be assigned to the closest or best fitting part of a BI system.

The information on the framework is based on the theoretical knowledge base of the study. It should be noted that the framework could hold information that is only useful for a generic BI system project but not necessarily for the project of the case organization. As mentioned, if any shortcomings are found in the demonstration and evaluation phase, the framework will be complemented. In principle, the framework can only be extended and improved. However, if there is information that is not needed or utilized in the case project, each instance must be handled individually.

7.4 Development of the artifact

The first version of the framework (Figure 8) was designed by considering the problems and objectives of the case project and the subject of BI systems in general. Additionally, the initial visualization was presented in two parts (Figure 8a & Figure 8b) for better understandability. In this subsection, the initial framework will be presented and each part of it will be examined and further explained. Most information on the framework could be fitted into the visualization, but naturally not everything. Before examining the visualization, there is one major set of activities that should be discussed before going into more technical details. Those activities are defining of the system in a big picture and setting the requirements for it, which are something that should be done at the start of any BI project (Chaudhuri & Dayal, 1997).

Defining accurate requirements for BI systems can be difficult until there is available test data (Inmon, 1992). However, if the implementation is done in an iterative manner, it isn't a problem if changes in the initial requirements happen (Reinschmidt & Francoise, 2000). Few most important questions that organizations should ask themselves at the start of the project are "What are the reasons for using the BI system and what is the expected value?", "Who uses the system and how the decisions are made in the organization?", and "What are the strategic goals for the BI system?" (Gangadharan & Swami, 2004). After the big picture macro requirements are defined, next up is the more detailed and technical micro requirements (Elena, 2011). This type of requirement can for example concern security, user access, data volume, data storage, and performance-related requirements (Ranjan, 2009). Next up, the rest of this subsection will focus on the design choices, requirements, and critical factors that were presented in the visualization of the framework, and it will proceed in a similar fashion to the visualization.

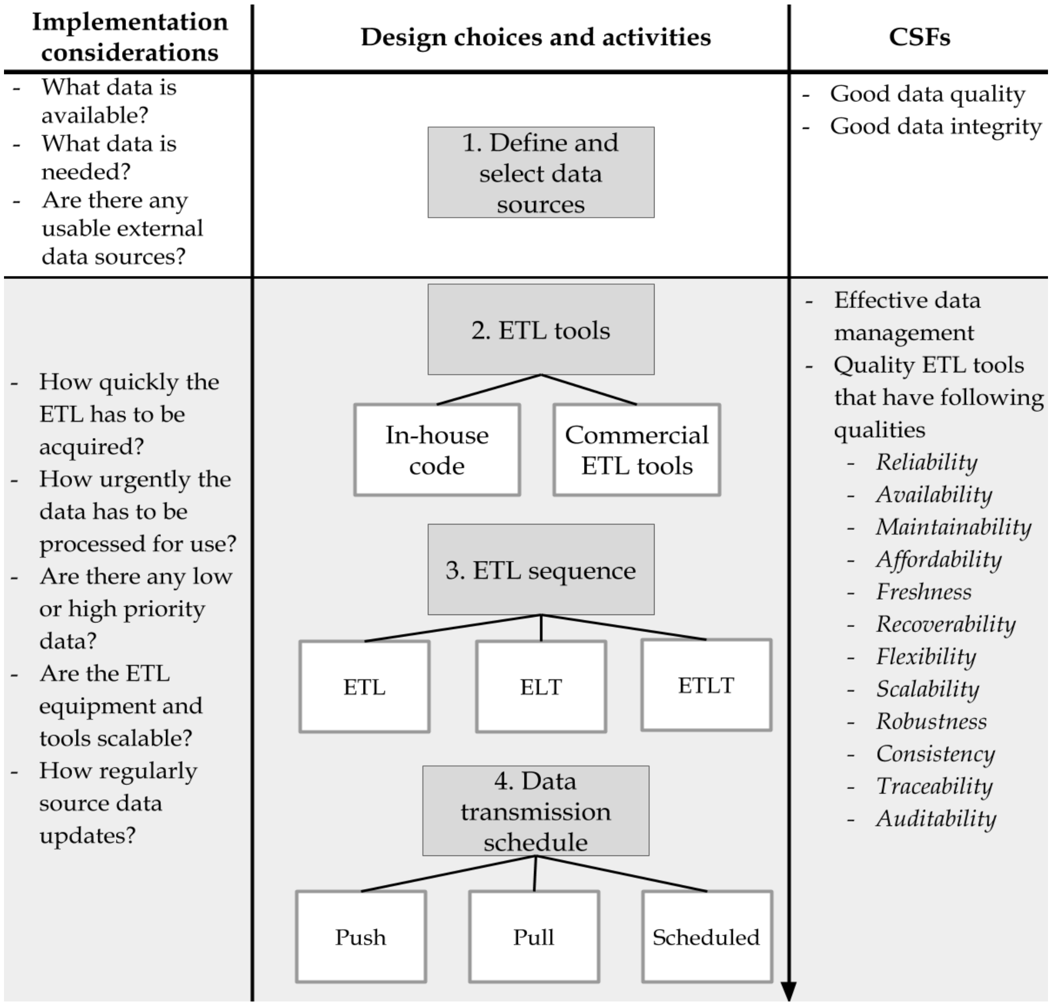

FIGURE 8a Framework for BI system projects adopted from Ariyachanda & Watson, 2008; Chaudhuri et al., 2011; Negash & Gray, 2008; Ranjan, 2009; Turban et al., 2014

7.4.1 Define and select data sources

Source data is the key element of BI systems because it will be processed into information and later on into knowledge. Selecting the right data source is an important task because it can affect the quality of decision making (Moss & Atre, 2003). The first considerations concerning data and the source of it are to define what kind of data would be required and what data is available. There is no reason to flood the BI system with useless data, even if it would be easily available. If the most obvious data sources do not fulfil the desired qualities of source data, organizations could search alternative sources, such as external data sources, or they could redefine their requirements (Ranjan, 2009). Organizations should seek a balance between the quality and the quantity of data since the lack of one and the abundance of another is not very useful (Popovič et al., 2009).

In addition to defining what data is available, wanted, where it is from, and what is the quality of it, the structure of the data should be examined. There are three common data structures, which are unstructured, semi-structured, and structured data. Each structure has its own qualities and they are processed differently. It is useful to define the structure of the source data because this information is needed when designing ETL and data warehouse (Negash & Gray, 2008).

7.4.2 ETL

It is possible that the design and implementation of ETL could seem trivial, but it should not be considered as such (Dayal et al., 2009). It is far too common to underestimate the required effort for ETL and the design of it because in practice it is a laborious process (Jun et al., 2009). Organizations should make sure that ETL receives enough attention. For example, the requirements for the ETL should be defined thoughtfully, because they are needed for various design-related decisions, that are examined next.

The first important decision in the ETL implementation is to consider whether to acquire ETL technology by purchasing commercial tools or to develop the tools in-house. In general, the major differences between the two are found in costs, flexibility, and complexity. A more detailed comparison is presented below (Table 6) and additional information can be found in the fifth section.

TABLE 6 ETL tool comparison (Jun et al., 2009)

ETL tools	sufficient	easy	good	short	low	high
In-house coding	good	difficult	good/unsure	long	high	low

In order to make a decision, an organization should understand its own capabilities, needs, and resources. If an organization is confident in the capabilities of their developers and they have previous experience of ETL tools, choosing in-house coding would be a possible option. This approach could lead to reduced costs and increased flexibility of the tools. However, if the organization has either a need for an urgent implementation or a lack of skilled developers, they should consider purchasing commercial ETL tools. Well-known and tested products are naturally the safest option (Jun et al., 2009).

