Net Present Value of Investments in Data Warehousing and Differences Across Firm Sizes

Master Thesis I:

The shift in data organization: Uncovering influences on the Net Present Value of investment in data warehousing and differences across firm sizes

Submitted to Università Commerciale Luigi Bocconi (Milano, Italy) in October 2023
Defended at Università Commerciale Luigi Bocconi (Milano, Italy) in December 2023
Thesis Supervision: Prof. Aldona Kapačinskaitė

Master Thesis II:

Case Study: Evaluating the Net Present Value of a Data Warehouse at Transalb, a Small German Trucking Company

Submitted to Stockholm School of Economics (Stockholm, Sweden) in November 2023
Thesis Supervision: Prof. Bo Becker

Ferdinand Netsch

Master of Science Thesis in Finance, Double Degree
Stockholm School of Economics & Università Commerciale Luigi Bocconi
November 2023

Disclaimer

Both theses have been written and submitted in accordance with the framework of the Double Degree Program in Finance between Stockholm School of Economics (Home School) and Università Commerciale Luigi Bocconi (Host School). Both theses can be read independently.

The former thesis was submitted as an independent piece of work in June 2021 to Università Commerciale Luigi Bocconi and was defended ibidem in July 2021. The thesis is in line with both universities standards and regulations and constitutes 18 ECTS.

The latter thesis, constituting 12 ECTS, was written in order to fulfil the additional requirements at Stockholm School of Economics and was submitted in September 2020.

Master Thesis I

The shift in data organization: Uncovering influences on the Net Present Value of investment in data warehousing and differences across firm sizes

Ferdinand Netsch
Matricola: 3153697
Master's Thesis (M.Sc. Finance)

Abstract

In today's digital era, the competitiveness of businesses may depend on their ability to make use of data. Central to this is the establishment of a data warehouse, a unified repository that consolidates data from various sources. This study delves into the strategic and financial implications, with a primary focus on understanding how a firm's size influences the perceived Net Present Value (NPV) of such investment. The research reveals that such decision is not just a technological initiative but a strategic commitment that can shape a company's competitive landscape for years. Various factors are identified that influence the success of data warehouse implementations. These range from organizational elements like vision and culture to technical aspects such as source systems and development technology. Factors including company size, cloud service usage, outsourcing, and General Data Protection Regulation (GDPR) awareness, have been identified as increasingly relevant in today's business environment. Also, the choice of implementation approach, whether emphasizing an enterprise-wide view or starting with individual data marts, was found to have a profound impact on the perceived NPV. The applied qualitative research approach, involving two contrasting case studies of an SME and a global corporation, provides insights into real-world implications. The results underscore the importance of a clear vision in guiding data warehousing projects, the state of source systems, and firm culture as significant determinants of success. Furthermore, the study reveals that while company size doesn't necessarily hinder the adoption of data warehousing, it does influence the perceived NPV. Larger corporations face higher costs and increased risks. However, if managed correctly, they can derive exponential value from their data. In contrast, Small- and Medium Enterprises (SME) can utilize cloud solutions and outsourcing to reduce upfront costs and implementation challenges. Testing an implementation with a proof-of-concept can reveal organization-specific implementation issues at an early stage.

Università Bocconi
MILANO

Università Commerciale Luigi Bocconi
Department of Finance
October 2023

Contents

Abbreviations

AI Artificial Intelligence. 9, 23, 27, 49, 62, 64

BI Business Intelligence. 7, 9, 12–15, 18, 20–23, 29

CAPM Capital Asset Pricing Model. 37

CSF Critical Success Factor. 7, 13, 14, 18, 49, 64

GDPR General Data Protection Regulation. 7, 26, 27, 30, 38, 42, 60–64, 66

ML Machine Learning. 9, 23

NPV Net Present Value. 7, 8, 22, 24–28, 39–41, 43, 44, 47, 50–52, 54, 60, 62–65, 67, 68

SME Small and medium-sized enterprises. 17, 23, 24, 26, 47, 51, 53–57, 63, 65–67

SWOT Strengths, Weaknesses, Opportunities, and Threats. 65, 66

TDWI The Data Warehousing Institute. 23

TV Terminal Value. 39, 40

1 Introduction

In the digital era, the competitive landscape is increasingly shaped by the ability of companies to harness the power of data. Central to this capability is the establishment of a robust data foundation, often realized through the creation of a data warehouse. This repository consolidates data from diverse sources, ensuring a unified and consistent view of information, as opposed to fragmented insights from disparate sources.

The literature review section delves into the strategic implications of data warehouse implementation, with a particular focus on the Net Present Value (NPV) as a pivotal measure for informed decision-making. The seminal works of Inmon (2005) and Kimball et al. (2008) are explored, shedding light on the contrasting methodologies for building data warehouses. The strategic commitment associated with data warehousing is further elucidated, drawing on insights from the study of Ghemawat (1991) and the research of Cassiman, Ricart, and Valentini (2022). These commitments, while promising competitive advantage, also pose challenges in terms of flexibility and adaptability in the fast-evolving digital landscape.

The study then transitions to an in-depth examination of the factors influencing the success of data warehouse and Business Intelligence (BI) implementations. Foundational research by Wixom and Watson (2001) and Yeoh and Koronios (2010) among others offers insights into the Critical Success Factor (CSF) from both technical and organizational perspectives. The research landscape is diverse, with studies employing various methodologies, from regression analysis to interviews, to identify and validate these CSFs. The study further delves into emerging factors, such as company size, implementation approach, cloud service usage, extent of outsourcing, and General Data Protection Regulation (GDPR) awareness, some of which have become increasingly relevant in recent years.

A crucial segment of the literature review is dedicated to dissecting the NPV constituents, which include implementation costs, ongoing costs, expected benefits, discount factor, and time horizon. This comprehensive understanding of NPV components will later facilitate a nuanced analysis of how different factors influence these components, ultimately impacting the overall NPV of data warehousing projects.

The methodology section outlines the structured approach adopted for the empirical part of the study. Rooted in Eisenhardt (1989)'s case study approach, this segment aims to discern the factors influencing the NPV of data warehousing investments. The research design, data collection methods, and data analysis techniques are meticulously detailed, ensuring a robust and empirically grounded theory.

The discussion of results offers a comparative analysis of two distinct cases, representing both an SME in the healthcare sector and a global insurance company. These cases, with their unique characteristics and challenges, provide multifaceted insights into the real-world implications of data warehousing decisions on the NPV.

2 Literature Review

In this part of the study, a common understanding of the NPV and data warehousing in general will be established. Then, the topic is looked at from a strategic lens to see what the implementation means for the strategy of the business conducting it. Subsequently, the study will dive into the factors given by the literature that were identified as influential on the successful outcome of such implementation. Lastly, the constituents of the NPV on the specific case of a data warehouse implementation project are uncovered.

2.1 Strategic Implications of a Data Warehouse Implementation

2.1.1 The NPV as a Decisive Measure

The NPV is a cornerstone in financial decision-making, offering a lens through which the potential profitability of investments can be assessed. Defined as "the value of all future cash flows (positive and negative) over the entire life of an investment discounted to the present" (TheCorporateFinanceInstitute 2018), the NPV provides a snapshot of an investment's worth in today's terms, taking into account all anticipated future inflows and outflows.

The essence of NPV lies in its ability to incorporate the time value of money, acknowledging that a dollar today holds in general, with exceptions, more value than one in the future. A positive NPV suggests a potentially profitable investment, while a negative one indicates the opposite.

Compared to other appraisal methods, the NPV's comprehensive approach stands out. It considers all future cash flows over an investment's lifespan, ensuring a holistic view of its potential value. The constituents of the NPV, which form the backbone of this calculation, will be explored in detail in the subsection 2.3 NPV Constituents.

2.1.2 Definition of Data Warehousing

IBM (2023) defines data warehousing as ”a system that aggregates data from different sources into a single, central, consistent data store to support data analysis, data mining, Artificial Intelligence (AI), and Machine Learning (ML). A data warehouse system enables an organization to run powerful analytics on huge volumes [...] of historical data in ways that a standard database cannot. Data warehousing systems have been a part of BI solutions for over three decades” (IBM 2023).

Data warehouses play a pivotal role in modern business operations, facilitating organizations in the efficient storage, retrieval, and analysis of vast amounts of data. From a business perspective, data warehouses are often termed as Decision Support Systems, highlighting their role in facilitating informed decision-making (Westerman 2001). Their prominence in the corporate realm is evident, with an impressive 95% of Fortune 1000 companies either having a data warehouse or being in the process of developing one as of 2001 (Wixom and Watson 2001).

2.1.3 Data Warehousing as a Form of Strategic Commitment

Pankaj Ghemawat (1991)'s perspective on commitment offers a profound understanding of the strategic decisions companies make in their pursuit of competitive advantage. At the heart of Ghemawat (1991)'s argument is the notion that for a company to establish a competitive edge, it must be willing to commit to developing capabilities that are not only superior to those of its competitors but also sustainable over time. This commitment, however, is not without its challenges and implications.

Ghemawat (1991, p.22) defines commitment as ”the tendency of strategies to persist over time”. This definition underscores the enduring nature of strategic decisions and the long-term implications they carry. When a firm commits to a particular strategy, it is essentially making a promise to maintain that strategy over an extended period, even in the face of changing market dynamics or unforeseen challenges. Such a commitment is not merely a declaration of intent; it is a binding pledge that shapes the firm's future actions and decisions.

The decision to commit, as Ghemawat (1991) points out, is fraught with complexities. Commitments, by their very nature, tend to be resource-intensive and challenging to reverse. This is particularly true in the context of data warehousing, where the allocation of significant resources is a prerequisite. The establishment of a data warehouse is not a one-off project but a long-term strategic move that demands continuous investment in terms of finances, human capital, and technological infrastructure - among other factors. Thus, the decision to invest in data warehousing is emblematic of a firm's commitment to harnessing the power of data for competitive advantage.

Ghemawat (1991) identifies four primary causes of commitment: lock-in, lock-out, lags, and inertia. Each of these factors plays a crucial role in shaping the nature and extent of a firm's commitment to a particular strategy.

Lock-In Lock-in, as conceptualized by Ghemawat (1991), is a multifaceted phenomenon that plays a pivotal role in the persistence of strategic commitments. It is rooted in the presence of "sticky" factors that not only influence historical decisions but also shape the trajectory of future strategic moves. Ghemawat (1991) identifies three key characteristics that these factors must possess to induce lock-in: durability, specialization, and untradeability.

- **Durability**: Central to the concept of lock-in is the inherent durability of certain strategic factors. Durability implies that a factor retains its relevance and utility over prolonged periods, continually influencing the strategic trajectory of a firm. Within the realm of data warehousing, durability manifests in the sustained value the warehouse delivers. By consistently providing decision-makers with timely and relevant information, a durable data warehouse becomes an indispensable asset.
- **Specialization**: The second characteristic, specialization, implies that these factors are tailored to support specific strategies. If a factor were generic and applicable to all strategies, it would not create any form of lock-in. Specialized factors, by their very nature, are best suited to specific strategic paths, making deviation from those paths less optimal. For instance, a data warehouse designed to cater to a particular industry's needs or a specific business model would be a specialized asset, making it challenging for the firm to pivot to a radically different strategy without significant reconfiguration.
- **Untradability**: While durability and specialization matter, they don't induce lock-in if the asset can be easily traded. In the context of data warehousing, the high degree of customization makes this evident. Data warehouses are tailored to an organization's specific systems and processes, making them inherently non-transferable. Such bespoke solutions, once integrated, anchor the organization to its strategic path, emphasizing the lock-in effect.

Lock-out Lock-out, as described by Ghemawat (1991), is the counterpart to lock-in. While lock-in focuses on the challenges associated with disposing of sticky factors, lock-out highlights the hurdles a company faces when trying to initially acquire or repurpose these factors after they have been abandoned.

Consider a firm that either refrains from investing in a data warehousing solution or divests from it. While the initial decision might be influenced by various factors, the real challenge emerges when competitors have already established sophisticated data warehousing solutions. These competitors, having harnessed the power of their data, gain significant advantages in insights, agility, and decision-making. For the firm that lagged behind or divested, trying to catch up becomes not just a matter of investment but also of time. The gap widens as competitors continue to refine and leverage their solutions, making it increasingly difficult for the lagging firm to bridge the divide and regain competitive parity.

Lags and Inertia Ghemawat (1991) identifies lags as the temporal delays that organizations face when adjusting their stocks of sticky factors to desired levels. In the context of data warehousing, these lags can manifest in multiple ways. For instance, once a firm recognizes the need to upgrade or modify its data warehousing infrastructure, the actual implementation of these changes can take considerable time. This delay can be attributed to various factors, such as the intricacies of the technology, the need for specialized skills, or the alignment with other organizational systems. The presence of such lags means that even if a firm desires to pivot its strategy or adapt to new market conditions, the inherent delays in adjusting its data warehousing capabilities can act as a deterrent, further solidifying its commitment to the existing setup.

Inertia, on the other hand, delves into the organizational psyche. Ghemawat (1991, p.26) notes that "organizations are widely presumed to have a built-in bias toward inertia". This inherent resistance to change is not just a function of tangible factors like resources or technology but is deeply rooted in organizational culture, mindset, and routines. The prevailing literature often focuses more on the challenges of inducing change within organizations than on the risks of becoming too rigid or ossified. In the sphere of data warehousing, inertia can be observed when organizations continue to rely on legacy systems or outdated data practices, not necessarily because they are the best fit for the current needs, but because of a deep-seated resistance to change. This inertia can be detrimental in a rapidly evolving data landscape, where adaptability and agility are paramount.

Digital Commitment Dynamics: Balancing Competitive Edge and Flexibility Research by Bruno Cassiman, Ricart, and Valentini (2022) on digital platforms provides a contemporary lens through which to view the intricacies of strategic commitment, especially in the digital age. Cassiman, Ricart, and Valentini (2022) emphasize the delicate balance that firms must strike between the sustainability of competitive advantage and the value of flexibility. This trade-off is particularly salient in the realm of data warehousing. On one hand, a robust and specialized data warehouse can provide a firm with unparalleled insights, driving a sustainable competitive edge. On the other, the very architecture that supports this advantage can become a source of rigidity, limiting the firm's ability to pivot in response to changing market dynamics.

Cassiman, Ricart, and Valentini (2022)'s insights resonate deeply with the challenges of committing to data warehousing. The uncertainty surrounding such commitments is twofold: firstly, there's the question of whether the firm is committing to the right "sticky" factors that will yield long-term benefits. Secondly, even if the right factors are identified, there's the challenge of effectively developing them. This mirrors Inmon (2005)'s assertion that organizations must have a clear vision of what a data warehouse can achieve for them and the opportunities it presents before its implementation.

Furthermore, Cassiman, Ricart, and Valentini (2022)'s research underscores the importance of architectural choices in determining a platform's flexibility and adaptability. Drawing parallels with the digital platform realm, the example of booking.com's inability to swiftly respond to Airbnb's value proposition due to its architectural constraints highlights the "stickiness" of such decisions. Similarly, in the context of data warehousing, architectural choices made during the initial stages can have long-lasting implications. A firm might find itself constrained by its data warehouse's architecture, limiting its ability to adapt to new data sources, integrate novel analytics tools, or pivot to emerging business models.

2.2 Factors on Data Warehouse Implementation Success

This chapter dives into the literature on data warehousing and BI success factors, outlining the most important contributions towards this area. Subsequently, the study will explore in more detail the factors that were identified as most influential.

The work by Wixom and Watson (2001) stands out for its rigorous examination of these factors. Surveying 111 organizations across diverse industries, Wixom and Watson (2001) embarked on a comprehensive exploration of the relationships between various potential success factors and the different dimensions of data warehousing success. The study's methodology was structured around a multi-tiered regression analysis, as illustrated in Figure 1.

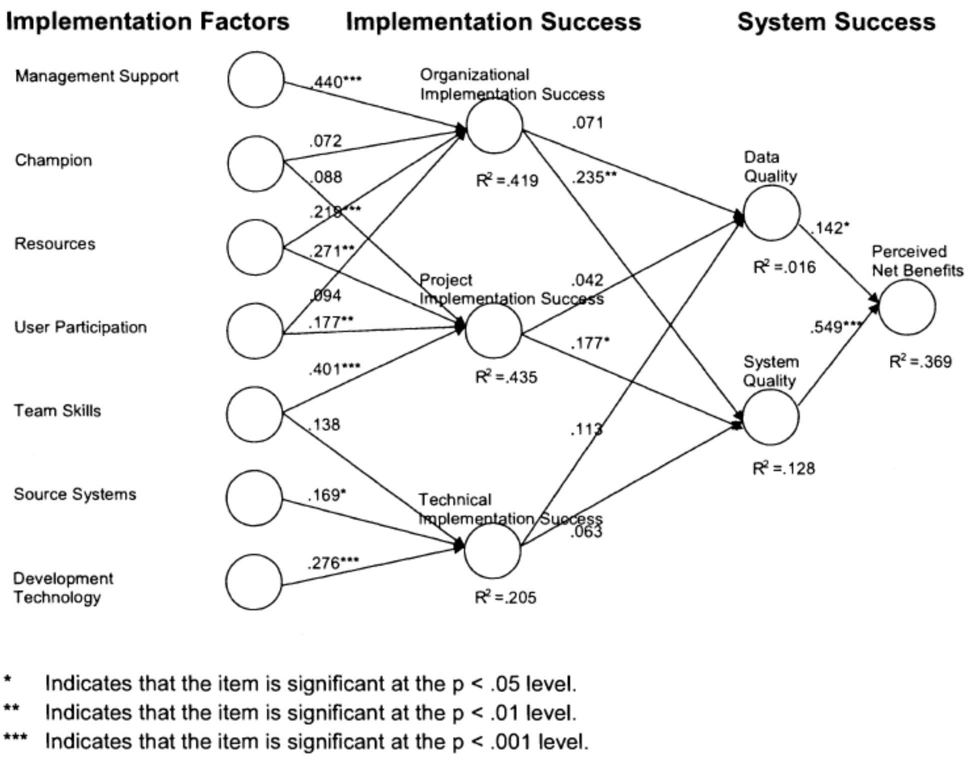


Figure 1: Wixom and Watson (2001) Success Factors for Data Warehouse Implementation.

The figure represents a cascading model where Wixom and Watson (2001) first regressed responses from the survey, which ranged from one to seven (from "strongly disagree" to "strongly agree"), on implementation success. This was further broken down into three constituents: organizational, project, and technical implementation success. Each of these constituents was then regressed on the most promising of the seven identified potential success factors.

In the subsequent stages of the model, Wixom and Watson (2001) regressed system success, which was composed of data quality and system quality, on the aforementioned implementation success constituents. Lastly, perceived net benefits were regressed on data quality and system success, providing a holistic view of how initial factors could potentially influence the ultimate benefits derived from data warehousing.

This multi-tiered approach adopted by Wixom and Watson (2001) provides a nuanced understanding of the intricate web of factors and their interplay in determining the success of data warehousing implementations. The study's findings serve as a foundational reference for subsequent research in the domain.

The study by Yeoh and Koronios (2010) complements Wixom and Watson (2001)'s work by adopting a different methodology, specifically the Delphi interview method, involving BI experts from five companies across four industries. The study aimed to reach a consensus on the importance of various CSFs using a five-point Likert scale. The results underscored the significance of management support, clear vision, and user participation, among other factors.

Yeoh and Koronios (2010)'s research further validated these CSFs through multiple case studies, revealing that successful BI implementations were characterized by systems that were stable, easy to use, fully functional, and responsive. Importantly, the study found that a business-oriented approach to managing CSFs was pivotal. Organizations that aligned their BI initiatives with specific business needs from the outset were more likely to achieve successful implementations. This business orientation served as a meta-CSF, dictating the effectiveness of other success factors.

Yeoh and Koronios (2010)'s findings contribute to the understanding that non-technical factors, including organizational and process-related aspects, often outweigh technological and data-related factors in determining BI implementation success. Therefore, Yeoh and Koronios (2010)'s work serves as a critical extension of existing literature, emphasizing the need for a business-driven, organization-focused approach to BI implementation.

Building on the foundational works of Wixom and Watson (2001) and Yeoh and Koronios (2010), the study by Mungree, Rudra, and Morien (2013) introduces a more contemporary perspective on the CSF in BI implementations. Utilizing a survey-based research approach that includes interviews with 16 BI consultants, the study emphasizes that BI is not solely a technological endeavor but a complex interplay of technology, human capital, and strategic alignment. The study identifies ten CSFs crucial for BI implementation success that we will dive into in more detail in later chapters.

The study by El-Adaileh and Foster (2019) offers a meta-analysis of 38 articles, distilling the collective wisdom on BI implementation success factors. As illustrated in Figure 2, the review identifies management support as the most frequently cited factor, affirming its central role.

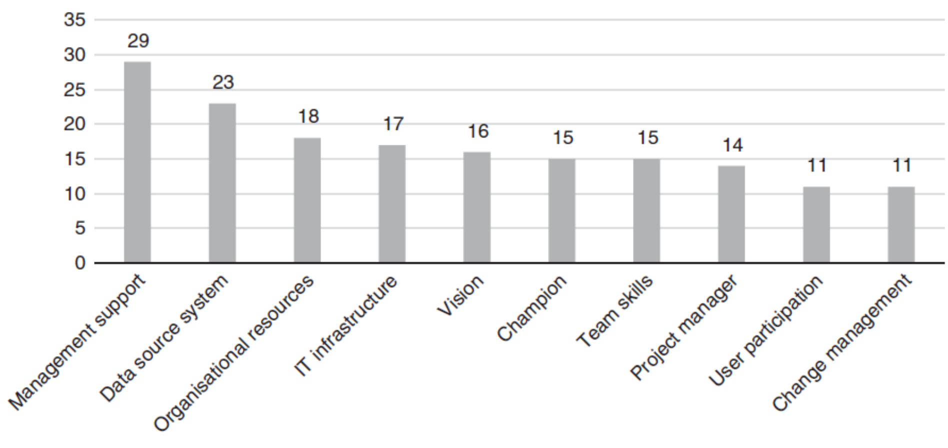


Figure 2: Most common implementation factors for BI success.

The study furthermore points out that while much of the existing literature focuses on organizations in developed countries, there is a need for more research on BI implementation in developing countries, where different challenges and priorities may exist.

In synthesizing these diverse studies, it becomes evident that while there is a consensus on certain factors like management support and user participation, the significance and definition of these factors can vary. This variability is influenced by the methodologies employed, the evolving technological landscape, and the specific contexts of the studies.

Interestingly, at most very few studies take factors like firm size, usage of cloud computing solutions, or outsourcing into account. Some of the studies date back multiple decades and are therefore less representative for the considerations of IT decision makers nowadays.

2.2.1 Operational Effectiveness Factors

As done by Wixom and Watson (2001), this study will group the most influential factors into different categories. While Wixom and Watson (2001) show two different groupings for organizational and project implementation success factors, here they are merged together into one group, the operational effectiveness factors.

Management Support According to Wixom and Watson (2001, p.23), "Management support is widespread sponsorship for a project across the management team" and "motivates people in the organization to support the data warehousing initiative and the organizational changes that inevitably accompany it".

The study by Xu and Hwang (2007) offers a contrasting perspective on the determinants of data warehousing success. Based on 98 questionnaires completed by data warehouse professionals, the study employed regression analysis to understand the influence of eleven implementation factors on eight distinct success variables. The detailed results of this regression analysis can be observed in Figure 3.

The table presents a matrix where the rows represent the eight success factors, and the columns detail the eleven implementation factors. Each cell within the matrix provides the regression coefficient, with the p-value in brackets. A standout observation from the table is the non-significance of the "top management support" factor across all success variables.

This finding is initially surprising, given the traditionally held belief in the paramount importance of top management support in ensuring the success of data warehousing im-

Data Warehousing Success (Dependent Variable)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	Intercept	R2
Easy to use	.192 (.035)		.225 (.027)	.288 (.003)			-.443 (.000)	.210 (.051)		.210 (.042)		1.022 (.057)	.365 (.000)
Speedy information retrieval			.243 (.014)		.294 (.007)				-.208 (.066)	.185 (.062)		1.997 (.000)	.212 (.000)
More information							.241 (.007)			.299 (.000)		1.491 (.000)	.232 (.000)
Better quality information				.376 (.000)			.276 (.005)					1.449 (.000)	.346 (.000)
Improved productivity			.215 (.010)			.147 (.069)				.213 (.019)	.203 (.010)	1.049 (.014)	.369 (.000)
Better decisions	.138 (.098)					.151 (.090)	.510 (.000)	-.155 (.101)			.201 (.014)	.696 (.142)	.447 (.000)
Improved business processes										.199 (.045)	.391 (.000)	1.561 (.000)	.279 (.000)
Increased competitive position	.205 (.050)						.342 (.006)		-.196 (.107)		.558 (.000)	.097 (.862)	.418 (.000)

[1] **Independent variables for the regression models are:**

F1: clearly defined business needs/benefits	F7: project management (teamwork)
F2: top management support	F8: practical implementation schedule
F3: user involvement/participation	F9: proper planning / scoping of project
F4: source data quality	F10: adequate funding
F5: proper development technology	F11: measurable business benefits
F6: adequate IS staff and consultants	

[2] **The first number in each cell (except the last column) is the regression coefficient; the second number in parenthesis is the p-value**

Figure 3: Xu and Hwang (2007) Success Factors for data warehouse Implementation.

plementations. Xu and Hwang (2007) posits that while top management might have fully embraced the concept of data warehousing, making it a non-critical factor in this context, its effect could be indirect and not captured in the model. This perspective finds partial resonance with Wixom and Watson (2001)'s findings, which suggested that top management support influences organizational implementation success, which subsequently affects system quality.

Due to this finding, there is no doubt that Management Support is a very important factor to consider for data warehouse implementation success.

Champion "A champion actively supports and promotes the project and provides information, material resources, and political support" (Wixom and Watson 2001, p.23).

Even though this factor was ranked number six in El-Adaileh and Foster (2019)'s ranking, Wixom and Watson (2001)'s analysis shows that there was no significant influence on neither organizational nor project implementation success. One possible reason is that "[a] single warehouse champion may abandon the project at the first sign of trouble [...] and has limited influence and understanding outside his or her own area of the organization" (Wixom and Watson 2001, p.36).

Furthermore, Mungree, Rudra, and Morien (2013) does not analyse the factor on its own, but includes it in the factor of appropriate team skills. For this reason the meaningfulness of this factor is less clear and for the above reasons not included explicitly into the operational effectiveness factors.

Resources According to Wixom and Watson (2001), resources include the money, people, and time that are required to successfully complete the implementation project.

While this factor shows great promise in the literature and empirical studies, the reason for its success might be that it is too broadly defined, therefore capturing the explanatory power of multiple underlying factors. While Wixom and Watson (2001) approached the factor with the questions of adequate funding of the project, enough team members, and enough time for completion, the regression on this factor raises multiple questions:

- How are the results of the regression influenced if not only big corporations are included in the population but also Small and medium-sized enterprises (SME)?
- Which time frame was considered? Does the time frame also include ongoing costs up to a point in the future? If not, it is difficult to compare data warehouses in the cloud with investment-intensive on-premise solutions.

- What split between employees and consultants making up the implementation team members was considered?
- When is the project considered complete? As learned from Inmon (2005) and Kimball et al. (2008), a data warehousing project can take a very different amount of time depending on the two different implementation approaches.

Due to these unanswered questions, this study will broaden its focus on this factor by looking also at factors that are ultimately expected to affect its outcome: the size of the implementing company, the extent of cloud service and outsourcing usage, and the implementation approach.

In the following sections these factors are looked into in more detail. Undeniably, however, Resources are one of the fundamental factors of data warehousing implementation success and deserve consideration.

Vision Regarding this factor, El-Adaileh and Foster (2019, p.125) note "that a system for BI has to be tied closely to the strategic vision of a company".

Mungree, Rudra, and Morien (2013, p.7) split this factor into multiple factors: "alignment of BI strategy with business objectives" and "Clear vision and well-defined information and systems requirements". Since the differentiation seems to be negligible, El-Adaileh and Foster (2019)'s approach will be inherited and the two will be combined into one single factor.

As previously outlined, Yeoh and Koronios (2010) also emphasize the importance of aligning the BI system closely with specific business objectives from the very beginning of implementation. According to Yeoh and Koronios (2010), this alignment serves as a kind of meta-CSF, and its presence or absence can be a decisive element in determining the overall success of the BI initiative.

Team skills, User Participation, Change Management, and Project Management Concerning team skills, Wixom and Watson (2001) ask the participants about the data warehouse implementation team having the right technical and good interpersonal skills. Notably, they do not differentiate between consultants and employees.

Wixom and Watson (2001) also include the factor user participation in their regression, but do not include change management nor project management as included by El-Adaileh and Foster (2019). According to them, "user participation occurs when users are assigned project roles and tasks, which leads to a better communication of their needs and helps ensure that the system is implemented successfully" (Wixom and Watson 2001, p.24).

In a Forbes article, James Lupton (2022), who serves as the CTO of a consultancy specializing in data and analytics strategy, delves into the key pitfalls that often lead to the failure of data warehousing initiatives. Lupton emphasizes that the most common reason for failure is a lack of user focus. Instead of addressing the needs and delivering value to the end users, these initiatives often become overly concentrated on technological aspects. This results in an inability to secure further investments or to engage a consistent user base from various departments. Lupton argues that the failure to align the project with the specific needs of a target customer, such as a marketing team looking for customer insights, turns these data warehouses into costly liabilities over time, rather than assets that add value to the organization.

Team skills and user participation both being significant factors for the project implementation success according to Wixom and Watson (2001) and among the top ten most cited factors according to El-Adaileh and Foster (2019), are included in the operational effectiveness factors.

Change management, according to Mungree, Rudra, and Morien (2013), involves development iteratively with strong user involvement. Here, the lines blur in the literature, mixing in with the user participation factor. Seemingly, less change management is needed, when users are included in the development to begin with.

"The term 'project management' is in reference to the ongoing management of the plan for implementation. As well as stages of planning it involves, therefore, the allocation of responsibilities to a variety of stakeholders, definition of critical paths and milestones, human resource planning, determination of success indicators and training" according to El-Adaileh and Foster (2019, p.126).

Since change management interferes with user participation, it will be disregarded in the sense of a factor for operational effectiveness. Project management, being less similar to user participation and team skills factors, will be included.

2.2.2 Technical Factors

While this study already touched on Yeoh and Koronios (2010, p.31)'s finding "that non-technical factors, including organisational and process-related factors, are more influential and important than technological and data-related factors", they shall also be included in the analysis for a comprehensive view. It's essential to note that although non-technical factors like organizational alignment and process-related aspects may carry more weight, according to Yeoh and Koronios (2010), technical factors should not be disregarded. They serve as the backbone of any BI or data warehousing initiative.

Source Systems Wixom and Watson (2001) identified a notable challenge in data warehousing projects, specifically the issue of inconsistent data definitions across disparate data sources. The inconsistency complicates the task for teams who are responsible for reconciling and uploading the data into the warehouse.

This problem can make data warehousing projects especially difficult to implement for several reasons. First, resolving these inconsistencies often requires additional time and effort, thus increasing the overall cost of the project. Second, the main advantage of a data warehouse is to serve as a single source of truth for an organization. If the data in the warehouse is inconsistent or incorrect due to differing data definitions from multiple sources, it undermines the very purpose of having a centralized data warehouse (Wixom and Watson 2001).

In the context of source systems, the diversity and disparity of applications and systems present a unique set of challenges for data warehousing projects. When data is drawn from a variety of different applications and systems, the complexities involved in data integration multiply. This is in line with Wixom and Watson (2001, p.30)'s survey questions, in which participants are asked to consider whether "the data sources used for data warehouse were diverse and disparate applications/systems" and whether "a significant number of source systems had to be modified to provide data for [the] data warehouse".

As with inconsistent data definitions, diversity can be a double-edged sword. On one hand, disparate systems can offer a broader range of data, potentially leading to richer insights. On the other hand, they can create hurdles in terms of data standardization and compatibility. Data from different systems might be formatted differently, use varying scales or units, or even provide conflicting information, all of which require careful attention to reconcile.

Additionally, modifications to existing systems may be necessary to allow them to interface effectively with the new data warehouse. These modifications are not just technical challenges but can also involve administrative overhead, further stretching the budget and timeline of a data warehousing project.

Mungree, Rudra, and Morien (2013) refer to the Source Systems factor in their framework as "Effective Data Management", including among others the dimensions "good data quality at source systems" and "integration of data from multiple sources". As the second most cited factor for implementation success according to El-Adaileh and Foster (2019) and significant at the five percent level regarding the technical implementation success dimension according to Wixom and Watson (2001) it is included in the technical factors.

Development Technology The concept of "IT Infrastructure" as outlined by El and "Development Technology", as described by Wixom and Watson (2001), share a common core focus on the role of technology in the successful implementation of a data warehousing project. El-Adaileh and Foster (2019) highlight the importance of a reliable, secure, and adaptable IT infrastructure, not just as a combination of hardware and software, but as a complex ecosystem that aligns with long-term business objectives.

Similarly, Wixom and Watson (2001) underscore the critical role played by development technology, specifically the tools and methods used for the data warehouse project. These tools not only need to be effective but also must integrate well with existing technologies within the organization. While these factors are articulated slightly differently, they both emphasize the technological backbone necessary for the success of BI and data warehousing projects.

Given these similarities and the significance of this technological aspect as evidenced particularly by Wixom and Watson (2001)'s research, these factors are consolidated into a single development technology factor in this framework. Even though this aspect is not among the top ten most cited factors in El-Adaileh and Foster (2019)'s study, its undeniable impact on technical implementation success justifies its inclusion.

The development technology factor is particularly interesting from a strategic perspective since it embodies the tension between commitment and flexibility according to Ghemawat (1991). In the realm of data warehousing, the choice of development technology — comprising the tools, methods, and platforms used — is in itself a form of strategic commitment. This choice is not merely a technical decision but a strategic one that aligns with long-term business objectives, echoing Ghemawat (1991)'s notion of lock-in and path dependency. Once a firm commits to a specific set of technologies, it is implicitly committing to a particular trajectory that is difficult to reverse without significant cost and effort. These technologies often have characteristics of durability, specialization, and untradeability, making them "sticky" factors that influence both current and future strategic moves.

However, the rapid pace of technological advancement in the last decade adds a layer of complexity to this commitment. Unlike more stable assets, development technology can become obsolete or less competitive in a short span of time, challenging the durability aspect of Ghemawat (1991)'s lock-in. This rapid evolution also brings Cassiman, Ricart, and Valentini (2022)'s insights into focus. The value of flexibility, especially in the digital age, cannot be overstated. Firms must balance the allure of a competitive edge gained through cutting-edge development technology with the need for adaptability. The architecture and tools chosen today might constrain a firm's ability to integrate new data sources, employ emerging analytics tools, or adapt to changing business models tomorrow. This issue will be revisited upon examining the implementation cost component of the NPV.