Coming to the next choice, which is choosing the sequence of the extract, transform, and load. The most common approach for this is the E, T, and L sequence, but the other two options are ELT and ETLT (Dayal et al., 2009). The ETL approach should be sufficient choice in most cases, but the alternatives have their own strengths. If the system that is running the ETL tools is having poor performance, or if scalability is an issue, the ELT could be the right choice. In this approach, the data is loaded to the warehouse, where it will be transformed (Dayal et al., 2009). If the data warehouse of the system has better performance than the ETL equipment like it usually is, the transformation can be done more efficiently (Vassiliadis & Simitsis, 2008). The second alternative to the ETL is the ETLT. The strength of the ETLT option is that the source data is transformed into two phases. If an organization clearly has data that must be available very urgently, but it also has data that is not as heavily prioritized, then the ETLT could be the right and the most beneficial option of performing the extracting, transforming, and loading (Vassiliadis & Simitsis, 2008).

The last design choice considering the ETL is the data transmission schedule, which is likely the most straightforward choice of the three. There are three options for the schedule, which are periodic, pull, and push transmission. The periodic transmission is a simple and commonly used option (Dayal et al., 2009) and it most likely the safest choice. However, if the organization needs fresh and up to date data as soon as it is available, the push transmission would be the clear winner. The pull transmission would be the best option if the source data updates irregularly and the organization requires more flexibility in the transmission schedule (Vassiliadis & Simitsis, 2008).

The success of BI systems depends on many factors, that were discussed earlier in the study. As mentioned, the critical success factors of BI systems that were most directly fitting to certain parts of the system are listed in the framework. The CSFs that were affiliated with data sources and ETL are good data quality, data integrity, effective data management, and quality ETL tools.

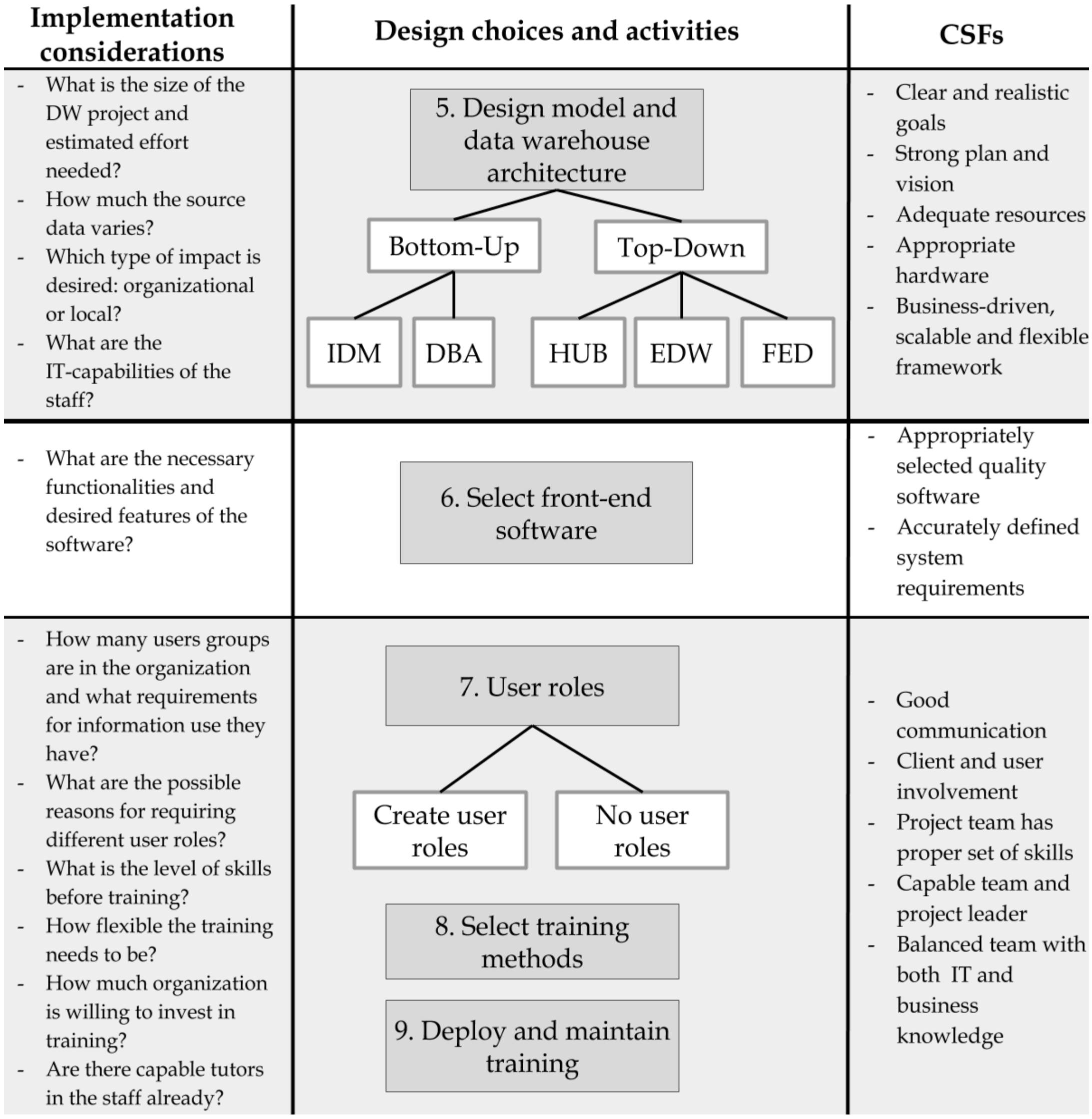

FIGURE 8b Framework for BI system projects adopted from Ariyachanda & Watson, 2008; Chaudhuri et al., 2011; Negash & Gray, 2008; Ranjan, 2009; Turban et al., 2014

7.4.3 Design model and data warehouse architecture

When designing BI system architecture, it would be helpful to consider which one of the two common design models fits better for the requirements of the organization. It should be noted that the choice for the design model is not as unambiguous as the theory make it seem. For example, if an organization considers that the top-down strategy would be the best approach for them in general, it does not necessarily mean that they are locked in with that choice and they must aim to fulfil every characteristic of that approach. The design model does not give practical help with the implementation of a BI system, but it can guide the organization in the right direction with the design and the architecture (Turban et al., 2014).

Design model choice is naturally organization-specific, but there are some characteristics and factors that can be used to make a decision. The models were discussed more comprehensively in the fourth section, but the most important factors are discussed here. Firstly, the Kimball model, which is also called the data mart approach or bottom-up strategy. This approach is considered as the small-scale option for data warehousing and BI systems. The scope, costs, development time, and difficulty are smaller or lower than in the top-down model. The data mart approach of Kimball is a better fit for small to average size projects and it focuses on the local and departmental areas (Turban et al., 2014).

The Inmon model, also known as the EDW approach and top-down strategy, is mostly for large-scale projects. This approach is associated with projects that have longer development time, high costs and it aims to cause organizational impact. In general, the top-down strategy is a better fit for large organizations, that have a high number of users (Turban et al., 2014). As can be perceived, the main factors that separate the design models are scope, costs, size, and time. An organization should first examine which model fits best for their needs, and then consider the BI architecture. As was said earlier in the study, the architectures are not necessarily tied to the design model, but they can steer organizations in the right direction. With the help of various studies, it was possible to construct a table for comparing different BI architectures. The following table (Table 7) compiles and presents various factors and measures of the five BI architectures that have been used in this study.

TABLE 7 BI architecture comparison adopted from Ariyachandra & Watson, 2006; Ari-yachandra & Watson, 2008; Turban et al., 2014

	IDM	DBA	HUB	EDW	FED
Associated design model	Kimball/Bottom-up	Kimball/Bottom-up	Inmon/Top-down	Inmon/Top-down	Inmon/Top-down
Information quality	Low	High	High	High	Low
Typical project scale	Small	Small	Large	Large	Medium
Development cost	Low	Medium	High	High	Medium
Main impact area	Departmental	Departmental	Organizational	Organizational	Organizational

There are multiple factors and measures that separate the common BI architectures. The previously presented information about the design models and BI architectures can possibly guide an organization to the right path in architecture selection, but it has become evident that it is very difficult if not impossible to be able to recommend the absolute best architecture for every situation and context. For example, if an organization would prefer a large-scale BI system with a wide organizational impact, according to the design models and the characteristic of BI architectures, the best fit for this description would either be the EDW or the HUB architecture (Ariyachandra & Watson, 2008; Turban et al., 2014). However, if we examine the BI architecture scores (Table 3), it can be perceived that the data mart bus architecture has marginally scored highest in the organizational impact category.