Moreover, the speed at which technology evolves can induce lags as described by Ghemawat (1991). Even if a firm recognizes the need to pivot its technology strategy, the time required to transition from one set of development technologies to another can be significant. This lag can result in lost opportunities and can serve as a deterrent to adapting to new market conditions. Inertia, too, can be a significant barrier. Organizations may resist migrating to newer technologies due to the built-in bias towards maintaining existing systems and practices, even when they are no longer the most efficient or effective.

2.2.3 Emerging Factors

This section highlights those factors that have not been explored in as much detail as others and are considered important for further investigation of data warehouse implementation success.

Company Size The factor of company size in the context of data warehousing and BI implementation has been relatively underexplored in the literature compared to other factors. However, it emerges as a significant determinant in the study of Ramamurthy, Sen, and Sinha (2008). The study argues that larger firms are generally better positioned to afford the substantial resources — both financial and human capital — required for successful data warehousing implementation. Larger firms can also leverage economies of scale and scope, making the massive initial investments more justifiable. Moreover, the larger customer and supplier base of such firms provides more opportunities for leveraging the initial investment in data warehousing infrastructure.

Ramamurthy, Sen, and Sinha (2008) point out the importance of a company's ability to learn and adapt when adopting new technology. In simpler terms, it's not just about having the money to invest in data warehousing; the company also needs to have the right skills and culture to make the most of this technology. This is especially true for data warehousing, which isn't just a one-time purchase but a concept that needs to be integrated into the company's overall operations and strategy.

However, the landscape is changing. Increasingly affordable data warehousing technologies are making it easier for SME to invest in these solutions. This is corroborated by a survey by The Data Warehousing Institute (TDWI), which found that 53 % of data science leaders have fewer than 5,000 employees. It suggests that smaller firms, often more agile and adaptive, are increasingly becoming Insights Leaders (TDWI 2017).

For SME, cloud computing offers a particularly attractive option for data warehousing. The flexibility of cloud solutions allows SME to purchase additional IT resources as needed, effectively managing the rapid changes in today's business models. This flexibility replaces the need for extensive management activities, backup, recovery, security, and performance and capacity management with a more manageable monthly cost (Fernandes and Bernardino 2016).

The literature therefore presents a nuanced view of the role of company size. While larger firms have the advantage of resources and economies of scale, smaller firms benefit from agility and the decreasing costs of technology. This dichotomy aligns with previous studies that have found inconsistent findings regarding the influence of firm size on innovation adoption. Some suggest that smaller firms may be more efficient at generating and adopting innovations, while others argue that larger firms can more effectively mobilize the resources required for innovations (Ramamurthy, Sen, and Sinha 2008).

Extent of Cloud Service Usage The advent of cloud computing has fundamentally altered the landscape of data warehousing and BI implementation. The cloud offers a scalable, flexible, and cost-effective solution for data storage and analytics, making it an increasingly popular choice for organizations. According to a 2020 study by TDWI, 66 % of respondents indicated that most or all of their analytics and AI/ML initiatives are running in the cloud, and 69 % are using the cloud for data management (TDWI 2020).

This trend is particularly significant for SME, which often lack the resources for large-scale, on-premises data warehousing solutions. Cloud-based data warehouses offer SME the ability to scale their data storage and analytics capabilities in line with their business growth, without the need for significant upfront investment. A Microsoft (2023) report suggests that SME face a choice between adopting big data tools for future extensibility or maintaining traditional SQL-based solutions for cost efficiency and ease of maintenance. Interestingly, a hybrid approach is also emerging, which combines the ease of migrating existing data with the opportunity to add big data tools for specific use cases.

The cloud's pay-as-you-go model offers several advantages, including:

- **Flexibility**: Organizations can easily scale their data warehousing solutions up or down based on their needs, without incurring sunk costs.

- **Cost Transparency**: The monthly subscription model allows organizations to understand their costs upfront, aiding in more accurate financial planning.
- **Reduced Time-to-Market**: Cloud-based solutions can be deployed more quickly than traditional on-premises solutions, enabling faster realization of business value.

However, while the cloud offers the advantage of reducing upfront capital expenditure, operational costs can accumulate over time, affecting the long-term NPV of the investment.

Extent of Outsourcing In the intricate landscape of data warehousing, it's important to clarify the distinct roles of outsourcing and cloud usage, especially since these two factors are often blurred in the literature. While this study previously delved into the topic of cloud usage, focusing on the infrastructure where the data warehouse is hosted, this section will separately examine outsourcing. Outsourcing pertains to the delegation of specific tasks or entire functions related to data warehousing to third-party vendors. By treating these two factors independently, this study aims to provide a clearer understanding of their unique impacts and considerations in the realm of data warehousing.

Understanding outsourcing in data warehousing is not about rigid categories but rather about a continuum that reflects the varying degrees of external involvement. At one end of this continuum, an organization might opt for a fully in-house approach, relying on internal employees who have been specially trained for data warehousing tasks. At the opposite end, an organization might outsource both the initial setup and the ongoing management, effectively treating the data warehouse as a data-science platform as a service (TDWI 2017). Between these two extremes, numerous hybrid models exist, such as using external consultants for the initial setup while keeping the ongoing management in-house. No matter where an organization falls on this continuum, the importance of having an internal project leader to co-manage with the external partner is crucial. This ensures that expectations, resources, and timelines are clearly defined and met (Westerman 2001).

The skills required for effective data warehousing are diverse and complex, encompassing technical, system, and business skills. This complexity often poses a challenge for SME that may lack the resources to acquire these skills in-house. Preston and Brohman (2002) suggest that outsourcing can be a viable solution to this problem, providing access to specialized skills that are otherwise scarce or expensive to develop internally. However, while technical and system skills can be more easily outsourced, business skills are often specific to the organization and may require a closer, more collaborative relationship with the service provider for effective transfer.

Outsourcing comes with its own set of opportunities and challenges. On the upside, it offers flexibility, scalability, and access to specialized expertise. However, as Payton and Handfield (2003) note, outsourcing can complicate the already challenging landscape of data warehousing implementation. Risks include reduced control over data and limitations on customization. The economics of outsourcing can also vary based on the organization's specific needs. For example, if an organization requires a large volume of standard reports, the transaction costs associated with monitoring an outsourced service may outweigh the benefits, making it more economical to keep this function in-house (Preston and Brohman 2002).

In terms of best practices, both studies by Payton and Handfield (2003) and Preston and Brohman (2002) offer valuable insights. Before selecting a data warehouse vendor, organizations should establish success metrics that can later be used to renegotiate contract terms if performance falls short. Even when outsourcing, maintaining some level of internal expertise is crucial for supporting the outsourced functions and ensuring a trustful relationship between the organization and the vendor. Additionally, forming a cross-functional committee can be invaluable for driving critical processes and ensuring alignment with various organizational needs and objectives.

Implementation Approach While there are various methodologies for implementing a data warehouse, this section will focus on the two primary approaches: those championed by Bill Inmon (2005) and Ralph Kimball et al. (2008). It is worth noting that many other approaches are often considered derivatives or variations of these two foundational methodologies. Therefore, by examining the Inmon (2005) and Kimball et al. (2008) approaches, this study also indirectly covers a broader spectrum of implementation strategies.

The Inmon (2005) methodology advocates for a top-down approach, where an enterprise-wide data warehouse serves as the central repository for all organizational data. Data marts, specialized subsets of data warehouses, are then created based on this centralized warehouse. This approach ensures a high degree of data consistency across the organization but can be resource-intensive and time-consuming. For larger firms with substantial resources, the Inmon (2005) approach may offer a more integrated and consistent data architecture. However, the initial investment required can be substantial, affecting the NPV negatively in the short term but potentially offering long-term benefits through economies of scale and scope.

In contrast, Ralph Kimball et al. (2008)'s methodology suggests a bottom-up approach, starting with the creation of individual data marts for specific departments or functions. These data marts can later be integrated to form a comprehensive data warehouse. This approach allows for quicker implementation and may require a smaller initial investment, making it more accessible for SME. However, Kimball et al. (2008)'s approach runs the risk of creating data inconsistencies across different departments, especially when these data marts are developed in isolation. Over time, the costs of managing and reconciling these disparate data marts can accumulate, potentially making this approach more expensive in the long run.

The choice between the Inmon (2005) and Kimball et al. (2008) approaches can significantly impact the NPV of the data warehousing investment. The Inmon (2005) approach, with its higher initial costs, may result in a lower NPV in the short term but offers the potential for greater long-term value through data consistency and integration. Conversely, the Kimball et al. (2008) approach may offer a higher short-term NPV due to its lower initial investment and quicker time-to-value but may incur additional costs in the long term due to data inconsistencies and the need for reconciliation.

GDPR Awareness While existing literature may not extensively cover the impact of GDPR on data warehousing, insights from the study by Li, Yu, and He (2019) on the impact of GDPR on global technology development offers valuable perspectives.

Firstly, GDPR compliance can significantly increase the operational costs of maintaining a data warehouse. Li, Yu, and He (2019) emphasize that compliance necessitates a thorough internal assessment of technology platforms and data architectures, including data warehouses. Organizations may need to reengineer existing systems to reduce the risk of non-compliance, involving significant investments in manpower and resources. This increase in operational costs could negatively affect the NPV of the data warehouse investment.

On the other hand, GDPR compliance could serve as a form of risk mitigation. Non-compliance with GDPR can result in substantial fines and damage to customer trust (TDWI 2023b). By investing in GDPR compliance measures, companies can potentially avoid these costs, thereby preserving or even improving the NPV of their data warehouse over the long term.

Moreover, GDPR's stringent requirements for data protection and consumer rights, such as the Right to be Forgotten and Right of Access to Data, add another layer of complexity. The development of holistic search tools that can identify and extract personal data across various platforms and systems may be required. This could further escalate operational costs, thereby influencing the NPV calculations.

On the cybersecurity front, GDPR mandates companies to implement reasonable data protection measures. This could lead to an increased demand for cybersecurity professionals and data protection officers, adding to the operational costs. However, this should also be viewed as a form of risk mitigation, as non-compliance could result in substantial fines and loss of consumer trust.

For managers, these increased operational costs should be weighed against the benefits of having greater control over data and enhanced data security. While compliance may entail higher upfront and ongoing expenses, it also provides a structured framework for data governance that can mitigate risks and potentially lead to long-term savings. Li, Yu, and He (2019) also point out that compliance with GDPR can serve as a competitive advantage, especially in a landscape where consumer trust is increasingly important.

Furthermore, Li, Yu, and He (2019) discuss the impact of GDPR on emerging technologies like AI, blockchain, and cloud computing. These technologies often rely on massive data sets, and GDPR's stringent data handling and processing regulations could inhibit their development and increase costs. For data warehouses that integrate such emerging technologies, this could mean additional compliance challenges and costs, affecting NPV.

2.3 NPV Constituents

As shown in Figure 4 the NPV consists of five components. The y-axis depicts the relative value of a cash in- or outflow from a company that implements a data warehouse. The expected value of implementation investment or cost is negative and therefore marked red. It is expected to be of higher value in the first periods of the implementation time frame than the expected values of ongoing costs. On the long run the expected values of ongoing costs are expected to be of similar magnitude. After the first few periods with only cash outflows the first benefits of the implementation show yields and continue to grow until they reach a level of similar magnitude, too. These expected values of benefits can be monetary and non-monetary and therefore need to be valued differently.

Including more or fewer periods can have an affect on the NPV and therefore needs to be analyzed. The question here is when the implementation is considered complete. Furthermore, the riskiness of the expected cashflows needs to be evaluated. While an expected value by itself can already be obtained by multiplying the expected cashflow by its probability, this expected value should furthermore be discounted to reflect the uncertainty over time. Discounting the netted cashflows of every period throughout the

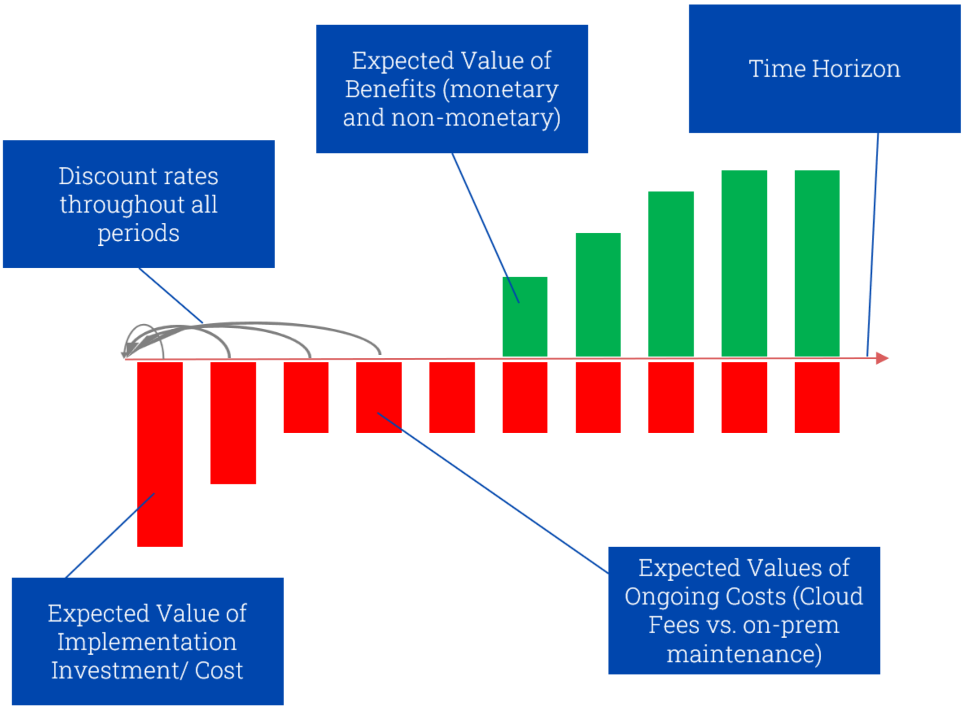


Figure 4: Components of the NPV.

time horizon to the current moment in time will then lead to the decision measure NPV, which is used to evaluate undertakings. If the NPV is negative, the benefits are not worth the cost. If it is positive, it should be compared to other investment opportunities to decide on the most promising option.

2.3.1 Expected Value of Implementation Costs

The construction phase is often the most time-consuming and resource-intensive part of building a data warehouse. Westerman (2001) notes that this phase will use the most human resources and will prove to be the most time-consuming. Following, the up-front costs that the literature provides concerning data warehousing are dissected.

Proof of Concept Starting with a limited implementation focused on a single subject area can serve as a proof of concept. This approach can help secure a positive NPV for the initial investment, potentially making it easier to justify further investment. Westerman (2001) advises that the first implementation should be limited to only one subject area, e.g. sales or procurement. Inmon (2005) concludes that data warehouses are often built incrementally, allowing for the initial iteration to be done quickly and at a relatively low cost. This incremental approach enables organizations to explore the benefits and justify the development costs over time.

Hardware and Software Licenses For on-premises solutions, the initial purchase or lease of hardware and software constitutes one of the most significant upfront costs. This includes the costs associated with servers, storage, and software licenses for database management systems and other essential tools. However, in cloud-based solutions, these hardware costs are mitigated as the infrastructure is provided by the cloud service provider, and clients typically pay as they go or based on their usage. Software licensing models might also differ in cloud environments, often leaning towards subscription-based models. Kimball et al. (2008) specifically list hardware and software license purchases or lease expenses as key cost factors.

Internal Development Resources The internal team responsible for constructing the data warehouse can also significantly contribute to the initial costs. This includes the salaries and benefits for various roles such as the project leader, database analyst, GUI programmer, data warehouse programmer, and "host" programmer. Westerman (2001) outlines these roles as part of the construction team. With outsourcing these costs can be mitigated and shifted towards costs for external resources.

External Resources If the internal team lacks specific expertise, the hiring of external consultants or vendors may be necessary, adding to the initial cost. If the data warehouse is outsourced entirely (BI as a service), this will be the only major ongoing cost component (Kimball et al. 2008).

Education and Training Both the project team and the broader business community may require specialized training to effectively understand and utilize the data warehouse. This could involve formal training programs or workshops. Kimball et al. (2008) include education for both the project team and business community as a cost factor.

Data Preparation The cost of preparing data in the appropriate format for the data warehouse can be a significant factor. This includes the activities of cleaning, transforming, and loading the data into the warehouse (Mitchell, Ker, and Lesher 2021, p.20).

Interface Maintenance The choice between using an ETL tool and building the interface manually can have a significant impact on costs. Inmon (2005) notes that if an organization chooses to build the interface manually, the costs of implementation and maintenance can skyrocket. However, using an ETL tool can mitigate these costs over time.

The implementation of a data warehouse is a complex and resource-intensive endeavor, but it's important to note that many of these costs are one-time investments. Inmon (2005) emphasizes that although the infrastructure for a data warehouse is expensive and difficult to build, it only needs to be built once. After its construction, the data warehouse serves as a flexible and reusable foundation for the organization.

2.3.2 Expected Value of Ongoing Costs

As with the implementation costs, now the ongoing costs are broken down into the main aspects including an explanation on each. This can be divided into three parts, as some of the costs vary between on-premise data warehousing and cloud solutions (Inmon 2005; Kimball et al. 2008; Westerman 2001).