When approaching the data warehouse choice, the design models can provide a helpful first step. If an organization prefers bottom-up strategy and considers it as a better fit, then in principle the IDM and the DBA architectures are something to contemplate. This is because the main characteristics of the bottom-up strategy are small to average project scale and the emphasis on departmental impact. The IDM and the DBA architectures utilize data marts, which could be considered as a small-scale variant of data warehouses, and data marts typically focus on one business area (Turban et al., 2014). The choice of whether to use independent data marts or data mart bus architecture could potentially be trivial. IDM scored lowest in every category of the BI architecture study by Ariyachandra and Watson (2006) and it can be considered as an inferior solution. The benefits and good qualities of IDM are low costs and short development time (Turban et al., 2014). Unless an organization has a lack of resources and a need for an ad-hoc solution, it wouldn't be advisable to select the IDM architecture. DBA architecture can provide the same good qualities of the data mart approach as the IDM, but the DBA is much more versatile. The main reason for this is that DBA is scalable, and by increasing the number of connected data marts, the organizational impact increases as well (Ariyachandra & Watson, 2010). DBA architecture is the better choice for bottom-up strategy in most cases.

If an organization prefers the top-down strategy, there are three recommended architectures, which are enterprise data warehouse architecture, hub and spoke architecture, and federated architecture. These three have a lot in common, but they do have some characteristics of their own, that can help with the choosing process. Starting with the federated architecture, which differs from all the other options, because it leaves the legacy systems intact. FED could be considered as a realistic approach for BI architecture, because it is not the most optimal way, and that is how projects often turn out to be (Turban et al., 2014). FED had the second-lowest scores in the BI architecture comparison, so there is only one reason for recommending it to an organization (Ariyachandra & Watson, 2006). If an organization has a good reason for implementing and combining the new BI system with the existing systems, then the federated architecture can be recommended. However, if the preferred design model of an organization is the top-down strategy, but there are not any legacy systems in play, it would be recommended to consider either EDW or HUB.

Both EDW and HUB use a central data warehouse with a large organizational impact. The main difference between the two options is that in addition to a data warehouse, the hub and spoke architecture also utilizes data marts. EDW is a very scalable option and it provides an enterprise-wide repository (Ariyachandra & Watson, 2010). If an organization prefers the top-down strategy, EDW will most likely provide good value for them. However, the one major downside of the top-down compared to the bottom-up strategy is the lack of departmental impact (Turban et al., 2014). Even though EDW is the most popular architecture and could be considered as a safe choice, there is a possibility that an organization would require more emphasis on certain business areas (Sen & Sinha, 2005). In this type of situation, the hub and spoke architecture would be recommended over the enterprise data warehouse architecture because it provides more departmental impact. To conclude the design model and BI architecture section, an important notion is given once more - even if an organization prefers one design model, it does not necessarily limit the selection of BI architecture to two or three. The choice of the design model is only used as a basis to make recommendations. The critical success factors that should receive special emphasis in the implementation of data warehouses are setting clear and realistic goals, strong plan and vision, adequate resources, appropriate hardware, and good framework.

7.4.4 Front-end software

Choosing the right front-end software is an important choice because the capabilities of the software could affect productivity, effectiveness, information quality and decision making of an organization (Amara et al., 2012; Tutunea & Rus, 2012). The choice should at least be based on the organization's own research, but preferably to testing or seeing the software in use. Possible methods to evaluate the software efficiently are demos, trials, presentations, and third-party analysis. The measures and factors that should be considered in the evaluation process are the abilities to sort, filter, and aggregate information in various models and visualizations, and the abilities to export and import information in various formats, including the industry standards (Amara et al., 2012).

Another factor in the choosing of the software is the provider of it. This naturally is not as important as the functionalities and the performance, but the factor should still be considered. The main achievable benefit from using the provider as a factor in the software choice is that if an organization already uses products from the same provider, there could be added value from increased combability (Olszak & Ziemba, 2007). However, it isn't advisable to make the provider of the software a high priority in the choosing process, because it is not likely that increased combability can make generally inferior software a good choice.

Costs are an important factor in the software choice, but as the provider choice, it should not be the first measure to base the choice on. It should be noted that the initial costs of the software are seldom the only costs (Olszak & Ziemba, 2007). The additional costs such as licences, updates, maintenance, and support should be taken into account when comparing prices. One possible option that can provide an opportunity to reduce costs is to consider using open-source software. If the open-source software fulfils the requirements that the organization has set for the front-end software, the option could be very viable. Additionally, if the organization possesses sufficient capabilities, the open-source software could be customized to the liking of the organization (Tutunea & Rus, 2012). Lastly about the front-end software are the CSFs that were best fitting for this part of the BI system. The factors were appropriately selected software and accurately defined system requirements.

7.4.5 Users

There are three main activities and choices concerning users in the visualization of the framework (Figure 8), and they are user roles, training methods, and deployment and maintaining of training. Beginning with the user role topic, which is mostly concerning the front-end software. In principle, simple and effortless access to information is an important feature of a BI system, but there can be good reasons to make limitations to it (Gangadharan & Swami, 2014). A common reason for creating different user roles are limiting user access because some user groups do not need access to some features. For example, if an organization has users who use the software but do not require editing rights, it could be beneficial to purchase view-only licenses for those users to reduce costs and to make sure only certain employees can make commits. However, if a BI system contains information that should not be visible to everyone, instead of limiting features for certain users, the viewing rights of sensitive data ought to be limited. Another way of dividing user roles is to consider using power users. If an organization possesses capable users, who are familiar with BI technologies and understand both IT and business sides of the system and the project, it could be recommended to divide user access to power and normal users (Chenoweth et al.,2006). This way the power users would have unlimited access to features and information, and it reduces the chance that incapable users could not do any unintentional damage to the system.

Keeping the system secure and overseeing access to it is one of the major challenges of BI (Gangadharan & Swami, 2014), and organizations should seek to find a balance between rigorous security and liberal policies on user access. The last note about user access is that the user roles are not an answer to everything. If an organization has two or more user groups that have very different needs for the front-end software and information use, it could possible that the organization needs multiple software (Howson, 2006; Isik et al., 2011).

Easy but secure user access is an important quality in BI systems, but even more important is the competency of the users. Even if a BI system has been designed, implemented, and tested properly, and the system is found to be working and usable, the success of BI projects depends a lot on the users and employees (Gangadharan & Swami, 2004). The users of a BI system should possess various competencies, for example, marketing, BI software and architecture knowledge, database administration, project management, and business analysis (Debortoli et al., 2014). Naturally, if an organization does not possess these competencies, the users must be trained. However, even if all the major competencies are satisfied on a sufficient level, training should still be conducted, because, in the information management field, continuous development is required for surviving in the market (Turban et a., 2014).

Every organization should create a training plan, which demonstrates how the training will be conducted. Every project is different, and that is why organizations should consider their own needs for training and which methods could work best for them. The most common training types are public classroom training, onsite training, conferences, online training, and one-on-one training (Boyer et al., 2010). Briefly, the main reasons for choosing the previous methods are the following. Public classroom training and onsite training can provide good engagement and practicality to the training. Onsite training is naturally closer to the real work environment, but on the other hand, public classroom training can be more effortless to organize. Both are most likely safe and all-around choices. Conferences usually are comprehensive, and they have people attending from various organizations. This enables the possibility to network and discover ideas from another perspective (Boyer et al., 2010). If an organization values such matters, conferences could be a good training method. Online training can be considered as cost-efficient and very flexible. Especially remote workers could benefit from online training. If an organization considers online training as a sufficient enough method, it could be a good choice, but if not, online training could be used as a supportive training method. One-on-one training could be for example an experienced user training a new employee, and in this case it most likely would be an informal situation with a supportive atmosphere. This type of training can be an efficient method in many ways, especially in costs, because it could be possible for the trainer to also carry out daily tasks meanwhile.