On-Premise and Cloud Ongoing Costs

- **Staff Maintenance and Training:** Costs associated with IT staff for software maintenance, data management, and addressing integration issues. This also includes periodic training to keep up with updates, best practices, and platform-specific nuances.
- **Software Licenses and Updates:** Expenses related to periodic software license renewals, software updates, patches, and additional modules or features.
- **Security and Compliance:** Expenditures for security software or services, periodic security audits, compliance checks, certifications, and ensuring GDPR compliance and data protection measures.
- **Backup and Disaster Recovery:** Costs for implementing and maintaining backup solutions, as well as disaster recovery plans and infrastructure.
- **Data Integration and ETL Processes:** Ongoing expenses related to data integration tools, ETL processes, and potential modifications to accommodate new data sources.
- **User Access and Support:** Costs associated with user management, access controls, and providing support to end-users, including helpdesk services for the usage of the data warehouse and its interface.

On-Premise Only Ongoing Costs

- **Hardware Maintenance:** Costs for maintaining, repairing, and eventually replacing server hardware, storage devices, and network equipment.
- **Energy Consumption:** Expenses related to powering the servers, cooling systems, and other infrastructure components.
- **Facility Costs:** Rent or mortgage for the physical space housing the data center, including costs for cooling, fire suppression, and physical security measures.
- **Hardware Depreciation:** Accounting for the decrease in value of the hardware assets over time.
- **Redundancy Systems:** Expenses for systems that ensure high availability, such as redundant power supplies and network connections.

Cloud Only Ongoing Costs

- **Subscription Fees:** Periodic fees paid to the cloud service provider, which can vary based on usage, data volume, or specific services utilized.
- **Data Transfer Costs:** Charges associated with transferring data in and out of the cloud environment.
- **Managed Services:** Additional costs if opting for managed services where the cloud provider handles certain administrative tasks.
- **Scalability Costs:** Expenses incurred when scaling resources up or down based on demand. Kimball et al. (2008) note that these can include costs for loading additional data, accommodating expanded user populations, new release upgrades, and technology to support higher performance demands.
- **Service Add-ons:** Costs for additional services or features that can be added to the basic cloud package, such as advanced analytics or machine learning capabilities.
- **Data Retrieval Fees:** Some cloud providers charge fees for retrieving data, especially if it's archived or in "cold storage".

Ongoing costs in data warehousing are a critical aspect that organizations must consider for long-term sustainability. Contrary to the general trend in information technology where initial setup costs are the most significant, Inmon (2005) argues that in the context of data warehousing, ongoing maintenance costs can far outweigh the initial infrastructure costs. This is especially true when using cloud services.

Westerman (2001) highlights the rapidly decreasing costs of data warehousing technology. He provides the example of Wal-Mart's first data warehouse database, which cost around $20 million in 1993, stating that a similar database would cost around $750,000 eight years later due to advancements in technology.

Cost-saving Measures Ongoing costs are largely dependent on the way the data warehouse is built and which data is loaded into it. The literature gives best practices on keeping the ongoing costs low while maintaining the advantages offered by the data warehouse.

Inmon (2005) provides a breakdown of memory types and their associated costs and retrieval speeds. While this information might be more relevant for on-premises solutions, it offers insights into the trade-offs between cost and performance. For instance, main memory is very fast but also very expensive, whereas options like magnetic tape are not fast but are much less expensive.

To manage costs, Inmon (2005) suggests a dual-level architecture where only aggregated data is stored in the data warehouse, while detailed data is stored in slower, inexpensive storage. This approach allows organizations to maintain detailed data without incurring high storage costs.

Another cost-saving measure is the cyclicity of data, which refers to the time it takes for changes in the operational environment to be reflected in the data warehouse. Inmon (2005) notes that a less tight coupling between the operational environment and the data warehouse can reduce technology costs. This is due to the fact that multiple changes are made to the data in the operational environment until it is completed. Storing each step along the way would incur additional costs that can be mitigated. Here again one can see the trade-off between the timely accessibility of data and cost effort.

Inmon (2005) underscores the significance of granularity in the data warehousing environment. Granularity refers to the level of detail or precision of the data stored in the warehouse. This decision on granularity is pivotal, as it directly influences both the storage costs and the depth of insights that can be derived from the data.

The implementation team faces a delicate balancing act. On one hand, storing data at a very detailed level (low granularity) allows for in-depth analyses, enabling businesses to glean nuanced insights from their data. This could be invaluable for certain use cases, such as fraud detection, where minute details can be crucial. On the other hand, storing data at such a detailed level can quickly escalate storage costs, especially in large organizations with vast amounts of data. Conversely, storing data at a higher level of granularity, such as monthly or quarterly summaries, can significantly reduce storage costs. This approach is more efficient and can be sufficient for many analytical purposes, especially for trend analysis over longer periods. However, the trade-off is that some finer details are lost, potentially limiting the depth of insights that can be derived.

The challenge for the implementation team is to determine the optimal level of granularity that meets the organization's analytical needs while also being cost-effective. This decision should be informed by the specific business use cases and analytical requirements of the organization. It's not just about saving costs; it's about ensuring that the data stored provides real value to the business.

2.3.3 Expected Value of Benefits

In the context of data warehousing, organizations stand to gain a myriad of benefits, both monetary and non-monetary. These advantages not only translate to direct revenue streams and cost savings but also foster a culture of innovation, enhance decision-making, and streamline operations. While monetary benefits are more straightforward, non-monetary benefits offer a broader perspective on the transformative power of data warehousing. The true value does not come from the warehouse itself but lays in how the data is utilized. After giving a brief overview of potential monetary and non-monetary benefits this section will dive into a more detailed explanation thereof.

Monetary Benefits

- Increased revenue due to new sales (enabling new market opportunities with better demand insights), cross-selling, and up-selling.
- Quicker time to market leading to increased revenue.
- Increased profit due to increased response rate to mailings, reduced customer churn, and decreased raw material costs.
- Enhanced customer service or quality levels, leading to higher customer satisfaction and potentially reduced costs.
- Reduction in costs due to data democratization and better data use.
- Savings from reduced research time.
- Direct earnings from optimized operations, such as improved inventory management.

- Significant reduction in the cost of information as e.g. time saved for report creation.

Data warehousing has emerged as a transformative tool for organizations, offering a plethora of monetary benefits that extend beyond mere cost savings. One of the most direct advantages is the potential for increased revenue. For instance, with the insights derived from a data warehouse, businesses can tap into new market opportunities, gaining a clearer understanding of demand patterns. This clarity can lead to more effective cross-selling and up-selling strategies, as highlighted by Kimball et al. (2008).

Furthermore, the speed at which products or services are introduced to the market can significantly influence revenue streams. A quicker time to market, facilitated by the insights from a data warehouse, can lead to substantial revenue increases. As an example, one can consider the potential revenue from launching two new products a month earlier than planned. If every month saved equates to an additional $50,000 in revenue per product, the financial implications are considerable (Kimball et al. 2008).

Profit margins can also see a boost due to the enhanced response rates to marketing campaigns, such as mailings. By targeting the right audience with the right message, businesses can increase sales, reduce customer churn, and even manage their raw material costs more efficiently. The ability to pinpoint low-margin products and eliminate them or to decrease promotional spending based on data-driven insights can further augment profit margins (Kimball et al. 2008).

Customer service, often a significant differentiator in today's competitive market, can also be enhanced through data warehousing. By understanding customer preferences and behaviors, businesses can tailor their services, leading to higher customer satisfaction. This not only retains existing customers but can also reduce acquisition costs for new ones. For instance, if a company can reduce its customer acquisition cost by $75 for each of the 4,000 new customers acquired annually, the savings are substantial (Kimball et al. 2008).

Operational costs can be significantly reduced through effective data use. One of the key strategies in this regard is data democratization, which involves providing everyone in an organization access to the best available data. This approach not only empowers employees to make informed decisions but also eliminates redundant efforts and costs associated with multiple departments seeking the same data. For instance, one can consider a scenario where multiple departments in a large organization each subscribe to the same external data service because they are unaware that other departments are already accessing the same information. This could lead to duplicated costs for the same data subscription. With data democratization, a centralized data warehouse can subscribe to the service once, and then all departments can access the data they need from a single source. This not only ensures consistency in the data being used across the organization but also results in substantial cost savings by eliminating redundant subscriptions (TDWI 2023a). Furthermore, the efficiency gains from having immediate access to required data can reduce research and decision-making time, leading to further operational savings (Westerman 2001).

Optimized operations, such as improved inventory management, can also lead to direct earnings. Using the example of a merchandise manager, by leveraging data warehousing, they can identify inventory flow problems, adjust order frequencies, and consequently increase sales and profits (Westerman 2001).

Another often overlooked benefit is the reduction in the cost of information. Inmon (2005) suggests that data warehousing can lower the cost of information by approximately two orders of magnitude. This means that while one organization might spend a significant amount to access a piece of information, another with a data warehouse could access the same information at a fraction of the cost.

Non-Monetary Benefits

- Building a culture of growth and innovation by data-driven insights.
- Enhanced decision-making process through actionable insights.
- Creation of a single source of data, leading to consistent and reliable information.
- Reduction in data latency, analysis latency, and decision latency.
- Ability to reconcile data discrepancies easily.
- Enhanced system quality through flexibility and integration.
- Ability to compete on a global market by aggregating insights from multiple markets.
- Reduction in dependency on suppliers and customers, as e.g. in contract negotiations due to better market insights, not only horizontally but also vertical across the value chain.
- Reduction in the challenges of data sharing, especially in the context of mergers and acquisitions.
- Enhanced ability to link data with other data, maximizing its value.

Non-monetary benefits, while not directly translating to financial gains, have a profound impact on the overall efficiency, competitiveness, and adaptability of an organization.

One of the most salient advantages of data warehousing is the cultivation of a culture rooted in growth and innovation. As Brynjolfsson, Hitt, and Kim (2011) highlight, firms that emphasize data-driven decision-making exhibit superior performance, with their productivity being notably higher than counterparts that don't leverage data as effectively. This culture of data-driven insights fosters an environment where decisions are made based on concrete evidence rather than intuition, leading to more informed and strategic choices.

Enhanced decision-making is another pivotal benefit. The ability to access and analyze comprehensive data sets allows organizations to derive actionable insights, streamlining processes and optimizing strategies. Kimball et al. (2008) emphasize the transformative power of data warehousing in refining the decision-making process, ensuring that decisions are not only timely but also rooted in accurate and comprehensive data.

The establishment of a single, consistent, and reliable source of data is another significant advantage. Watson and Haley (1998) note that one of the primary motivations for organizations to adopt data warehousing is the desire for better access to information and a unified data source. This centralized approach eliminates discrepancies and inconsistencies, ensuring that all departments and teams work with the same set of data, thereby reducing errors and enhancing collaboration.

Latency, in its various forms, poses a significant challenge to businesses. Watson's work delineates the components of action distance, emphasizing the importance of reducing data, analysis, and decision latency. Data warehousing effectively addresses these latency issues, ensuring that data is not only promptly available but also quickly analyzed, leading to swift decision-making.

Inmon (2005) underscores the flexibility and integration that data warehousing offers, emphasizing its role in enhancing system quality. With a robust data warehouse in place, organizations can easily accommodate future needs and strategic shifts without extensive reprogramming, as highlighted in the survey of Denmark's largest companies by Rikhardsson and Kræmmergaard (2006). This adaptability is crucial in today's rapidly evolving business environment, where agility and responsiveness are paramount.

Furthermore, data warehousing plays a pivotal role in global competition. The ability to aggregate insights from multiple markets allows organizations to craft strategies that cater to diverse audiences, ensuring their competitiveness on a global scale (Rikhardsson and Kræmmergaard 2006).

In mergers and acquisitions, the seamless integration and sharing of data emerge as paramount challenges. As Payton and Handfield (2003) mention, the vastness of global business mergers intensifies the complexities of data management across expansive supply chains. This complexity is further exacerbated when organizations attempt to extend their internal systems to encompass a broader supply chain, including suppliers, business partners, and customers. Particularly, when organizations venture into (in)outsourcing without established success metrics, the barriers to data sharing and integration become even more formidable. The challenge is not just limited to external integrations; even internal data sharing among various departments of an organization can pose significant hurdles, especially when the goal is to achieve a cohesive integration with external supply chain entities.

The intrinsic value of data amplifies exponentially when interconnected with other datasets. Agrawal et al. (2012) underscores that data integration is pivotal in harnessing this enhanced value. In today's digital age, where the majority of data is generated digitally, there lies both an opportunity and a challenge: to shape data creation for easier future linkage and to seamlessly connect previously generated data. Beyond integration, the journey of data encompasses analysis, organization, retrieval, and modeling. The analysis phase, in particular, faces hurdles due to the limitations of current algorithms and the intricate nature of the data. Ultimately, presenting these insights in an accessible manner to those without technical expertise is essential for translating data into actionable knowledge.

2.3.4 Discount Factor and Associated Risks

Determining the appropriate discount rate for a data warehousing project is challenging due to the unique risks associated with such projects. The literature does not provide guidance on establishing a discount factor for data warehousing projects. One potential approach is treating the project like any other investment and applying the Capital Asset Pricing Model (CAPM).

The CAPM as e.g. described by Brealey et al. (2006) offers a method to determine a risk-adjusted discount rate for investments. The formula is given by:

$$\text{Risk-adjusted discount rate} = \text{Risk-free interest rate} + \text{Expected risk premium}$$

Where the risk premium is calculated as:

$$\text{Risk premium} = (\text{Market rate of return} - \text{Risk-free rate of return}) \times \text{Beta}$$

And the Beta of the project is:

$$\text{Beta} = \frac{\text{Covariance}}{\text{Variance}}$$

However, for a data warehousing project, obtaining a Beta value is challenging due to the lack of returns data from comparable projects. This is mainly based on the fact that data warehouse implementation projects are unique in many ways. Therefore, an alternative approach to establishing a discount rate is necessary.

Risk-adjusted Discount Rate Based on insights from Kimball et al. (2008), Inmon (2005), and others, the following risks are the major ones among those associated with data warehousing projects:

1. **High Failure Rates:** Despite the potential benefits of data warehousing, many projects fail to achieve their objectives. Ramamurthy, Sen, and Sinha (2008) as well as Xu and Hwang (2007) summarize failure rates of multiple reports ranging from 41% to as high as 90%. The high fluctuation between these figures is mainly caused by the differences in the definition of failure.

2. **High Implementation Costs:** Implementing a data warehouse is costly, with one study from 1999 reporting an average cost of $2.2 million (Gagnon 1999). As discussed in previous chapters, this value might not be representative anymore due to the possibility of cloud service usage.

3. **Technological Obsolescence:** The rapid pace of technological advancements can quickly make a data warehouse obsolete (Inmon 2005).

4. **Data Breaches:** As previously discussed, data breaches can lead to significant fines and reputational damage. Combining many sources into one can expose firms to the risk of facilitated disclosure of confidential data on a large scale. Additionally, GDPR compliance, while serving as a risk mitigation strategy, can increase operational costs and necessitate system re-engineering (Li, Yu, and He 2019; TDWI 2023b).

5. **Regulatory Changes:** Changes in data regulations can impact the operation and utility of a data warehouse. This includes e.g. the need for tools to support consumer rights like the Right to be Forgotten, and mandatory data collection and access reporting (Li, Yu, and He 2019).

6. **Implementation Challenges:** As highlighted in previous chapters, both organizational and technical aspects can pose challenges to successful implementation.

Given these risks, a base discount rate (as e.g. the Weighted Average Cost of Capital) can be adjusted to account for the riskiness of the data warehousing project. Each risk can be assigned a probability and potential impact, and the cumulative effect can be used to adjust the base discount rate. This risk-adjusted discount rate, while tailored to the specific circumstances of the project, remains subjective and hinges on the expertise and judgment of the individuals conducting the assessment. It's crucial for organizations to ensure that this assessment is as comprehensive and informed as possible, drawing from both internal insights and external benchmarks where available.

2.3.5 Time Horizon

The time horizon is an essential factor in determining the NPV of a project. In essence, it refers to the duration over which future cash flows from an investment are anticipated and used in the NPV calculation. Depending on the scope and nature of the project, time horizons can vary widely.

For data warehousing projects, the time horizon is often a matter of debate. As Ramamurthy, Sen, and Sinha (2008) note, the average project effort for data warehouse is about 4.46 person-years. The concept of 'person-years' refers to the amount of work done by one person in one year. If a project is said to require 4.46 person-years, it could mean a single individual working on the project for 4.46 years, or four individuals working for a bit over a year, or any combination therein.

Different methodologies might advocate for shorter or longer time horizons. For instance, Kimball et al. (2008)'s approach generally focuses on faster results, whereas Inmon (2005)'s methodology is more long-term. As Inmon (2005) suggests, when organizations opt for a short-term approach, it can be challenging to justify a data warehouse due to the high long-term costs associated with building multiple multidimensional database environments. Ghemawat (1991) similarly underscores the point that sometimes it's quicker and potentially more cost-effective to proceed with larger steps than multiple smaller steps.

When performing Discounted Cash Flow (DCF) modeling, a standard practice is to plan as far as one can reasonably estimate. After reaching the limit of reasonable forecasting, a Terminal Value (TV) is typically calculated. The TV represents the estimated value of all future cash flows beyond the forecasted period and assumes that cash flows will continue indefinitely with a constant growth rate. The NPV considering all these factors is:

$$NPV = -IC + \sum_{t=1}^{n} \frac{MB_t - OC_t}{(1+r)^t} + \frac{TV}{(1+r)^n}$$

Where the TV is given as:

$$TV = \frac{(MB_n - OC_n) \times (1 + g)}{r - g}$$

And:

- MB_t are the monetary benefits for period t.
- MB_n are the constantly occurring monetary benefits from period n onward.
- OC_t are the ongoing costs for period t.
- OC_n are the constantly occurring ongoing costs from period n onward.
- IC is the initial investment or implementation costs.
- r is the risk-adjusted discount rate.
- n is the number of periods that can be reasonably forecasted.