Lastly, about the training, it should be remembered that it is not a one-time deal. Training should be considered as an on-going process, because new users might always join the organization and because the field of BI requires continuous development (Negash & Gray, 2008; Turban et al., 2014). The CSFs that were the closest fit with training, users and user roles are good communication, client and user involvement, capabilities of the project team and its leader, and the balance between IT and business knowledge.

This was the initial version of the framework, and the utility of it will be demonstrated and tested in the next subsection.

7.5 Demonstration

Before the demonstration phase, theoretical knowledge of BI systems was already acquired, and gathering information about the case organization was not necessary, because the author and the project group were very familiar with the project and the organization. In addition to theoretical knowledge, new empirical knowledge was required.

The objective of the demonstration phase is to prove that the artifact is working and usable in the purpose of it. The proving can be done by exposing the artifact to a real business environment (Peffers et al., 2007). In this study, the utility of the framework is tested in the case project. The approach in this was to use the framework as a guide in various problems and considerations that were faced in the case project. If the framework did provide sufficient help with a specific design choice or a problematic consideration, then the said area of the framework was left in the original form. However, if the help provided was not sufficient enough or if that specific problem was not included in the framework in the first place, then that area had to be updated and improved. The demonstration phase proceeded with the pace of the project, and it took months to cover every component and area of the framework from start to finish. The overall process was done in cooperation with the researcher and the case organization.

The actual way of finding insight into the utility of the framework was not very rigorous and scientific, because the topic was examined in an open discussion. Each part of the framework was discussed and notes about the discussion were kept. Like the discussion itself, the notes were also flexible in format. Even though various project members took a part in the demonstration phase, the memos and the discussion were heavily managed and lead by the author. The discussions and meetings were held either physically in the workplace or remotely online. Additionally, and unfortunate fact about the demonstration and evaluation phases was that it was done internally to the organization and the project group.

The more precise schedule for the demonstration phase can be presented, but the whole process did not proceed in a clear or linear manner. The main work for the implementation of the BI system was done in six months, and the emphasis of each month is represented below.

- Month 1: Data sources
- Month 2: Data sources
- Month 3: Data sources, ETL
- Month 4: Data sources, ETL, Data warehousing
- Month 5: Data warehousing, Front-end software
- Month 6: Front-end software, users and training

As seen, some of the parts of the framework were developed and demonstrated simultaneously Additionally, the approach for the progress of the project was not to complete one part before moving onto another but to maintain flexible and iterative practices. There was not a clear and simple moment when it could have been stated that one part of the project is now completely ready and needs no further development. The previous list of months solely explains which were the main points of emphasis for each month, but in practice, each part of the project and framework were discussed and developed over the entire timeframe. The further subsections go through the demonstration phase of each part of the framework and discuss the utility of it.

7.5.1 Data sources and ETL

Even though data sources were a problem to some extent in the project of the case organization, there was not much to be considered about them. The framework guides that the data sources should be selected and defined carefully. In the project, there was a new data type that had a very simple structure and format, but the problem was the source of it. Data was coming from various different departments, that were not connected to each other in any way, and the departments had different employees, working methods, capabilities to manage source data, and data collecting methods. The organization decided to maintain the old methods of source data management and did not attempt to develop them by utilizing the framework. However, the data structure was not defined before but doing so was necessary for creating a structure and format standard for the departments.

Moving on to the ETL part of the framework, which had three different design choices. The first choice, which concerns ETL tools, was fairly unnecessary for the case organization because the existing ETL tools of the IT provider could be used for this project as well. Even though the organization possessed the ETL tool, it had to be evaluated first before it was clear that it was usable in this project as well. The framework did provide questionnaires that helped with defining of the qualities and requirements of ETL tools, and that is why the ETL tool section was not completely unnecessary. Additionally, the choice between different ETL sequences was not particularly necessary for the case project. The main reasons for using a different sequence than the traditional ETL were scalability issues and data types with different levels of urgency. The amount or the urgency of the source data was not very high, and that is why the common ETL sequence was the simplest option for the organization.

The data transmission schedule choice was the most useful part of the ETL section in the framework. Even though the choice is not the most crucial for the functionality of the BI system, it proved to be a tough and discussed question for the project. The case organization debated all the three options for the schedule, which were push, pull, and scheduled. Each option had its own benefits and weaknesses, but the irregular updates of the source data turned out to be the deciding factor. The scheduled transmission was selected, and the choice was able to be done with the help of the framework.

Conclusion about the data sources and ETL would be that the framework provided some help with the topic, but the main problems related to the sources could be considered as being outside of the research scope. The main issue was finding ways to deliver and extract the data from the various departments that had very suboptimal systems and data management methods.

The simplicity of the source data and the existence of previous ETL tools made some of the design choices and considerations very trivial, but those parts of the framework were still kept unchanged. For example, the fact that the organization already had usable ETL tools before the BI system project began, is most likely a not common situation. Overall, the data source and ETL section of the framework did provide help, but the guidance it gave couldn't be considered as critical for the case project.

7.5.2 Design model and data warehouse

The correct design model was hard to define for the case project. There were many defined qualities and requirements of the project that applied to top-down strategy, but also many that applied to bottom-up strategy. The final decision for the design model turned out to be the top-down strategy for the following reasons. Starting with the data modelling and philosophy of the strategy. The top-down strategy is described to be data-driven and the data modelling is usually relational (Turban et al., 2014). Additionally, in this strategy, user access is very limited, and the primary audience is IT professionals. The last deciding factor was the desired data and information flow, which in the top-down strategy proceeds from the one large enterprise-wide data warehouse to the front-end.

Moving on to the data warehouse architecture, which was a relatively straightforward choice for the case organization. The selection of the top-down strategy directs the architecture choice more towards enterprise data warehouse (EDW), hub and spoke (HUB), and federated architectures (FED), although the architectures based on data marts were not completely ruled out just because of the design model. The organization felt that they did not have a need for business area-specific data marts, and the desired impact of the BI system and data warehouse was more organizational than departmental. With this description alone, the framework would recommend the EDW architecture. However, the organization had an issue with the new source data type, which was coming from multiple legacy systems. A reasonable or pleasing solution for replacing the old data source systems could not be found, and they were included in the system. The additional information on the characteristics and requirements for the data warehouse architecture changes the recommendation from EDW to FED. Naturally, the common architectures that are used in this study are only directive and the framework does not determine that an organization should select one architecture and implement it exactly like in the theory. The case organization used the information of the framework to define its requirements for the architecture and for understanding what kind of alternatives there were. The desired architecture that the case organization defined for themselves was close to the textbook example of federated architecture.

The framework provided help with the design model choice and with the architecture choice, and the information and guidance about both of the choices could be considered as sufficient. However, the connection between the choices did not receive clear justification. Defining and considering of the design model could have helped with the architecture choice but stating that it clearly does so would be exaggerated. On the other hand, mentioning about the connection between the two choices most likely can provide some utility, and it is not harmful or misleading. The framework already emphasizes that the two choices have a connection or relation, but it might not be very strong. Because of this, the connection is left unchanged to the framework, and possible further studies could examine the relation more comprehensively.

7.5.3 Software and users

As it was discovered in the theory, successful choices and the implementation of the front-end and users are very important to the functionality and effectiveness of BI systems. Starting with the software choice. The framework guides to make the decision based on the desired qualities and functionalities, provider, and cost of the software. The case organization started the choosing process by defining what they wanted from the software. Going into very specific details is not necessary, but in general, the organization wanted a software that was capable of exporting reports in various formats and had good abilities to visualize information. Likely most of the BI software would fulfil these requirements. Going to the last factors in the software choice, which were very important to the organization. Software by Microsoft had been used for years in the organization, and the users were experienced with its Office-tools. If the BI software by Microsoft would fulfil the required functionalities, the organization wanted to select it no matter the possible competitors. Microsoft Power BI did not excel in performing calculations to data but fortunately for the organization, it had good capabilities in visualization. The costs of the software were considered as acceptable, so the decision was made with ease.