Data warehouse projects present unique challenges due to high initial costs, potential for failure, and benefits that may not be immediately realized. Regularly revisiting and adjusting modelled values based on real-world project performance is crucial. As earlier chapters emphasized, understanding implementation costs, ongoing costs, benefits, and risks will significantly influence the chosen time horizon and discount-factor calculation.

3 Methodology

This section sets out the structured approach for the empirical part of the study. The goal of the empirical part is to gain an understanding of the factors influencing the NPV of an investment in data warehousing and how they are perceived by experts in this field at the current time.

3.1 Research Design and Approach

In understanding the nuances of data warehousing investments, particularly their influence on NPV across varying firm sizes, this study will embrace a qualitative research design based on a case study approach as outlined by Eisenhardt (1989).

In her studies, Eisenhardt (1989) gives detailed instructions on the best practices of conducting case study based research. This study acknowledges these steps. Prior to data collection or conducting interviews, it is imperative to define the research question and a priori constructs.

Research Question The following question shall be answered:

How does firm size influence the perceived NPV of investments in data warehousing?

This question was chosen since it was discovered in the Literature Review part of this study that some factors are not yet as well explored as others. Particularly the size factor appears fruitful to be further investigated. Therefore, research on its influence on the NPV of an investment in data warehousing, amongst other factors, shall be conducted.

A Priori Constructs Eisenhardt (1989) is advocating for a balanced approach in case study research. While researchers should be open to emergent themes and unexpected findings, they should also have some predefined constructs based on existing knowledge to guide and ground their investigation. Five emerging factors were already identified, influencing the constituents of the NPV, and providing a good structure to base the case study on. While based on the literature review, one might have theories on the interrelation between these topics, one wants to "avoid thinking about specific relationships between variables and theories as much as possible, especially at the outset of the process", as Eisenhardt (1989) notes, to avoid bias and to retain theoretical flexibility.

Next in the process, a population should be defined from which the research samples - the cases covered in this study - are drawn.

Population and Case Selection To identify interrelations between the firm size and the other four emerging factors and the NPV constituents, the population shall entail reviews by professionals working on implementations of data warehousing on a daily basis. With this expertise, conclusions and concepts can be drawn from experience.

Furthermore, it should be guaranteed that other factors that might significantly impact the results are constrained. Only cases among established implementations will be selected to ensure that findings are grounded in real-world experience rather than theoretical constructs or beliefs.

As mentioned earlier in the study, through the emergence of new technologies, like cloud computing, the data warehousing landscape underwent a shift. Even though factors influencing the successful outcome of data warehouse implementations were already sufficiently researched on, they are mostly based on a technological environment from decades ago. The focus will be solely on timely implementations, thus limiting the population to a time horizon of the last five years for the commencement of the implementation.

The constraint to limit the population on a specific industry would make theoretical sense, but is practically difficult to arrange. Overly limiting the population would make a generalization of results difficult and given the nature of the investigation, the concept of data warehousing as a general, non-industry specific practice, such restriction is not appropriate.

As Eisenhardt (1989, p.537) emphasizes, "it makes sense to choose cases such as extreme situations and polar types in which the process of interest is 'transparently observable'". To shed light on the size factor among the five emerging factors, this study focuses on cases that are particularly binary concerning these factors. It therefore makes sense to look at firms that are either particularly small or big, used mainly on-premise vs. mainly cloud computing solutions, outsourced the implementation vs. trained an implementation team in-house, are particularly GDPR aware or unaware, and used either the Inmon (2005) vs. the Kimball et al. (2008) implementation approach. In that sense, cases are not chosen randomly, but based on the a priori established constructs, the theoretical approach.

3.2 Data Collection Methods

In the exploration of data warehousing expertise, a population of 50 professionals specializing in this domain was identified through LinkedIn.com. These individuals were meticulously selected based on their demonstrable work experience in the field of data warehousing. It is noteworthy that the selection criteria did not impose any geographical limitations, resulting in a diverse group of experts hailing from various countries.

Upon reaching out to the identified 50 data warehousing professionals on LinkedIn.com, only a subset responded to the inquiry. Out of those who responded, just two experts were deemed suitable for the study based on the detailed criteria and constructs previously established. Their selection was determined by their alignment with the desired parameters: their involvement in recent implementations of data warehousing (specifically within the last five years) and their association with either extreme ends of the predetermined factors, such as firm size, technology usage (on-premise versus cloud computing), approach to implementation (outsourcing versus in-house training), GDPR awareness, and adherence to either the Inmon (2005) or the Kimball et al. (2008) implementation approach.

It's pivotal to underscore that while this might appear as a small sample size, the intent was to ensure a comprehensive exploration of cases that transparently demonstrated the parameters of interest. Such a focused approach is in alignment with Eisenhardt (1989)'s emphasis on the value of selecting cases where processes are "transparently observable," with the aim to derive meaningful insights grounded in genuine, hands-on experience rather than broad generalizations.

The Case Studies This research integrates two case studies, each providing distinct perspectives on data warehousing implementations from experienced professionals.

The first case involves a professional with a title of "Lead Cloud Consultant," focusing on Data Engineering and Warehousing. This expert is affiliated with a notable German IT consultancy firm known for its services in Cloud, DevOps (Development and Operations), and Data Engineering. With nine years in the field, this consultant played a pivotal role in leading the implementation team on a data warehousing project for a healthcare company. The insights provided stem from both his accumulated experience and specifics from the aforementioned project.

The second interviewee has more than 25 years of experience, 24 of which are with the same company where he helped implement the data warehouse in multiple iterations. The company is providing insurance services for more than 100 years and acts in multiple countries globally. Given the time with the company and being involved over such a long period with the same data warehouse implementation makes this participant's insights valuable.

The Questionnaire Prior to reaching out to the potential participants, the questionnaire was created. In Appendix A the questions, and also the answers of both participants can be viewed.

The questionnaire was designed to understand which factors the expert deems as important for a data warehousing implementation and how they perceive the impact on the NPV of the data warehouse implementation. It is worth mentioning that in order to keep flexibility in the answers, the initial opening question was asked on "which Factors [were] deem[ed] important for a data warehousing implementation". After this initial question, the following questions ask specifically about the perceived importance and the impact on the NPV of the five emerging factors.

After this general part of getting insights into the experience of the expert, the following questions ask specifically about the (last) data warehouse that the expert was involved with. Question 17 asks about the size of the firm in terms of employees, where a scale according to the study of Ramamurthy, Sen, and Sinha (2008) was applied. Question 22 was specifically designed to keep the interviewee open in his answers while also touching on previously mentioned aspects. It ensured that there were no angles left untouched that were deemed important by the interviewee.

Questions 23 until 39 were taken from the framework as introduced by Wixom and Watson (2001). They ensure to cover other factors at play that could influence findings from previous answers. The reason why this exact structure with the same wording was chosen is comparability. Since there are quantitative results from Wixom and Watson (2001), one can draw conclusions based on his findings and combine them with the insights of this study.

Questions 40 to 44 make the perceived NPV and its constituents quantifiable by asking about ex-post fulfilment of expectations.

The last question was put last because it falls into the NPV section but cannot easily be quantified. It aims at gaining an understanding of the approval process of the data warehouse implementation through the applied discount rate, also as a part of the NPV calculation.

The approach to mix qualitative with quantitative questions was inherited from Eisenhardt (1989)'s recommendation on the synergy between these two methods. By leveraging both, this study aims to harness the quantitative strength in identifying potential patterns and relationships while using qualitative data to delve deep into understanding the rationale and underlying theories.

The Interview Process After initially reaching out to each participant to confirm their participation in an interview about trends in data warehousing, the interview sessions were scheduled. Each interview included only the participant and the interviewer. During the interview, the interviewees were first briefed on the topic of the thesis before starting to fill the questionnaire together with the interviewer in a Microsoft Teams meeting. The briefing did not outline the factors that the questionnaire focuses on just yet to avoid bias.

During the interview, field notes were taken that noted down the most important findings and interrelations between factors. This enabled to steer the second interview into a direction of interest without changing the initial format.

This concept is called iterative questioning, and builds on Eisenhardt (1989)'s insights.

It emphasizes the value of flexibility in data collection, especially in theory-building research. This approach, while offering richer insights, requires careful balance to ensure the broader research goals remain intact. During the interviews, the line of questioning was adjusted based on the first participant's answers to delve deeper into emergent themes. For example, the second interviewee was asked after his initial response to question nine, whether he thought about anonymisation as a means of ensuring that information is not spread as easily. He then went on to give insights into potential threats that can emerge through the employees, enhancing the insights.

3.3 Data Analysis Techniques

In line with Eisenhardt (1989)'s approach to theory-building, this research adopted a dynamic process of shaping hypotheses. The two case studies were analyzed using iterative comparisons of emergent frames with case evidence. Through this meticulous process, potential themes, constructs, and possible relationships between variables began to surface. The objective was to ensure that the derived theoretical framework was closely aligned with the empirical data collected.

This method encouraged the continuous juxtaposition of theory and data, refining constructs and sharpening their definitions. Such an approach, despite being more judgmental due to the qualitative nature of data, leans heavily on the authenticity and depth of insights. The absence of a vast array of cases did not deter the efficacy of the research; rather, it allowed for a more focused and transparent analysis. This iterative comparison and hypothesis-shaping ensures that the emergent theory is both empirically valid and robustly grounded.

Triangulation After the best practice provided by Eisenhardt (1989), triangulation serves as a means of validation. By cross-referencing interview data with information from annual reports, websites, news articles, and other sources, this method ensures the reliability and validity of the data collected. Some statements given by the participants were cross-referenced and validated using data and information published online. A deeper examination of the sources and implications will be presented in the discussion section.

Comparative Case Analysis To fully appreciate the depth and breadth of the findings, the two case studies were juxtaposed, underlining both their congruities and disparities. To encapsulate these observations, a comprehensive table was crafted, rooted in Eisenhardt (1989)'s guidelines, presenting the evidence underlying each construct in a structured manner.

Hypothesis Verification As hypotheses took shape, they were diligently evaluated against each case's empirical evidence. Where evidence reaffirmed the relationships between constructs, the hypothesis was strengthened. In instances where the case data diverged from the hypothesis, it was not perceived as a flaw. Instead, these divergences were seen as catalysts, prompting refinements to, or extensions of, the budding theory.

Engaging Broader Literature The expansive landscape of related literature served a dual purpose in this research. First, by juxtaposing the study's findings against opposing literature, it forced a deeper introspection, ensuring that conclusions were not drawn hastily. Ignoring such literature would jeopardize the study's credibility. Simultaneously, literature echoing similar findings was brought to the fore, offering additional weight to the study's conclusions and weaving it seamlessly into the existing tapestry of knowledge.

4 Discussion of Results

In this section the results obtained from conducted research and interviews shall be discussed. To begin with, a comparison between the two cases is necessary to establish an understanding of the different situations at hand. Subsequently, this section looks at already well established factors based on the literature regarding data warehouse implementations and implications from the interviews. Then, focus is laid on the emerging factors identified in the literature review and the findings from the conducted interviews to commence further investigation.

4.1 Case comparison

Table 1 shows a comparison of both cases from a meta perspective. The two cases are different from multiple perspectives, but especially in terms of the factors identified worth researching in the Literature Review. While participant 2 was involved for over 20 years with the same data warehouse in the same company, his insights come from an internal view of a company that implemented a data warehouse. On the other hand, participant 1 comes from the perspective of an external consultant giving valuable insights into practices and challenges from a broad spectrum of different implementation experiences.

While the implementation chosen by participant 1 concerned a German healthcare com-

	Case 1	Case 2
Narrative view	External Consultant	Internal Employee
Industry	Healthcare	Insurance
Number of employees	801-1,500	24,000 (3,001-6,000 in DE)
Outsourcing Degree	86%	20%
Cloud Usage Degree	100%	20%
Time Frame	3 Months	18 Months
Completion Date	(11.10.2023)	01.07.2020
% Completion	71%	100% (71% in overall project)
Architecture	Data Mesh, Inmon (2005)	Kimball et al. (2008)

Table 1: Meta comparison between Cases 1 and 2

pany in the SME league, participant 2 works for a global player in the insurance field with more than 24,000 employees worldwide. Since the implementation of the data warehouse was only concerning Germany, interviewee 2 guessed the affected employees in the range of 3,001-6,000 - which is the number of employees working for this company in Germany. Not only from the number of employees these two cases differ, but also among the degrees of outsourcing and cloud service usage.

The healthcare company used mainly outsourced expertise, whereas the global insurance provider could resort to internal experts. A fact check with web searches revealed that the insurance provider indeed employs multiple experts in different fields, among those

- IT data warehouse Developer (25 years of experience in multiple positions)
- Digital Transformation Lead (15 years of experience in multiple positions)
- Cloud Developer (ten years of experience in multiple positions)

while also making use of external experts, especially in the cloud development field. A fact check regarding the healthcare company was not possible as the name of the company was not revealed to the researcher.

Time-wise the implementation took less time for the SME compared to the insurance provider. While case 1 was at the time of the interview (07. September 2023) not yet complete, case 2 already completed on the 1. July 2020. This shows that both cases are within the targeted five years as discussed in the Methodology section.

The phase of the implementation can play a major role when looking at different factors for data warehouse implementation success and the implications on the NPV. Mungree, Rudra, and Morien (2013, p.8) note that "building on the findings, [...] the relevance and importance of the different factors vary according to the phase of the implementation. [...] Some particular factors are most important when the project gets started (e.g management support, defining scope, committed and informed executive business sponsor), while some other factors become vital in the following stages of the project (e.g. user oriented change management, adequate resources)". Since both of the cases have already reached a late stage of the implementation, the comparability and comprehensive view on the project as a whole can be guaranteed.

An interesting aspect was revealed when participant 2 stated 71% completion of the data warehouse, even though he earlier stated that the completion date was already in the past. This seemingly paradox statement was resolved by his explanation that the data warehouse implementation consists of many phases and that a data warehouse is an ongoing project that will never be 100% completed. Since the start of data warehouse with an on-site solution in the early 2000's, many iterations were done in extending and amending the data warehouse structure. This was also found to be reflected in the yearly financial report, stating the "expansion of a data warehouse" for each of the past several years.

Case 1 deployed a very novel implementation approach, called "Data mesh", which is similar to the Inmon (2005) approach, according to participant 1. Machado, Costa, and Santos (2022, p.264) describe this architecture as one, in which "data is organized into domains, [...] data teams manage themselves and carry out their own work in an agile and product-oriented way". It would be beyond the scope of this thesis to dive deeper into this type of architecture, but it can be concluded that research on the data warehouse architecture approach is still in turmoil.

4.2 Findings on Established Factors

The second question of the questionnaire asked openly about the perceived most important factors for a successful data warehouse implementation. The important point is that the participants were not shown any research on this topic as a briefing before the interview. Therefore, the answers reflected the truly "top of mind" factors from the experience of both participants.

Vision Interestingly, both participants named the Vision factor as crucial for a successful data warehouse implementation. The first participant gave examples such as "the need for better data insights, and modernization as means of staying competitive". While these examples are valid, participant 2 added that the firm size affects the granularity of these goals: "The bigger the company, the more sure it has to be what exactly [it wants] to achieve with the implementation". As an example for a specific use case, he named "sales reporting". The comparison between both answers regarding the granularity of the example confirms this distinction. Both answers from the participants underscore the concept of Yeoh and Koronios (2010), thinking of Vision as a meta-CSF.

Culture Both participants also named the firm culture as an important factor. While participant 1 states that the organization needs to be changed and trained, participant 2 notes that the culture can be a risk, when departments are asked to share their data. He elaborates that the data warehouse can make issues in the organization transparent and managers could go against that. Even though the culture factor is by this name not directly covered in the literature, it ties closely to other operational effectiveness factors such as management support, user participation, and change management. Whether the management stands behind the decision of implementing a data warehouse has a big impact on the motivation of the users who should later work with it. Similarly, the engagement in the implementation process furthers understanding for the changes and leads users to value the benefits. Looking at these answers, Wixom and Watson (2001)'s results on the impact of management support and user participation on the organizational and project implementation success are underscored, therefore demonstrating their timelessness.

Culture and inertia are deeply intertwined concepts when it comes to organizational change and adaptation. The firm culture, as highlighted by the participants, plays a pivotal role in how an organization responds to new initiatives, such as the implementation of a data warehouse. Ghemawat (1991)'s concept of inertia underscores this by suggesting that organizations inherently resist change due to their built-in biases. This resistance can be attributed to cultural aspects like fear of transparency or the reluctance of departments to share data. Therefore, understanding and addressing cultural barriers is crucial in overcoming organizational inertia and ensuring successful implementation of new systems or strategies.

Source Systems Participant 1 also named data quality as an important factor for success. He confirmed that an organization running solely on Microsoft Excel tables will not be able to implement a data warehouse to reap benefits, or only with mayor difficulties, due to the inconsistencies. He gave an example of a small company that wanted to make use of AI but did not even have an ERP system to use as a data source, solely relying on Microsoft Excel. While a solid data foundation is necessary for advanced analytics and building machine learning models, not even a data warehouse could be constructed on the basis of their current conduct. In the light of organizational strategy, this highlights the concept of "Lock-out" as described by Ghemawat (1991). Before being able to make use of the latest technological developments, the organization first has to establish more solid practices and systems. This delays the development, therefore giving competitors that committed to these topics earlier a competitive edge.