Considering information about software, there was nothing that was lacking from the framework and the decision was able to be made. However, the case organization had a very clear vision of what they wanted, and they did not need much help with the choice. It remains unclear how much help the framework actually provided with the topic.

After the software choice, the next topic in the framework is user access. Even though the BI system of the case organization contained data about the whole organization, the front-end software and the system was managed and used by one department only. The idea was that the department of economics would manage and use the software and create reports, and the rest of the organization could view the reports if necessary. The framework mentioned two common reasons for limiting user access. The first one was to limit access so that incapable or unnecessary users would not access writing rights and the second reason was to limit access so that only specific users could view and edit sensitive information. As a public sector organization, the case organization strives for transparency and the source data had only values that were considered as non-sensitive. This is why limiting access based on the second reason was forsaken. However, the limitation based on the capabilities and necessity seemed useful. Most of the users of the department were very capable of accounting and performing various calculations in traditional software such as Microsoft Excel, but they lacked previous experience with any BI software. Also, the general information technology-related skills of the employees were lacking.

There were not many options in the consideration of the user access, because the organization had only a few capable users for the software, and that amount was considered sufficient. This led to the choice of creating and using the power user role. The few users who had previous experience of similar BI tools, or they were in general more capable of using technology, were chosen as the power users. These users were the ones to create spreadsheets and various reports, and the rest of the department had basic licenses that allowed viewing and changing variables in the reports. The limitation was viewed more as an asset than as a deprivation, because it allowed being more cost-efficient by using less expensive licenses and because there were only a few employees who were intended to be heavy users. Overall, the choices regarding user access were able to be done with the guidance of the framework and the help was considered as sufficient by the organization.

The last topic in the front-end software and users- section of the framework is training. For training, the framework recommended selecting training methods that fit the culture and way of working of the organization and then maintain and develop the training in the upcoming years. The presented training methods were examined, and classroom and one-on-one training were the ones that felt natural for the case organization. The power users that possessed sufficient capabilities were able to provide most of the training to other users, but it was not sufficient enough method alone. In addition to classroom and one-on-one training, online courses were utilized for supplementary training for users who needed it.

Information that was provided about the training plan and different training methods was considered sufficient enough about those topics, but there is one consideration that was not included in the framework. There are many employees of the organization, that use the software more or less, but only a few heavy users. Every user does not need the same level of training, and some users might not need it at all. Questions about "Who needs training?" and "How much training each user group should receive?" were brought up. Guidance on concerns like those was not included, and the training section of the framework should be updated and improved.

To conclude the users and training section of the framework was updated (Figure 9) in order to fully answer the requirements of the case project. This was considered as the only major flaw in the framework, and with the improvement, it fulfils the requirements and needs of the case organization.

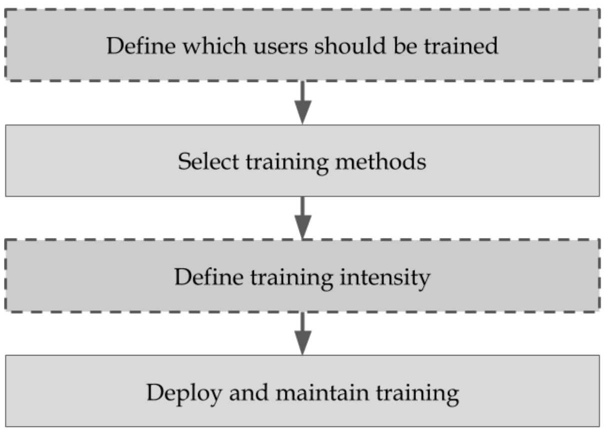

FIGURE 9 Updated framework: training

7.5.4 Public sector specific factors

One very important factor concerning the results of this study is that the case organization operates on the public sector. This must be noted because the demonstration phase of the study was done in the case organization and the results could vary if the test environment was the private sector.

The first hypothesis about the effect of the public sector was that it would limit or restrict the study in some way. However, the actual effect seemed opposite to the hypothesis. Some activities and design choices were most likely easier to do in the public sector. This is mainly because of transparency. The data of the case organization was not sensitive in any means, and most of it is reported publicly. The case BI system exports information primarily to internal use, but also for citizens (G2C), other governments (G2G), and businesses (G2B). Especially reports made for G2C and G2G are available and viewed by a large audience. Striving for transparency does not mean that the security of the data in BI systems would be neglected, but it could allow public sector organizations to have fewer considerations and less rigorous overseeing of the security. Again, data security and privacy considerations could have a stronger emphasis on the private sector. If this is true, it could be possible that security-related discussion and considerations are lacking from the framework. This is a possible limitation of the framework and the subject should be further studied. Overall, the business sector of the case organization was affecting the study unexpectedly little, but the topic requires further research.

7.5.5 Conclusion of demonstration

Generally, the framework was able to find the major components of BI system projects using previous studies and knowledge. The case project and the framework dealt with mostly the same design choices in the big picture, but the emphasis of these choices was usually different. For example, in theory data, warehouses are usually considered the most important and labour-heavy part of BI systems, and this is why the study provided a comprehensive amount of information about DW design choices. However, in this case project, the organization was able to find an easy way out with data warehouse implementation, and the information provided by the framework was actually greater than what was needed. As said, most parts of the framework had an adequate amount of information in them, but in few cases, the amount of help required and provided did not meet. The training was the only clear case where the organization required more information that was provided, and it was updated accordingly to an up to the mark level. Overall, the demonstration phase took very long and it was done over months while the case project was progressing. The final conclusion about the utility of the framework will be presented in the evaluation subsection next.

7.6 Evaluation

The more specific points of deficiency were addressed in the demonstration phase, but the more overall conclusion of the evaluation will be presented in this subsection. The evaluation phase of this study could be thought of as practical validation for the artifact. Some of the evaluation was already done in the demonstration phase because while the framework was tested in action, it was also analysed. The two phases were more or less fused together in this study.

As mentioned earlier, the utility of the framework was tested and evaluated by examining if it includes the major considerations of BI design and implementation and if it can provide help for them. The evaluation was done in open discussion in the organization, in which the main points of the framework were addressed. This method could not be considered as very rigorous but having a flexible and in-depth look into different topics was considered as a more effective and natural way for this context for example compared to having a long list of planned questions. However, there were few important questions that had to be answered about each part of the framework. The questions are the following: "Is this section of the framework a major consideration for the BI project?", "Does the organization require help for said consideration?", and "Did the framework provide sufficient help with the said consideration?".

When comparing the most important considerations that were selected to the framework based on previous studies and what the actual practice was in the case organization, there were not any major differences. Most of the framework was useful and needed in decision making,but was there also information that could be considered as useful but not necessary for the project. For example, information about ETL tools did not provide much utility for the project, because the existing tools that the organization already had were usable also for this project. The logic of the evaluation was to find out if there are any major issues in the project that the framework does not address and provide sufficient help to. The decisions about the sufficiency were done in the demonstration phase, in which each part of the framework was given a binary grade concerning the sufficiency. This means that the parts of the framework were ruled either as sufficient or not sufficient enough. This was considered up to the mark method for evaluation, but it leaves a few open questions about the utility of the framework. For example, it is known that the framework provided some help to the organization, but it remains somewhat unclear how helpful it was in general.

	How critical it was for the project	How much help was needed	How much help was received
Data sources	5	2	3
ETL	4	2	4
Design model	1	1	3
Data warehouse	5	3	5
Front-end software	4	3	3
User access	2	2	2
Training	3	3	2

TABLE 8 Framework evaluation

The ad-hoc evaluation that was used was considered to be a poor but sufficient method. This is why a table (Table 8) was constructed to give the evaluation a more scientific approach and more depth in general. Most of the contents of the table have been already mentioned in the demonstration and evaluation sections, but this type of conclusion was necessary to process and present the knowledge in an efficient way. The table demonstrates three different factors for each part of the framework. The first one describes how critical the part is considered to be for the project, the second one presents how much help the organization required, and the third factor evaluates how well the framework was able to provide guidance concerning that part. The evaluation was done by grading each part from 0 to 5, where 0 means not sufficient enough or not important, and 5 means excellent or very important. Additionally, if the grades for both help required and help received factors were exactly the same, it would indicate that the level of guidance and information was adequate.