Resources Participant 1 furthermore mentioned Resources as important, and specifically that implementations from his experience most often lacked sufficient time to complete the project. This also reflects in question 29, where he answered two out of seven on agreeing to "The data warehouse project was given enough time for completion". In comparison, participant 2 mentioned that "time was rarely an issue" giving the same question a seven out of seven. This hints on the conclusion that outsourced consultants generally deal with heavier time constraints than in-company developments. However, this finding might be firm-specific and can, due to the small sample size, not be confirmed. Investigation in the literature also did not yield any results to this topic.

Regarding the resource capital, participant 1 noted that a proof of concept for a data warehouse can be provided nowadays for a low four-digit euro amount, but he still regarded capital as an important factor for a successful implementation. Furthermore, he compared the importance with the size factor, which the following subsection will focus on.

4.3 Findings on Emerging Factors

From question three throughout twelve, the participants were asked on their perception of the importance of the five emerging factors - as introduced in the Literature Review part of this study - and their impact on the NPV of the data warehouse implementation.

4.3.1 Firm Size

According to participant 1, the firm size is not as important as capital and management support. This statement aligns with the answer of participant 2, who sees the factor not as a direct factor of success but as an influence on other success factors. Throughout the course of both interviews, this relation was discovered with multiple factors. In figure 5 the relations between the focused factors and each other, as well as the NPV of the data warehouse implementation are mapped. As an example of the relationships of the factors in figure 5, the previously discussed vision factor and the effect on the discount factor is shown in blue. Each connection line follows the bent angles until it arrives at its destination factor or NPV component.

Before diving into the relationships it is worth emphasizing that the two participants' statements by themselves are opposing the reviewed literature, in particular Ramamurthy, Sen, and Sinha (2008)'s findings that organizational size in terms of number of employees is significant and positively correlated with data warehouse adoption. This perspective will be evaluated by analyzing the relationships between the firm size and other factors.

NPV: Time Horizon Looking at the implications of the firm size towards the NPV of an implementation, one can see a difference of many months in the time horizon between the two cases. As learned from Ramamurthy, Sen, and Sinha (2008), the usual time frame for an implementation of a data warehouse should be around 4.46 person-years. Based on the indications by participants, this duration can deviate significantly with the size of the implementing company with only three to six months for the SME and 1.5 years for the big insurer. Both participants state that these values are already a bit over the expected time horizon, but nearly as expected. Participant 1 ensured that there must be a correlation between size of the company and the complexity and therefore time horizon of the implementation. Triangulation with experience data from a data warehouse solution provider yields three to twelve months (ScienceSoft 2023), based on a team of seven experts.

While more people might not necessarily shorten the time frame of the implementation, the number of people working on the implementation can have an impact on its speed. The participants were not asked how many experts were involved in either case. Furthermore, other factors as cloud computing usage, implementation approach, and outsourcing that differentiate between the cases makes them hard to compare on this dimension. For example, could the fact that outsourced consultants are put under more time pressure than internal resources be a mayor influence on the implementation time horizon.

Participant 1's statement, however, that smaller companies tend to have fewer systems to be integrated in the data warehouse solution, also speaks for the fact that the time horizon is indeed shorter. Ghemawat (1991)'s concept of lags, which refers to the temporal delays organizations experience when adjusting to desired levels, can be applied to this context. Larger firms, with their intricate systems and processes, might face longer lags due to the inherent challenges in integrating multiple systems into the data warehouse solution. On the other hand, smaller firms, with fewer systems to integrate, might experience shorter lags, allowing for quicker adjustments. This temporal difference in implementation, influenced by the firm's size, aligns with Ghemawat (1991)'s idea that organizational adjustments are subject to delays based on the stickiness of certain factors.

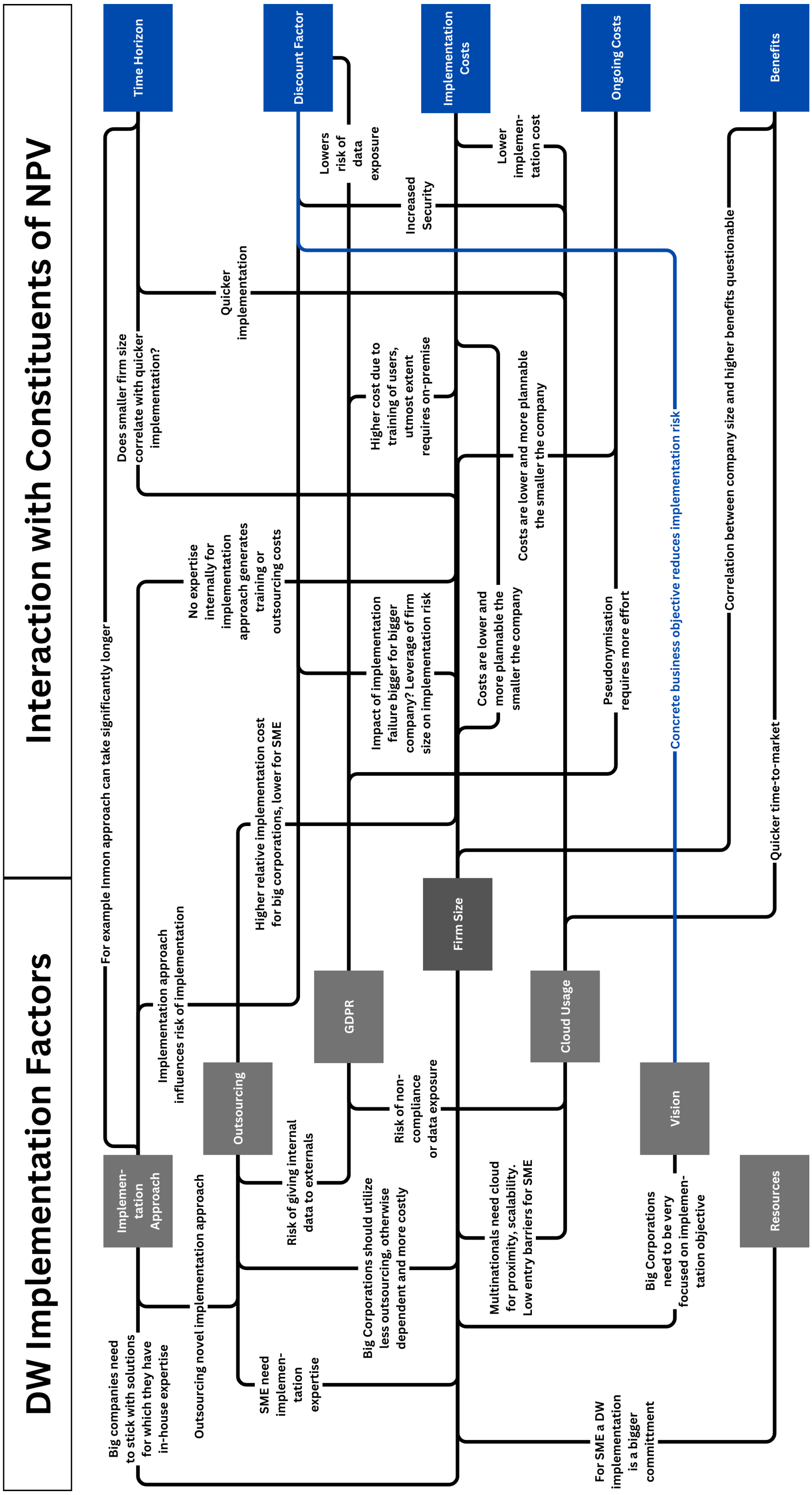


Figure 5: Interaction of Emerging Factors with each other and with NPV.

NPV: Benefits From a high-level perspective, larger corporations indisputably possess more data. This sheer volume of data invariably offers a granular depth of insights. As participant 1 accentuated, while big corporations might have to shoulder higher implementation costs, they have the potential advantage of drawing richer, more detailed insights from their vast data sets. This granularity can inform diverse business decisions spanning marketing, operations, customer engagement, and more. The driver of value, as observed, is a blend of the benefit derived and the amount of information, sometimes aggregated from various functional areas.

Contrarily, participant 2 posits that the benefits of a data warehouse are not contingent solely on the size of the firm. He underscores the core purpose of a data warehouse: to enhance decision-making validity. Indeed, the core objective of any data warehouse implementation, be it for an SME or a multinational, is to refine and support more informed, data-driven decisions. Yet, there is a vital angle to consider: the balance between granularity and cost.

A central question emerges: does a larger data volume equate to superior insights? It's essential to dissect the benefits of data warehouse across different areas and critically assess how firm size might influence each. Notably, insights represent just a facet of the data warehouse advantage spectrum. Each implementation, regardless of company size, needs a defensible business rationale. Thus, comparing the investment and benefits between different use cases can prove complex.

In terms of monetary benefits, larger corporations might leverage their data warehouse insights to capitalize on newer sales opportunities, mitigate customer churn, optimize operations, and more. The potential scale of revenue and savings might be proportionally larger, given their market reach and operational size. But juxtaposing this with the investment required, it's pivotal to ascertain if the benefits scale linearly with the associated costs.

However, non-monetary benefits often blur the firm size demarcation more than monetary benefits. Cultivating a culture of growth and innovation through data-driven insights or enhancing decision-making processes can be as applicable to an SME as it is to a global conglomerate. The creation of a unified data source, swift data reconciliation, or the strategic leverage in contract negotiations due to enhanced market insights are attributes that companies, regardless of size, can harness.

NPV: Implementation- and Ongoing Costs As both participants agree, the risks and costs in general of a data warehouse implementation are lower and more planable for a SME compared to a big corporation. This creates a "leverage", as participant 2 puts it, for the size of the company and resembles the "lock-in" after Ghemawat (1991) of such undertaking. For big companies, more stakeholders are involved in the process, more systems need to be integrated and a higher volume of data is handled on a daily basis. The size factor furthermore effects the costs generated and influenced by other factors. These aspects will be addressed in the subsequent sections.

NPV: Discount Factor As learned from the examples of the influence of culture and vision on the successful outcome of data warehouse implementation, the risk heightens when these factors are neglected. The implications of such neglect are further amplified in larger corporations due to the scale of operations and financial commitments involved. For instance, the granularity of vision, as pointed out by the participants, varies with firm size. Larger firms need a more defined vision due to the multiplicity of stakeholders involved. A nebulous vision or its poor execution could lead to a fragmented implementation, multiplying the risks. Similarly, cultural challenges, such as departments' hesitancy to share data, can be more pronounced in large entities where siloed operations and politics may be more prevalent. Participant 2's observation that a smaller company might waste 200k € if the data warehouse isn't adopted, while a larger corporation could potentially lose 20M €, underscores this point.

When considering the discount factor for NPV calculations, this disparity between small and large firms necessitates adjustments. Larger firms, due to their higher exposure to absolute monetary risks and the complexities introduced by their size, might require a higher risk-adjusted discount rate compared to SME. In essence, the size of a company doesn't just scale the potential monetary losses linearly, but it introduces layered complexities that could exponentially increase risks. As such, when adjusting the discount rate to factor in these risks, it's crucial to consider not only the direct financial implications but also the indirect risks posed by the multifaceted challenges larger corporations face.

When determining the discount factor for NPV, it was established that due to the individuality of data warehouse projects, one should begin with a base rate, such as the weighted average cost of capital, and then adjust for project-specific risks. As per the insights of participant 1, the base discount rate should be aligned with industry standards, given the distinctiveness of sectors like healthcare and associated high project returns. The data warehouse project would therefore need to provide a internal rate of return similar or preferably higher than other projects. Participant 2 suggested a different approach leaning on forward-planning the dividends of the stock. This approach is obviously not possible for a SME since the stock is not traded publicly and therefore dividend returns cannot be determined. Both approaches are valid and gave good insight into industry practice.

4.3.2 Outsourcing

As Ramamurthy, Sen, and Sinha (2008) argue, bigger firms have more abundant resources to cover the substantial data warehouse implementation costs. Therefore, an implementation is a bigger financial commitment for smaller companies - as also participant 1 notes in the interview. At the same time, SME are often too small to acquire internal expertise on data warehousing, therefore being reliant on outsourcing. This statement was confirmed by both participants.

While outsourcing is generally more expensive on an hourly basis, the absolute costs compared to acquiring in-house expertise is insignificant. As participant 1 notes, a trained data warehouse expert usually costs around six figures in yearly salary and above. This statement was triangulated with salary data from Glassdoor (2023), where an average around €120,000 in yearly salary for a data warehouse architect in Germany emerged as the current standard. Compared to an average salary of €150 hourly rate for a data warehouse consultant (Upwork 2023), it shows that outsourcing has a lower impact on implementation cost, especially since the cost increments are smaller.

Participant 2 introduces another significant point. Considering a corporation that has not yet invested in data warehouse staff or if this staff is busy with maintaining previously established infrastructure, the implementation costs are much higher than if utilizing outsourcing services, since new staff needs to be hired to conduct the implementation.

The next statement by participant 2 reveals a complication: Outsourcing the implementation on a large scale can create a dependency towards the outsourced solution provider. At worst, this will create issues regarding the previously covered User Participation factor. The danger is that the users will not stand behind the project at last, letting the implementation fail. To conclude, bigger corporations should try to utilize less outsourcing and more internal staff.

The validity of this concept can furthermore be confirmed when looking at the responses of both participants to the extend of outsourcing used in each case. While for the SME, participant 1 chose six out of seven in terms of usage of outsourced services, the big insurer only uses these for around 20% and lets internal resources do the main work.

4.3.3 Implementation Approach

Utilizing internal staff to implement a data warehouse requires the staff to be familiar with current best practices and frameworks. Often the ones making the decision about which approach will be chosen are not the same as the ones implementing it. According to participant 2 this can create problems of technological knowledge gaps requiring either outsourcing or internal up-skilling. This will not only increase implementation costs, but also increase the risk of failing the implementation due to lack of knowledge, which was approved by statements of both participants. For this reason, companies should decide to go with an implementation approach that the internal staff is already familiar with and keep staff trained concerning new emerging technologies.

The interviews furthermore shed light on which implementation approach is preferred in practice. Both interviewees agreed that the Inmon (2005) approach is generally more complex and therefore more risky and rarely done in practice. Also the triangulation with a comparative study between the approaches shows that this approach shows disadvantages compared to other implementation approaches (Yessad and Labiod 2016). Central to its disadvantages is its undivided commitment to a single, extensive project. This can raise concerns regarding project management, resource allocation, and risk mitigation. When embarking on a data-driven approach, as proposed by Inmon (2005), there isn't a clear division into smaller, more manageable sub-projects. Instead, the entire enterprise is viewed as a monolithic venture. As Yessad and Labiod (2016) pointed out, the Inmon (2005) method involves loading company data without prior knowledge of user requirements. This can be perceived as a risky venture since any misalignment between the data loaded and eventual user needs could result in significant wastage of resources. A lack of phased development could also mean that any issues or challenges are only discovered late in the project, making them more costly and time-consuming to address. In the realm of data warehousing, where the landscape is vast and multifaceted, this lack of granularity in project segmentation can be a pivotal factor in the decision-making process for organizations. As participant 2 put it, "if you are trying to eat the whole pig at once, you are going to choke on it".

Importantly, the leverage of risk discussed in the firm size factor part of this chapter is furthermore depending on the implementation approach. If an organization decides to implement a single data warehouse that connects with all departments, this becomes a bigger risk than if multiple Data Marts are built that can each be changed individually. Since SME companies are mainly using outsourced providers, which themselves keep up with the newest innovations, the "lock-in" according to Ghemawat (1991) of internal resources is of lower concern to them. However, since the architecture of the implementation of a data warehouse is not changed easily, they too are subject to the effect of "lock-in" when deciding on one approach or the other.

4.3.4 Cloud Usage

The word "cloud" was one of the most used words in both interviews, coming up on average eleven times by each participant. The reason for that is that it changed the data warehouse environment drastically over the last years. Compared to "outsourcing" which came up on average twice, it hints on the persistent importance of this topic.

Both participants acknowledged the importance of cloud computing in the context of data warehousing, albeit from slightly different perspectives. Participant 1 emphasized its relevance for both SME, due to the inherent benefits of flexibility and scalability, and for large corporations where managerial budget constraints can lead to a preference for cost-effective proofs of concept. The mentioned hypothetical budget of a low four-digit euro amount for a proof of concept underscores the notion that, from a budgetary standpoint, the cloud can offer economical solutions for testing and implementing data warehousing initiatives. Participant 1 emphasises that cloud lowers the entry barriers to build a data warehouse because no physical hardware needs to be bought.

Participant 2, while also highlighting the importance of cloud for SME, nuanced their perspective by indicating that SME have the choice between cloud and affordable on-premise solutions. However, for global companies, the situation is quite different. The logistics and expenses involved in establishing multiple on-premise data warehouses in each operational country can be prohibitive. In this context, the cloud becomes indispensable for multinationals, not only due to the benefits of scalability but also because of the necessity to maintain proximity to data sources in various locations.

When examining the hypothesis that a rising number of firms are transitioning to cloud storage, data from Statista (2022) supports this assertion. Their survey delineated that 60% of corporate data globally is now stored in the cloud, a noteworthy increase from 30% in 2015 as seen in figure 6.

Adding texture to this narrative, participant 1 confirmed that the healthcare SME too had transitioned from an on-premise data warehouse to a fully cloud-based solution. This real-world example confirms the cloud's pragmatic benefits being realized by smaller-scale enterprises.

Contrastingly, participant 2, speaking from the viewpoint of the global insurance provider, acknowledged the industry's gravitation towards cloud solutions but maintained some

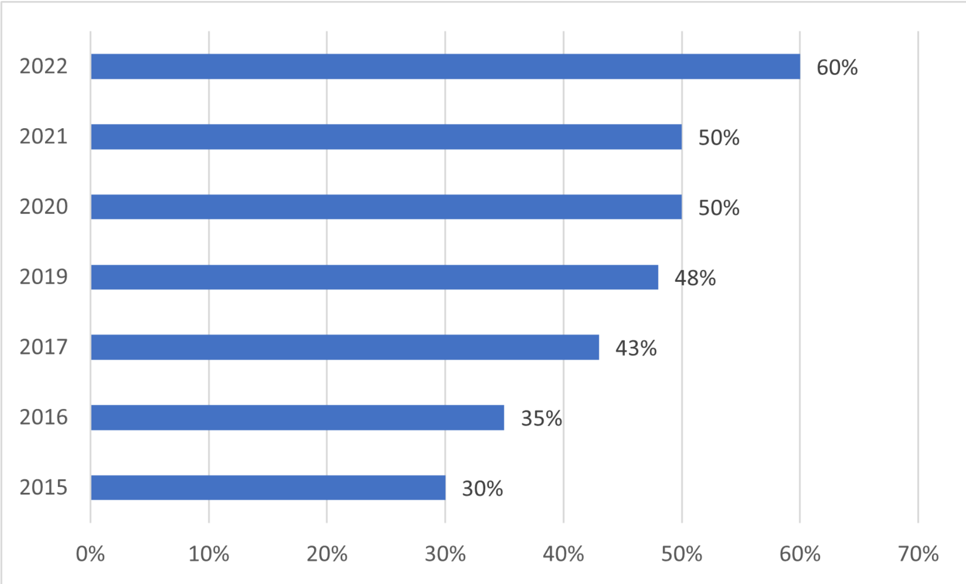


Figure 6: Share of corporate data stored in the cloud in organizations worldwide from 2015 to 2022. Source: Statista.

reservations. Their firm is in the exploratory phase, currently testing the waters with an AWS cloud computing proof of concept. Speaking for the German part of the enterprise, they still lean heavily towards on-premise solutions, with an 80-20 split between on-premise and cloud. A notable observation from participant 2 is the assertion that the cloud "hype" might be more of a managerial trend rather than an operational necessity. This perspective suggests that while the cloud's advantages are undeniable, the decision to migrate might sometimes be influenced by prevailing industry trends rather than purely strategic considerations.

Drawing further on the insights provided by participant 2, there emerges an intriguing contrast in the projected trajectory of cloud costs. While cloud solutions initially entered the market with a promise of affordability, several factors indicate this might not remain a constant. Participant 2 accentuates potential catalysts for this shift, notably the looming energy crisis and the associated workforce's escalating costs to maintain expansive cloud infrastructures. These insights parallel findings from TechTarget (2023b) which suggest that cloud costs are on the rise.

Indeed, the affordability of cloud computing, as participant 2 suggests, can be attributed to its current stage in market adoption. With fewer organizations initially utilizing it, costs remained relatively low. But as its prevalence grows, and as more organizations lean on cloud solutions demanding higher availability, there is a logical expectation of increased costs. This claim is further solidified by recent moves from industry giants. For instance, Microsoft's latest decision to hike the prices for Azure becomes particularly noteworthy (USCloud 2023).

On the technological front, the affordability narrative also intertwines with global manufacturing trends. Participant 2 points to the current low-cost production in China and notes that these costs, while low now, might not remain on the same level as the country develops. Shifts in geopolitical climates, labor costs, or trade policies could drastically influence these production costs, which would, in turn, impact the cost structure of cloud solutions.

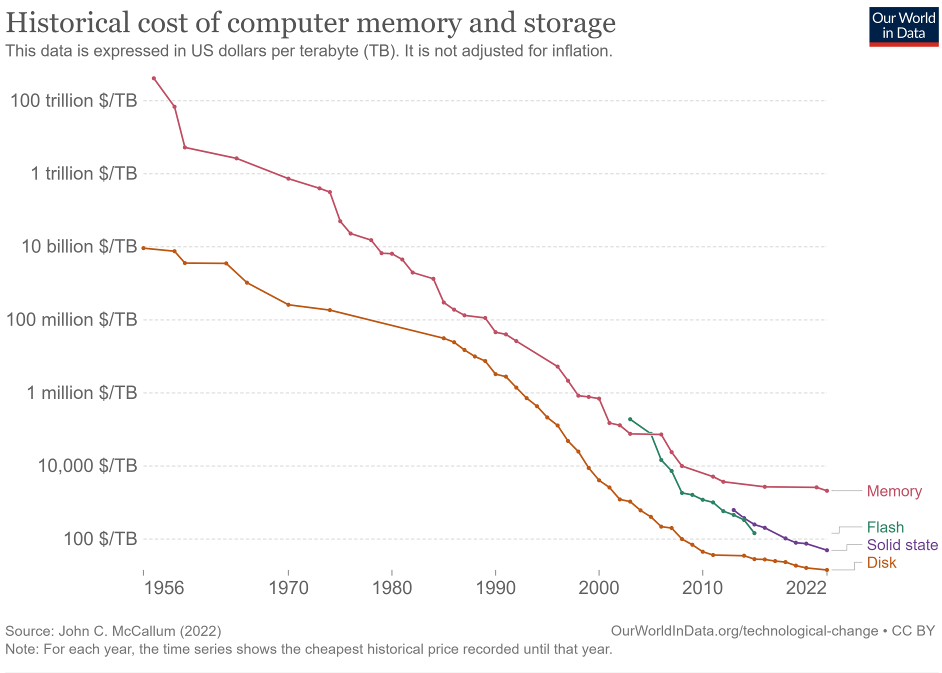


Figure 7: Historical cost of computer memory and storage in $ per terabyte, not adjusted for inflation.

However, juxtaposing this with the historical trend of storage costs paints a somewhat different picture. As depicted in the graphic 7 (OurWorldInData 2023), the cost of computer memory and storage has seen a consistent decline over the years. While this downward trajectory in hardware costs is undeniable, it doesn't necessarily insinuate a direct correlation with the holistic costs of cloud services. In fact, the divergence between rising cloud costs and falling hardware costs has been emphasized in an article by TechTarget (2023a).

An additional layer to this discourse is the strategic positioning by cloud service providers. Participant 2 alluded to the possibility that the present-day cost advantage of cloud services might be a calculated move, driven by political or market-share capturing intentions rather than sustainable pricing. Such tactical pricing strategies, while effective in the short-term, might evolve as market dynamics shift and as providers solidify their dominance.

Organizations also need to acknowledge the reliance that a cloud migration brings towards the provider, as participant 2 pointed out. The financial statements of the global insurance company from 2022 provide additional evidence that underscores this point. While the company recognizes the augmented cyber security standards associated with cloud usage, they are simultaneously ramping up their IT management to effectively mitigate the risks that arise with increased cloud reliance.

The NPV implications of cloud usage is clear. While implementation costs can be lowered significantly due to the omission of hardware purchasing, at the current time even ongoing costs can be lowered. This position is, however, unsure to be continued due to the reasons given above.

At the same time, the time horizon can be significantly shortened: Participant 1 noted that the first benefits can already be seen as soon as two months after the project kick-off. This has two implications: Firstly, the benefits come in earlier and are therefore discounted less. Secondly, the time-to-market is reduced, allowing the company to launch products faster and therefore generating revenue earlier.

Another point, as participant 2 posits, are the higher security levels that can be maintained, lowering the risk of data breaches and therefore lowering the discount factor. This is also mentioned in the yearly financial report of 2022 of the insurance company, stating that the company is increasingly using the cloud to take advantage of higher security standards.

4.3.5 GDPR Awareness

While cyber security standards might be heightened with cloud usage, this does not necessarily reflect in the same way on data governance. One of the notable concerns stemming from data governance, particularly in the cloud context, is the potential non-compliance with GDPR regulations, especially when data storage locations span multiple jurisdictions. The ubiquity of major cloud providers, such as AWS, often implies that data might be stored in servers located in the US or other non-EU regions. Such geographical data flows raise substantial GDPR concerns. However, it's pertinent to note that providers are increasingly becoming conscious of these implications, which is evident from AWS's decision to offer specific server locations, such as in Frankfurt, to allay such fears and ensure adherence to the GDPR framework. Participant 1's insights corroborate this strategic shift by cloud providers.

While both participants concurred on the overarching importance of GDPR, their emphasis varied slightly. For instance, participant 1 expressed that GDPR acts as a deterrent. He states that organizations, especially larger ones, have to be transparent about data sources, data access, and must often justify data-driven decisions. Such meticulous data governance often acts as an inhibitory force, making stakeholders apprehensive about potential legal repercussions, especially in light of instances where they might be held accountable. This is particularly pronounced in industries like banking, where multiple reasons might have to be provided for decisions such as credit denials. This account accentuates the profound implications GDPR can have on the decision-making processes of firms. Especially concerning is the insight by participant 1 who states that some banks are already moving away from the cloud and back to on-premise data warehousing solutions due to stricter GDPR requirements. While such occurrence could not be backed by articles found online, an article by RetailBankerInternational (2023) confirms the concerns in the industry.

Participant 2 also acknowledges the relevance of GDPR but adds a nuance, suggesting that while GDPR is significant, it might not be the primary determinant of data warehouse implementation success. This variance in perspective underscores the fact that GDPR's importance might be relative to the industry or the specific operational framework of an organization. Regulation is in general stricter concerning industries that deal with particularly protected data such as personal data.

Delving into the intricacies of GDPR compliance, participant 1 offers a multi-faceted approach to addressing these concerns:

1. On-premise solutions, while being expensive, offer the highest assurance in terms of data usage. However, they bring along inherent cyber security risks.
2. Opting for cloud solutions with not fully pseudonymised data storage. While cost-effective, they can be optimized to ensure data storage in GDPR-compliant regions, such as AWS's EU Central servers.
3. Employing cloud solutions but restricting data storage to pseudonymised data. This approach, though effort-intensive due to the need for pseudonymisation, preserves analytical insights while curtailing the legal exposure risks.

These delineated strategies offer actionable avenues for businesses to harmonize their data warehousing endeavors with GDPR mandates.

Triangulating these firsthand accounts with external sources brings added depth to this discourse. An article from Medium (2022) elucidates the cost implications of GDPR for data warehouse implementation. The article underscores the need for data warehouses to accommodate personalized data queries, which means enabling an individual's request to access all their saved data. Such functionalities require sophisticated metadata management and lineage capabilities, inevitably heightening the costs of implementation.

Regarding the impact on the NPV, the initial brunt of GDPR compliance is felt through escalated implementation costs, driven largely by the requisite training and sophisticated metadata management mechanisms. This emphasizes the monetary commitment organizations must shoulder to align with regulatory benchmarks.

The cost trajectory doesn't plateau post-implementation; organizations grapple with consistent overheads, particularly when they tread the path of pseudonymization. This approach, while safeguarding analytical acumen and legal standing, necessitates higher operational expenditures. The higher cost of GDPR ongoing training is also reflected in the 2022 financial report by the insurance firm of case 2.

However, GDPR compliance fortifies an organization's data defense, shrinking the likelihood of data breaches. This risk abatement can, from an NPV standpoint, lead to a reduction in the discount factor, enhancing the valuation of future cash flows.

5 Conclusions

In this section, the study is summarized, highlighting the key findings, limitations, and potential directions for future research.

5.1 Summary of the Study

The digital age has ushered in an era where data is at the forefront of business decision-making. As organizations grapple with the vast amounts of data generated daily, the need for efficient data organization and storage solutions becomes paramount. In the light of recent developments in AI this becomes even more important, since statistical models need to be trained on at least partially structured data. This study delves into the intricate world of data warehousing, exploring the strategic implications of its implementation and the factors that influence its success. At the heart of this investigation is the NPV of investment in data warehousing, a decisive measure that aids businesses in making informed choices to maximize returns on their investments.

Data warehousing is not merely a technological endeavor; it is a strategic commitment that can shape an organization's competitive landscape for years. The study examines the foundational perspectives of Inmon (2005) and Kimball et al. (2008), two pioneers in the field, whose approaches to building a data warehouse offer contrasting methodologies. While Inmon (2005)'s top-down approach emphasizes an enterprise-wide data warehouse, ensuring data consistency, Kimball et al. (2008)'s bottom-up approach starts with individual data marts, allowing for quicker implementation but potentially risking data inconsistencies. The choice between these methodologies is influenced by an organization's specific requirements, resources, and context.

Ghemawat (1991)'s insights into the nature of commitment, characterized by elements such as lock-in, lock-out, lags, and inertia, provide a comprehensive framework to understand the strategic implications of data warehousing decisions. These commitments, while offering potential competitive advantages, also come with the risk of reduced adaptability in a rapidly changing digital environment. Cassiman, Ricart, and Valentini (2022)'s research further nuances this understanding, emphasizing the balance between the allure of a competitive edge and the value of flexibility.

The study then transitions into a detailed exploration of the factors influencing data warehouse implementation success. Drawing from seminal works by researchers like Wixom and Watson (2001), Yeoh and Koronios (2010), and Mungree, Rudra, and Morien (2013), the study identified both established and emerging factors that play a pivotal role in the success of data warehousing projects. These factors range from organizational elements like vision, culture, and management support to technical aspects like source systems and development technology. The study also highlighted emerging factors, such as company size, extent of cloud service usage and outsourcing, the implementation approach, and GDPR awareness, which are becoming increasingly relevant in the modern business landscape and influence the success of the endeavour.

The methodology adopted for this research was rooted in a qualitative design, employing two case studies, questionnaires, and interviews to gather empirical data. This structured approach aimed to discern the intricate relationships between various factors and their impact on the NPV of data warehousing investments. Two contrasting case studies, representing both an SME and a global corporation, were analyzed to provide multifaceted insights into the real-world implications of data warehousing decisions.

The results of the study, derived from a comparative analysis of the case studies, insights from interviews, triangulation of hypotheses, and engagement with broader literature, offered a comprehensive understanding of the factors influencing the NPV of data warehousing investments. The findings emphasized the paramount importance of vision in guiding the goals and objectives of data warehousing projects. The state of source systems, the role of firm culture, and the availability of resources were also identified as significant determinants of success. Emerging factors, such as the influence of firm size and implementation approach on NPV, the implications of cloud usage and outsourcing, and the financial and strategic impact of GDPR compliance, were explored in depth, providing a contemporary perspective on the evolving landscape of data warehousing.

In essence, this study underscores the strategic significance of data warehousing in the modern business environment. As organizations navigate the challenges and opportunities of the digital age, the decision to invest in a data warehouse becomes a reflection of their broader strategic vision, risk appetite, and commitment to future growth. The insights derived from this research offer valuable guidance to businesses, helping them make informed decisions that align with their long-term objectives while retaining the flexibility to adapt to an ever-changing business landscape.

5.2 Key Findings and Contributions

A significant gap in the literature, as identified in this study, is the exploration of data warehouse implementation success factors from both a financial and strategic perspective. While previous research has extensively delved into the various determinants of data warehouse implementation success, the intricate interplay between these success factors, the NPV dimensions of the implementation project, and their strategic implications remains underexplored. This study bridges this gap by examining the interference between implementation success factors, NPV dimensions, and the broader strategic considerations associated with data warehousing decisions. By doing so, this research offers valuable insights for managers and decision-makers. It not only underscores the CSF that influence the benefits, costs, and risks associated with data warehouse implementation but also elucidates the strategic ramifications of such decisions.

First and foremost, organizations undertake the risk and costs of an investment in a data warehouse in order to gain a competitive edge. This competitive edge comes not only from the benefits of the data warehouse itself, but also from the lock-out effect that competitors experience that do not invest in such undertaking, as described by Ghemawat (1991). The lag in terms of time until such implementation is completed can limit the strategic options of companies, as experienced on the example of the company that wanted to make use of AI, but did not posses a sufficient data foundation to enable the training and use of AI models, like a data warehouse can provide it.

At the same time, this investment also comes with "lock-in" effects after Ghemawat (1991), making the implementing company reliant on expertise for the technology that was used, current GDPR regulation, as well as cloud and outsourcing providers. The challenge for managers is to balance the flexibility and the sustainability of this strategic commitment, as described by Cassiman, Ricart, and Valentini (2022). One suitable approach that was identified is gradually building Data Marts as suggested by Kimball et al. (2008). This approach is favorable especially for bigger organizations that need to include multiple departments in the data warehouse, as it keeps technological dependency, implementation costs and associated risks low. Furthermore, organizations can take advantage of low costs of proof of concepts using the cloud when thinking about implementing a data warehouse. This is especially advisable since literature and interviews indicate the crucial role of a vision associated with the implementation. The proof of concept can be tested on a specific use case that will, with low cost and risk associated, give an indication of the feasibility of the implementation. In this first glimpse at the data environment currently in place in the organization, underlying problems with source systems, another significant factor for the implementation success, can be identified early. Additionally, the company culture can be examined which can lower the risk of non-adoption and give an indication of the time lag due to inertia as described by Ghemawat (1991).

While Ramamurthy, Sen, and Sinha (2008)'s study highlights the role of firm size in data warehousing adoption, it primarily focused on the resource availability of larger firms. Since the emergence of cloud computing and outsourcing, this resource disadvantage does not keep SME from implementing data warehouses, as found through the conducted interviews and triangulation. Both participants indicated that the size of a company in terms of number of employees is not correlated with data warehouse adoption. Therefore, the size of a company does not create a per se barrier for implementation success, but has an impact on the NPV of such undertaking. A key finding is therefore that contrary to the findings of Ramamurthy, Sen, and Sinha (2008), resources are not on the forefront of implementation success, since organizations can make use of outsourcing and cloud computing, considerably lowering the up-front implementation costs for hardware, software, and the hiring of data warehousing professionals.