As it can be perceived, some of the parts had higher grades in the help received than the help required. This is because in some decisions and considerations the organization did not need much assistance, and the framework provided more information that was required. The only case where the help required was graded higher than the help received was training. This means that the organization received almost sufficient enough help from the framework, but not quite. An important note considering this deficiency was that this round of evaluation was done to the first version of the framework. The shortcomings in the training section of the framework were discovered in the demonstration phase, and it was edited accordingly. In the end, the additional and more in-depth round of evaluation did not provide much new knowledge, but it presented the whole situation from an additional perspective and summarized information more efficiently. The demonstration and evaluation subsections focused on the utility of the framework, but the next main section will examine the framework and the study in a more critical fashion and try to discover and present what was learned from this study overall.

8 DISCUSSION AND CONCLUSIONS

This section will examine the framework and the study in a critical fashion to understand what was learnt from it and what are the limitations of it. Firstly, the conclusion and contribution subsections discuss the new knowledge of the study and what kind of contribution it offers to the research area. Then the credibility and various limitations of the study are addressed. Lastly, further research topics are discussed briefly.

8.1 Conclusions

The objectives of this study were building of a comprehensive theoretical knowledge base, finding the major matters of BI system implementation projects, building a framework that can be utilized in BI projects, and finally helping and solving the problems of the case organization while inspecting possible public sector-specific factors that might affect the case project. All of the objectives were reached and fulfilled in a satisfying manner. In conclusion, the theoretical background knowledge was relatively comprehensive, and issues and design choices of the case project were solved. A beta version of the BI system was implemented while the study was conducted, and even though the system is not in a final version yet, it is safe to say that at least the minimum requirements of it are fulfilled.

The most basic things that were learned from this study were the characteristics and technologies of BI systems and BI projects. The study found the most important design choices and various considerations that organizations might face in their own BI projects, and some arguments and guidance was found and provided to help with these questions. Additionally, a generic example of a BI system was constructed and analysed in the first sections. This type of general knowledge of BI is useful but nothing new. The most interesting discoveries of the study were about the relationship of theory versus practice, and the public sector versus the private sector.

It was perceivable that in this case project the operating sector of the organization was considered to affect very marginally to the BI system design and implementation. As previously mentioned, the initial hypothesis was that public sector organizations would have to design their BI systems in a sector-specific manner. However, the case project followed very similar steps compared to the theoretical knowledge that was either sector-neutral or aimed more towards the private sector. Public sector organizations naturally have their own characteristics, that were examined in the second section, but in this case, the sector did not affect much to the BI system implementation. It must be noted that this was only one case system and the making of such statements requires testing in more case organizations and stronger proof in general to be ultimately true.

The credibility and critique of this study are addressed later in this section, but there is one important matter related that should be presented in this subsection. Likely the most major risk and critique for the credibility of this study would be a possible unintended bias. The author of the study knew the case organization from the past and had some idea of how the project might proceed and end up before writing the thesis. The main objectives of the study were to construct a framework solely based on theory, and then to improve and update it based on practice. As the evaluation and results imply, the first version of the framework was unexpectedly complete and close to the needs of a real business environment. There are at least four possible explanations for this. The first one is that the theoretical knowledge of previous studies would be comprehensive and well generalized that most of the BI projects would have the same considerations. The second reason could be that the BI projects in general are very similar. The third reason would be that it was coincidental, and the fourth possible explanation is the possibility that the first version of the framework wouldn't be purely based on theory, but that the author could have unintentionally known what to search from the theory based on the previous knowledge of the case organization. If that would be true, it fortunately does not undermine or worsen the framework in any way, and actually, it could only have a positive effect on the quality of it. However, this possibility must be addressed when making claims about the connection between the theoretical knowledge and practice.

8.2 Contribution

The knowledge in this study and the general contribution of it to the research field could not be considered as revolutionary, but it still offers something new and useful. Firstly, the theoretical knowledge that was accumulated for this study could be considered as comprehensive, although some of the topics were discussed only briefly. This study refers to and uses knowledge from a vast amount of studies. The theoretical part of the study offers a comprehensive summary of the research topic in a relatively understandable form. This study accumulated information about business intelligence, BI system technologies, BI architecture and design, and BI projects and users. One of the strengths was that the scope of the study was very large, and this is why it was possible to utilize background knowledge from various topics. Naturally, the large scope has some negative sides to it as well, but overall it was considered as a benefit.

Business intelligence implementation and design have been studied prior to this study, but the easiest way to justify this study would be the fact that having another research with a case organization provides the research field more data from a real business environment. Every case organization or project is different, and hence another case study is never useless. Even though the framework of this study is not a completely unique concept, it still contains some useful knowledge and new perspectives.

The study offers new knowledge in a few different ways. The first reason is the already mentioned case project, which is naturally a unique and uncharted testing environment and can be considered as new knowledge. Also, the framework that has been constructed in this study is not totally unique, but it could still be considered as a new creation that is the production of this study. Additionally, the study provides evidence and discussion for few less studied views. For example, the signification and importance of the operating sector of an organization in BI systems is a quite unique topic. Also, the discussion about the relation between theoretical knowledge and actual business environment is something that adds new perspective to this study.

8.3 Credibility

A high level of credibility, validity, and reliability has been one of the major objectives in the conduction of this study. One of the main activities that were repeated in order to avoid false claims and conclusions was continuous self- and peer-critique towards any statement and choice of the study. Even though the study aims to find and create new knowledge, the credibility of the study was naturally considered as the most important quality. Making of any strong claims was contemplated long before they were added into the final version.

The validity of different topics and research methods that were selected is by no means problematic and could be considered as sufficient at least. However, the study acknowledges that there is room for improvement. For example, the demonstration and evaluation methods were sufficient enough to make conclusions about the framework, but there were other possible methods that could have provided stronger evidence but were not reasonable to execute for different reasons.

In this type of research, the concern for reliability was not as major as it would be with traditional quantitative or qualitative research methods, but it still should be addressed. The main consideration of reliability in this study would be that each part of the framework was demonstrated and evaluated consistently. The major concepts of the framework were not equal in every way, for example, data warehouses received more attention and examination than data sources. However, every concept and part was still demonstrated and evaluated by using the methods and questions to address the reliability of the study.

The overall credibility and the reasons why the results of this study should be believed has been acknowledged throughout the study. Any claims or statements that were even marginally unsure, were critiqued and questioned in the main sections. Additionally, the most major risk to the credibility of the study was the fact that it had only one test case. It has been mentioned and taken into account that the framework worked in one case organization, but to make any strong claims about the utility of the framework, it should be tested in additional organizations.

8.4 Critique and limitations

The case organization was primarily satisfied with the framework and the study, but one major critique was brought forward. The framework did provide guidance with the design choices, but more contact with practicality was wished for. It remains somewhat unclear what type of improvements were wanted, but this was interpreted as a requirement for more practical and specific dictation. On one hand, as a scientific publication and a master's thesis, the study should maintain a certain level of theoreticality. On the other hand, this also is a design science research, which aims to find solutions for real-life problems. If the problem requires a higher level of practicality, it could be justified to do so. The issue was resolved by going through the main design choices with the organization off the record of the study.

A few additional limitations and critique were found after the main content of the study was ready. Testing the framework in an organization is definitely helpful and highly beneficial for the quality of the study, but to perfect the framework, it should be tested with multiple organizations. As earlier mentioned, if something major comes up in the project that is not in the framework, it would be added to the framework. The implementation project of the case organization most likely included many of the common considerations that BI projects have, but to remain critical, it should be noted that something could be missing.

Another limiting factor in the study was that the case organization did have some previous experience and existing solutions for performing business intelligence. An optimal context to test the framework would be an organization, that had only a marginal experience of BI and would start from scratch. This is because an organization with less experience would require more external help than an experienced organization. Even though the case organization had some previous knowledge of BI, it could not be considered as very experienced with the subject. Nonetheless, the case organization was a good fit for the study.

The last notion about the critique would be the scope of it. The study attempted to include a very large variety of topics, which has its own strengths and weaknesses. Naturally, having a large knowledge base in one place could be considered a good thing, but including many topics and a big spectrum of information might make the study hard to comprehend at first. Additionally, many of the topics included in this study were only scratched from the surface.

8.5 Follow-up research topics

As said, the framework was initially built by using theoretical knowledge from previous studies and it was tested in a public sector organization. The conclusion in this matter was that BI systems do not differ that much between the two sectors. However, to fully confirm the result, a potential idea for future studies would be to conduct a similar case study in a private sector organization. Additionally, the utility of the framework was concluded to be sufficient enough to help organizations in their implementation projects, but in order to justify this statement properly, the framework should be tested in multiple projects.

Additionally, there are multiple potential follow up research questions and topics concerning business intelligence systems. For example, how BI technologies and tools differ in the private and public sectors, and how frameworks or various guidebooks differ when they are targeted for a specific sector. And again, the framework would benefit from multiple rounds of demonstrations in different organizations. Though the quality of it would be considered sufficient, there is room for improvement.

9 SUMMARY

This is the last section of the study. It summarizes and explains the most important details of this study.

This study examined business intelligence and BI systems from various angles. The main objective was to create a framework that includes information about BI implementation projects and BI system design. Additionally, there were multiple objectives that had to be achieved in order to reach the main one. These objectives were (1) building of a comprehensive theoretical knowledge base, which explains business intelligence, BI systems and its components, and BI implementation projects, (2) finding solutions for the case implementation project, (3) answering the research questions, which were following.

- What are the most important design choices and considerations of BI system implementation projects?
- How to approach these design choices and considerations?
- Does a generalized framework that has been built for common BI projects work for a public sector organization?

Briefly, the approach for achieving the major objectives was the following. Finding of theoretical knowledge, creating a framework based on the knowledge, testing the framework in a real business environment, and finally evaluating the framework and discussing conclusions. The study followed design science guidelines in the conduction of it.

The building of the theoretical knowledge base began by finding general knowledge about business intelligence for understanding the basics of the topic and for setting up the rest of the study. The second section summarized very compactly as follows. The background of BI extends to DSS and MIS from the 1950's to today. The definition of BI can be expressed in various ways, but to quote Wixon and Watson (2010), "Business intelligence is a broad category of technologies, applications, and processes for gathering, storing, accessing, and analysing data to help its users make better decisions.". Utilizing BI efficiently can provide various benefits to an organization, for example, cost and time savings, better information and decision making, and improvement of business processes (Wixon & Watson, 2010). Additionally, the differences between BI in the public and private sectors were inspected. Multiple differences were found, which generated a hypothesis that BI implementation projects in public sector organizations would have some sector-specific concerns. However, later in the study, this hypothesis was questioned, because there were no major differences from a technical point of view.

After the base knowledge about business intelligence was presented, BI systems and different technologies associated with them were studied. The objective of the third section was to find and describe a common BI architecture and the components of BI systems. A BI system was divided into five parts, which were data source, ETL-layer, data warehouse, front-end software, and users. Data sources are connected and processed through ETL into the data warehouse, from which the data enters analytical software, which again is processed into knowledge by users (Jun et al., 2009). Especially important points in BI systems and technologies were various options and decisions about how to implement them. In the fifth section, various design and implementation choices for each of the BI system components were examined. Additionally, the most important considerations and success factors of them were attempted to find, because they were vital for the framework. Before moving on to the framework, more theory building was needed.

Information about the technological side of alone BI was not considered as enough for helping with the case project and for understanding the implementation side. This is why an additional round of finding theoretical knowledge was done in order to examine the characteristic of BI projects. For example, common activities, implementation design models, required resources, requirements, and critical success factors were examined. In general, BI projects are labour-heavy projects that require a lot of skills and other resources, but most importantly a lot of planning. Defining requirements and making implementation choices about different parts of the system could be considered as the main activities in BI system implementation (Chaudhuri & Dayal., 1997; Reinschmidt & Francoise, 2000). Additionally, two common approaches for BI systems in the form of design models were presented. The models were top-down and bottom-up strategies. The strategies vary for example in architecture, focus, and scope.

When the knowledge base was considered to be comprehensive enough, a framework was defined, designed, developed, demonstrated, and evaluated with the theoretical background knowledge and with the help of the DSRM process model (Peffers et al., 2007). The purpose or main objective of the framework was to provide a comprehensive information package of BI systems and how to implement and make various considering the topic in a big picture. A large amount of information from the theoretical part of the study was compressed into the framework, which presented the information in a directive manner. After the first version was done, the utility of it was tested in a real business environment. In other words, the framework was tested in a case BI implementation project to realize if it was able to provide help for every major problem of the case organization. The utility of the framework was considered sufficient or good in most cases, but minor updates were made. Inspecting the quality of the initial version of the framework leads to few conclusions.

The most trivial and basic realizations were about the structure, technologies, and design of BI systems. Additionally, a few more interesting perceptions were made. Firstly, it was perceived that the theory about BI systems was very close to the actual business environment. The reasoning for this is most likely the lack of variety in BI systems, but it could also result from coincidence, bias, or because the earlier studies were very comprehensive. Additionally, the original hypothesis about differences between BI in private and public sector organizations was strongly questioned. Both sectors clearly have different characteristics and points of emphasis, but it seems that the actual implementation, design, and technologies do not vary much between the sectors. However, the conclusions and any statements were made very carefully, because there was only one case organization. Even though having only one case organization to test the framework was the major limitation of this study, there was a positive side to it. The demonstration and evaluation phase were made over months in the case project and being focused on one organization made in-depth observing possible. The study provided some needed insight for BI system design in general, but more importantly for public sector BI implementation projects. A possible and recommended follow-up topic for this study would be testing the framework in additional organizations and comparing the differences between the two sectors. This would enable stronger proof for this research topic in general, and in the process, it would allow the making of stronger statements and conclusions.

REFERENCES

Airinei, D., & Homocianu, D. (2009). DSS vs. business intelligence. Revista Economica.

Amanpartap, S., & Khaira, J. S. (2013). A comparative review of extraction, transformation and loading tools. Database Systems Journal BOARD, 42.

Amara, Y., Solberg Søilen, K., & Vriens, D. J. (2012). Using the SSAV model to evaluate Business Intelligence Software.

Ariyachandra, T., & Watson, H. J. (2005). Key factors in selecting a Data Warehouse architecture. Business Intelligence Journal, 10(2), 19-26.

Ariyachandra, T., & Watson, H. J. (2008). Technical opinion Which data warehouse architecture is best?. Communications of the ACM, 51(10), 146-147.

Ariyachandra, T., & Watson, H. J. (2006). Which data warehouse architecture is most successful?. Business intelligence journal, 11(1), 4.

Ariyachandra, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. Decision support systems, 49(2), 200-212.

Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. Journal of information technology, 20(2), 67-87.

Arnott, D. (2008). Success factors for data warehouse and business intelligence systems. ACIS 2008 Proceedings, 16.

Azvine, B., Cui, Z., & Nauck, D. D. (2005). Towards real-time business intelligence. BT Technology Journal, 23(3), 214-225.

Backus, M. (2001). E-governance and developing countries, introduction and examples. International Institute for Communication and Development (IICD).

Bansal, S. K., & Kagemann, S. (2015). Integrating big data: A semantic extract-transform-load framework. Computer, 48(3), 42-50.

Blumberg, R., & Atre, S. (2003). The problem with unstructured data. Dm Review, 13(42-49), 62.

Bodislav, D. A. (2015). Transferring business intelligence and big data analysis from corporations to governments as a hybrid leading indicator. Theoretical and Applied Economics, 22(1), 257-264.

Boselli, R., Cesarini, M., & Mezzanzanica, M. (2011). Public service intelligence: evaluating how the public sector can exploit decision support systems. In XXI International RESER Conference.

Boyer, J., Frank, B., Green, B., Harris, T., & Van De Vanter, K. (2010). Business intelligence strategy. IBM Corporation.

Boynton, A. C., & Zmud, R. W. (1984). An assessment of critical success factors. Sloan management review, 25(4), 17-27.

Breslin, M. (2004). Data warehousing battle of the giants. Business intelligence journal, 7, 6-20.

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. ACM Sigmod record, 26(1), 65-74.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. Communications of the ACM, 54(8), 88-98.

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

Chenoweth, T., Corral, K., & Demirkan, H. (2006). Seven key interventions for data warehouse success. Communications of the ACM, 49(1), 114-119.

Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. ACM Transactions on Management Information Systems (TMIS), 3(3), 1-13.

Coman, M. (2009). Business Intelligence and E-governance. LESIJ-Lex ET Scientia International Journal, 16(1), 484-491.

Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). Data integration flows for business intelligence. In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (pp. 1-11). Acm.

Debortoli, S., Müller, O., & vom Brocke, J. (2014). Comparing business intelligence and big data skills. Business & Information Systems Engineering, 6(5), 289-300.

Elena, C. (2011). Business intelligence. Journal of Knowledge Management, Economics and Information Technology, 1(2), 1-12.

Hannula, M., & Pirttimaki, V. (2003). Business intelligence empirical study on the top 50 Finnish companies. Journal of American Academy of Business, 2(2), 593-599.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS quarterly, 75-105.

Howson, C. (2006). Seven pillars of BI success: BI tools may be getting better, but technology is only part of the story. Intelligent Enterprise-San Mateo-, 9(9), 33.

Howson, C. (2007). Successful Business Intelligence: Secrets to Making BI a Killer App.

Hostmann, B., Herschel, G., & Rayner, N. (2007). The evolution of business intelligence: The four worlds. Retrieved from Gartner database, 122-141.

Hüsemann, B., Lechtenbörger, J., & Vossen, G. (2000). Conceptual data warehouse design (pp. 6-1). Universität Münster. Angewandte Mathematik und Informatik.

Gangadharan, G. R., & Swami, S. N. (2004). Business intelligence systems: design and implementation strategies. In 26th International Conference on Information Technology Interfaces, 2004. (pp. 139-144). IEEE.

Gray, P., & Watson, H. J. (1998). Present and future directions in data warehousing. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 29(3), 83-90.

Grönlund, Å., & Horan, T. A. (2005). Introducing e-gov: history, definitions, and issues. Communications of the association for information systems, 15(1), 39.

Inmon, W. H. (1992). Building the Data Warehouse. 1992. QED Information Sciences, Wellesley, MA. Inmon, W.H. (1999) Building the Operational Data Store. Wiley Publishers- New York, 2nd edition.

Inmon, W. H., Imhoff, C., & Sousa, R. (2002). Corporate information factory. John Wiley & Sons.

Jun, T., Kai, C., Yu, F., & Gang, T. (2009). The research & application of ETL tool in business intelligence project. In 2009 International Forum on Information Technology and Applications (Vol. 2, pp. 620-623). IEEE.

Langseth, J., & Vivatrat, N. (2003). Why proactive business intelligence is a hallmark of the real-time enterprise: outward bound. Intelligent enterprise, 5(18), 34-41.

Lenz, H. J., & Thalheim, B. (2009). A Formal Framework of Aggregation for the OLAP-OLTP Model. J. UCS, 15(1), 273-303.

Lönnqvist, A., & Pirttimäki, V. (2006). The measurement of business intelligence. Information systems management, 23(1), 32.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. Decision support systems, 15(4), 251-266.

Moody, D. L., & Kortink, M. A. (2000). From enterprise models to dimensional models: a methodology for data warehouse and data mart design. In DMDW (p. 5).

Moss, L. T., & Atre, S. (2003). Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Professional.

Negash, S., & Gray, P. (2008). Business intelligence. In Handbook on decision support systems 2 (pp. 175-193). Springer, Berlin, Heidelberg.

Nutt, P. C. (2005). Comparing public and private sector decision-making practices. Journal of Public Administration Research and Theory, 16(2), 289-318.

Olszak, C. M., & Ziemba, E. (2007). Approach to building and implementing business intelligence systems. Interdisciplinary Journal of Information, Knowledge, and Management, 2(1), 135-148.

Olszak, C. M., & Ziemba, E. (2012). Critical success factors for implementing business intelligence systems in small and medium enterprises on the example of upper Silesia, Poland. Interdisciplinary Journal of Information, Knowledge, and Management, 7(2), 129-150.

Ong, I. L., Siew, P. H., & Wong, S. F. (2011). A five-layered business intelligence architecture. Communications of the IBIMA.

Palvia, S. C. J., & Sharma, S. S. (2007). E-government and e-governance: definitions/domain framework and status around the world. In International Conference on E-governance (No. 5, pp. 1-12).

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of management information systems, 24(3), 45-77.

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. uropean Journal of Information Systems, 27 (2), 129-139.

Popovič, A., Coelho, P. S., & Jaklič, J. (2009). The impact of business intelligence system maturity on information quality. Information research, 14(4).

Power, D. J. (2007). A brief history of decision support systems. DSSResources. com, 3.

Power, D. J. (2008). Understanding data-driven decision support systems. Information Systems Management, 25(2), 149-154.

Ranjan J. (2009). Business Intelligence: Concepts, Components, Techniques and Benefits. Journal of Theoretical and Applied Information Technology 9(1), 60–70.

Reinschmidt, J., & Francoise, A. (2000). Business intelligence certification guide. IBM International Technical Support Organisation.

Rudin, K. and D. Cressy (2003) "Will the Real Analytic Application Please Stand Up?" DM Review, (13)3, pp. 30-34.

Schewe, K. D., & Thalheim, B. (1993). Fundamental concepts of object oriented databases. Acta Cybernetica, 11(1-2), 49-83.

Sen, A., & Sinha, A. P. (2005). A comparison of data warehousing methodologies. Communications of the ACM, 48(3), 79-84.

Shin, B. (2003). An exploratory investigation of system success factors in data warehousing. Journal of the Association for Information Systems, 4(1), 6.

Sprague, R. H., & Watson, H. J. (1975). MIS concepts. Journal of Systems Management, 26(1), 34-37.

Trieu, V. H. (2017). Getting value from Business Intelligence systems: A review and research agenda. Decision Support Systems, 93, 111-124.

Turban, E., Sharda, R., & Delen, D. (2014). Business intelligence and analytics: systems for decision support. Pearson Higher Ed.

Tutunea, M. F., & Rus, R. V. (2012). Business intelligence solutions for SME's. Procedia economics and finance, 3, 865-870.

Vassiliadis, P., & Simitsis, A. (2008). Near real time ETL. Springer Annals of Information Systems, 3, 19-29.

Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. In International Conference on Design Science Research in Information Systems (pp. 423-438). Springer, Berlin, Heidelberg.

Watson, H. J. (2009). Tutorial: Business intelligence-Past, present, and future. Communications of the Association for Information Systems, 25(1), 39.

Watson, H. J., Rainer Jr, R. K., & Koh, C. E. (1991). Executive information systems: a framework for development and a survey of current practices. MIS quarterly, 13-30.

Wixom, B., & Watson, H. (2010). The BI-based organization. International Journal of Business Intelligence Research (IJBIR), 1(1), 13-28.

Wu, L., Barash, G., & Bartolini, C. (2007). A service-oriented architecture for business intelligence. In IEEE international conference on service-oriented computing and applications (SOCA'07) (pp. 279-285). IEEE.

Xu, H., & Hwang, M. I. (2007). The effect of implementation factors on data warehousing success: An exploratory study. Journal of Information, Information Technology, and Organizations, 2, 1.

Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. Journal of computer information systems, 50(3), 23-32.