The aim of this study was to answer the research question *How does firm size influence the perceived NPV of Investments in data warehousing?*. Figure 5 already gave a good understanding of the interconnectedness of the implementation success factors, and especially the role of firm size in this regard. The discussion part of this study explained how these connections interrelate with the NPV. Based on these findings, a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis can be obtained, comparing SME with Corporations, as seen in table 2.

Figure 5 and table 2 can be seen as answers to the research question, relating the constituents of the NPV to the firm size and revealing differences between companies of different size.

	SME	Corporation
S	• Costs of implementation are more planable and lower in absolute terms • Cloud usage lowers entry barrier to data warehouse • Data warehouse implementation time horizon is generally shorter	• Resources are more abundant • Economies of scale: Internal experts are more cost effective if trained well
W	• Resources are less abundant, data warehouse implementation has bigger financial impact • Dependency on outsourced data warehouse experts since internal expertise is too costly to establish • No internal resources drives up ongoing costs for service provider in relative terms	• Clear Vision or Goal of data warehouse implementation decides over implementation success • Technological path dependency: Internal sources need to be constantly trained on new technological trends • GDPR guidelines tend to be more strict than for SME • Outsourcing is compared to usage of internal experts more expensive, therefore hindered access to most recent developments in technology • Corporations are more prone to risk and costs of data warehouse implementation
O	• Cloud usage creates lower implementation costs, more flexibility and scalability • Time to market is quicker with cloud usage • Proof of concept is quicker established with cloud	• Cloud usage creates lower implementation costs, more flexibility and scalability • Time to market is quicker with cloud usage • Proof of concept is quicker established with cloud • Possibility to create more granular insights • Enhanced ability to link data
T	• Increase of Cloud data warehouse Cost • Dependency on outsourced expertise creates risk of giving away internal expertise	• Increase of Cloud data warehouse Cost • New technological developments that internal experts are not familiar with • Strict GDPR requirements creating need for on-premise data warehouse implementation comes with cyber security risk • Over-usage of outsourcing creates dependency • Changing regulation: GDPR makes cloud less utilisable e.g. Banks moving back to on-premise data warehouse • Scalability: If corporation becomes multi-national, cloud usage is inevitable, since it is too difficult to maintain proximity with on-premise solutions • Data warehouse benefits might not scale with the costs of implementation

Table 2: SWOT Analysis of data warehouse implementation, SME vs. Corporations

Bigger companies exhibit in general larger amounts of data, also from different topic areas. This poses both opportunities and challenges, as described by Agrawal et al. (2012), since to harness benefits from the data, the datasets need to be interconnected and accessible. Therefore, larger organizations need to walk the extra mile in order to reap additional benefits associated with the higher costs and longer implementation horizons compared to SME. If this is done correctly, however, the intrinsic value of data can amplify exponentially. At the same time, bigger corporations are prone to higher risks, justifying a higher discount rate for future cashflows. As learned from the interviews, the size of a company doesn't scale the potential monetary losses on a linear basis, but can be exponential. Furthermore, with the growth in size of a company, depending on the gathered data, GDPR regulations can become more strict, and the risk of dependency on outsourcing rises. For these reasons, bigger corporations can not rely on cloud usage and outsourcing on such a profound basis as SME can.

GDPR awareness is found to be hindering organizations in implementing data warehouses, depending on the industry and handled data. However, if done right, the accurate handling of GDPR can lower the discount factor of an implementation project, as it lowers the risk of data breaches and the accompanied reputational loss for the company. Multiple approaches were introduced in handling GDPR compliance, focusing on the processability and associated costs.

5.3 Limitations and Future Research Directions

This study, while offering a comprehensive understanding of the strategic implications of data warehousing and its influence on the NPV using a qualitative case study approach, has certain limitations that provide avenues for future research.

1. **Sample Size:** The primary limitation was the number of participants. The specificity of the research field and the criteria for case selection meant that only a limited number of participants could be included. This inherently restricts the generalizability of the findings. Future studies could benefit from a broader sample, encompassing a wider range of organizations and industries. In addition, as noted by El-Adaileh and Foster (2019), developing countries should be considered for the sample.

2. **Quantitative Analysis:** The qualitative nature of this study, while providing in-depth insights, did not allow for rigorous quantitative analyses, such as regression. A more extensive dataset would enable researchers to employ regression analysis

to discern relationships between various factors influencing the NPV of data warehousing investments. This would provide a more robust statistical foundation to the findings and could help in identifying patterns or trends that might not be evident in a qualitative study.

3. **Real Option Approach:** The study did not delve into the real option approach for valuing the decision to implement a data warehouse. As suggested by Taudes (1998) and Benaroch and Kauffman (1999), the traditional NPV can be expanded to a strategic NPV that takes into account the value of options from active management. Future research could explore this approach in the context of data warehousing, providing a more nuanced understanding of the financial implications of such decisions.

4. **Time Constraints and Outsourcing:** Preliminary findings hint at a potential relationship between outsourcing and time constraints in data warehousing projects. Outsourced consultants might face tighter time constraints compared to in-house teams. However, due to the limited sample size, this study could not confirm this relationship. Future research could delve deeper into this aspect, exploring how the nature of the project team (in-house vs. outsourced) influences project timelines and outcomes.

5. **Firm Size and Implementation Time:** The relationship between firm size and the time required for data warehouse implementation remains an underexplored area. While this study touched upon it, a more focused investigation could provide valuable insights into how organizational size impacts project timelines and, consequently, the NPV of the investment.

In conclusion, while this study has shed light on several critical aspects of data warehousing from both strategic and financial perspectives, there remains a plethora of avenues for future research. As the digital landscape continues to evolve, understanding the intricacies of data warehousing decisions will become even more crucial for organizations aiming to harness the power of data effectively.

References

El-Adaileh, Nadeem Ali and Foster, Scott (2019). "Successful business intelligence implementation: a systematic literature review". In: *Journal of Work-Applied Management* 11.2, pp. 121–132.

Agrawal, Divyakant et al. (2012). "Challenges and Opportunities with Big Data. A community white paper developed by leading researchers across the United States". In: *Computing Research Association, Washington*.

Benaroch, Michel and Kauffman, Robert J (1999). "A case for using real options pricing analysis to evaluate information technology project investments". In: *Information systems research* 10.1.

Brealey, Richard A et al. (2006). *Corporate finance*. Vol. 8. McGraw-Hill/Irwin Boston et al.

Brynjolfsson, Erik, Hitt, Lorin M, and Kim, Heekyung Hellen (2011). "Strength in numbers: How does data-driven decisionmaking affect firm performance?" In: *Available at SSRN 1819486*.

Cassiman, Bruno, Ricart, Joan Enric, and Valentini, Giovanni (2022). "Commitment and Competitive Advantage in a Digital World". In: *Strategy Science* 7.2.

Eisenhardt, Kathleen M (1989). "Building theories from case study research". In: *Academy of management review* 14.4, pp. 532–550.

Fernandes, Sérgio and Bernardino, Jorge (2016). "Cloud Data Warehousing for SMEs." In: *ICSOFT-EA*, pp. 276–282.

Gagnon, G. (1999). "Data warehousing: An overview." In: *PC Magazine*, pp. 245–246.

Ghemawat, Pankaj (1991). *Commitment*. Simon and Schuster.

Glassdoor (2023). *Data Warehouse Architect Salaries*. URL: https://www.glassdoor.com/Salaries/data-warehouse-architect-salary-SRCH_KO0,24.htm (visited on 09/17/2023).

IBM (2023). *What is a data warehouse?* URL: https://www.ibm.com/topics/data-warehouse (visited on 08/15/2023).

Inmon, William H (2005). *Building the data warehouse*. John wiley & sons.

James Lupton, Forbes (2022). *Why So Many Data Warehouse Projects Fail*. URL: https://www.forbes.com/sites/forbestechcouncil/2022/06/07/why-so-many-data-warehouse-projects-fail/ (visited on 06/07/2022).

Kimball, Ralph et al. (2008). *The data warehouse lifecycle toolkit*. John Wiley & Sons.

Li, He, Yu, Lu, and He, Wu (2019). *The impact of GDPR on global technology development*.

Machado, Inês Araújo, Costa, Carlos, and Santos, Maribel Yasmina (2022). "Data mesh: concepts and principles of a paradigm shift in data architectures". In: *Procedia Computer Science* 196, pp. 263–271.

Medium (2022). *GDPR Compliance in Big Data Infrastructure*. URL: https://medium.com/the-observatory-by-sifflet/the-gdpr-or-the-global-data-protection-regulation-is-an-eu-legislation-that-went-into-effect-on-ec5981d77644 (visited on 08/01/2022).

Microsoft (2023). *Modern data warehouse for small and medium business*. URL: https://learn.microsoft.com/en-us/azure/architecture/example-scenario/data/small-medium-data-warehouse (visited on 08/20/2023).

Mitchell, John, Ker, Daniel, and Lesher, Molly (2021). "Measuring the economic value of data". In.

Mungree, Deepshika, Rudra, Amit, and Morien, Diane (2013). "A framework for understanding the critical success factors of enterprise business intelligence implementation". In.

OurWorldInData (2023). *Historical cost of computer memory and storage*. URL: https://ourworldindata.org/grapher/historical-cost-of-computer-memory-and-storage (visited on 03/01/2023).

Payton, Fay Cobb and Handfield, Robert (2003). "Data warehousing implementation and outsourcing challenges: An action research project with Solectron". In: *Communications of the Association for Information Systems* 12.1.

Preston, David and Brohman, Kathryn (2002). "Outsourcing opportunities for data warehousing business usage". In: *Logistics Information Management* 15.3.

Ramamurthy, K Ram, Sen, Arun, and Sinha, Atish P (2008). "An empirical investigation of the key determinants of data warehouse adoption". In: *Decision support systems* 44.4, pp. 817–841.

RetailBankerInternational (2023). *The key issues with data privacy and the cloud*. URL: https://www.retailbankerinternational.com/comment/sophie-stalla-bourdillon-immuta-expert-comment-data-privacy-and-the-cloud/ (visited on 01/27/2023).

Rikhardsson, Pall and Kræmmergaard, Pernille (2006). "Identifying the impacts of enterprise system implementation and use: Examples from Denmark". In: *International Journal of Accounting Information Systems* 7.1.

ScienceSoft (2023). *How to Build a Data Warehouse from Scratch*. URL: https://www.scnsoft.com/analytics/data-warehouse/building (visited on 09/15/2023).

Statista (2022). *Share of corporate data stored in the cloud in organizations worldwide from 2015 to 2022*. URL: https://www.statista.com/statistics/1062879/worldwide-cloud-storage-of-corporate-data/ (visited on 03/28/2022).

Taudes, Alfred (1998). "Software growth options". In: *Journal of Management Information Systems* 15.1.

TDWI (2017). *Survey Says: Data Science Investment Can Pay Off Big Time*. URL: https://tdwi.org/articles/2017/02/28/data-science-investment-can-pay-off-big.aspx (visited on 02/28/2017).

— (2020). *Study Finds Three Out of Four Executives Lack Confidence in Their Data's Quality*. URL: https://tdwi.org/articles/2020/01/30/diq-all-executives-lack-confidence-in-data-quality.aspx (visited on 01/30/2020).

TDWI (2023a). *How to Monetize Your Data Assets.* URL: https://tdwi.org/articles/2023/07/19/biz-all-how-to-monetize-your-data-assets.aspx (visited on 07/19/2023).

— (2023b). *TDWI Insight Accelerator — Saving Time and Costs: How Small and Midsize Businesses Can Maximize the Power of Data.* URL: https://tdwi.org/research/2023/07/biz-all-insight-accelerator-small-and-midsize-businesses-can-maximize-the-power-of-data.aspx (visited on 07/28/2023).

TechTarget (2023a). *Cloud costs continue to rise among IT commodities.* URL: https://www.techtarget.com/searchcio/news/366548312/Cloud-costs-continue-to-rise-among-IT-commodities. (visited on 08/14/2023).

— (2023b). *Server prices drop while cloud costs rise in PPI data.* URL: https://www.techtarget.com/searchcio/news/366542095/Server-prices-drop-while-cloud-costs-rise-in-PPI-data (visited on 06/15/2023).

TheCorporateFinanceInstitute (2018). *Definition of the Net Present Value (NPV).* URL: https://corporatefinanceinstitute.com/resources/valuation/net-present-value-npv/ (visited on 03/13/2023).

Upwork (2023). *Data Warehouse Architect Consultant Hourly Fees.* URL: https://www.upwork.com/ab/profiles/search/?q=data%20warehouse%20architect (visited on 09/17/2023).

USCloud (2023). *15% Price Increase for Microsoft Cloud Starting April 1, 2023.* URL: https://www.uscloud.com/blog/15-price-increase-for-microsoft-cloud-starting-april-1-2023/ (visited on 04/01/2023).

Watson, Hugh J and Haley, Barbara J (1998). "Managerial considerations". In: *Communications of the ACM* 41.9.

Westerman, P. (2001). *Data Warehousing: Using the Wal-Mart Model.* ITPro collection. Elsevier Science. ISBN: 9781558606845. URL: https://books.google.com/books?id=dXazzPSLC6UC.

Wixom, Barbara H and Watson, Hugh J (2001). "An empirical investigation of the factors affecting data warehousing success". In: *MIS quarterly.*

Xu, Hongjiang and Hwang, Mark I (2007). "The effect of implementation factors on data warehousing success: An exploratory study". In: *Journal of Information, Information Technology, and Organizations* 2, p. 1.

Yeoh, William and Koronios, Andy (2010). "Critical success factors for business intelligence systems". In: *Journal of computer information systems* 50.3.

Yessad, Lamia and Labiod, Aissa (2016). "Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault". In: *2016 International Conference on System Reliability and Science (ICSRS).* IEEE, pp. 95–99.

A Appendix: Answers to Interview Questions

View results

Respondent

1	Anonymous	94:08 Time to complete

1. What is your current professional title (e.g. Date Warehousing Specialist) / describe your DW expertise briefly

Lead Cloud Consultant (specialized on Data Engineering/ Warehousing)

2. Which Factors do you deem important for a data warehousing implementation? (open question first, then later particular ones)

Data Integration/ Staging (Rohdaten in einen Platz packen), Data Quality, Need for better data insights, Cloud: easy to scale with lower implementation cost, Modernization: Staying competitive

3. Do you consider Firm size an important factor on the success of a DW implementation?

Firm size not as relevant as capital, management must be behind the project, organization must be changed, gdpr is a prohibitor because users are afraid that they will be charged, afraid of going in the cloud because of knowledge-outflow; data governance: the bigger companies are the more they have to be able to explain where the data comes from and who can access them, banks need to show multiple reasons why they declined credit

4. How do you perceive the impact of the Firm Size on the NPV of the DW implementation?

for small companies the entry barrier is lower to build DW because of cloud. capital wise the decision is still a bigger financial commitment compared to a big corporation. big corporation has more data therefore insights are better. big corporations not only have higher costs but can have more granular insights. driver of value is benefit and amount of information (also from different areas), small companies more niche. e.g. youtube bought and value of data in a few weeks paid off

5. Do you consider Outsourcing an important factor on the success of a DW implementation?

depending on the size of the company. if bigger they have better specialists anyway. smaller companies need consultants

6. How do you perceive the impact of Outsourcing on the NPV of the DW implementation?

smaller comapnies need outsourcing because otherwise internal training and expertise would be too expensive, for bigger corporations less significant. data specialist can cost 6 figures

7. Do you consider Cloud Usage an important factor on the success of a DW implementation?

important factor also for big corporations, because managers have budgets to reach their goals. for certain investment they have xxx benefit. take 1000€ for a proof of concept. time to market much faster

8. How do you perceive the impact of Cloud Usage on the NPV of the DW implementation?

big impact on time horizon because proof of concept can be established much quicker

9. Do you consider GDPR an important factor on the success of a DW implementation?

often required by regulations (banks), all publicly traded companies must be auditable, also small companies can be sued by individuals

10. How do you perceive the impact of GDPR on the NPV of the DW implementation?

> key must be given to un-pseudonymize, costs are higher because of anonymization, real-time analyses can be more expensive because must be near, depends on what data is saved

11. Do you consider the implementation approach acc. to Inmon vs. Kimball an important factor on the success of a DW implementation?

> did not build an inmon for a long time, data mesh: each department is responsible for its own data,

12. How do you perceive the impact of the implementation approach on the NPV of the DW implementation?

> inmon for sure more expensive, but cheaper on the long run, kimball easier to maintain because less complex, inmon expense can grow exponentially depending on growth and granularity

13. When did you start implementing your (last) data warehouse

> 04/07/2023

14. When did you end with data warehouse implementation (or expected)

> 11/10/2023

15. How far along are you in the implementation of a data warehouse (7 = done)

1	2	3	4	**5**	6	7

16. Which industry does the company operate in (between Manufacturing, Healthcare, Retail/Wholesale, Telecommunications, Financial Services/Banking, Insurance, Government, Utilities, Education/Publishing, Petrochemical, Transportation, Market Research, Reseller, Travel, Defense, Distribution, Consumer Products, Other)

Healthcare

17. What size is your firm in terms of employees (according to ramamurthy2008empirical)

- ○ 1-400
- ○ 401-800
- ◉ 801-1,500
- ○ 1,501-3,000
- ○ 3,001-6,000
- ○ 6,001-15,000
- ○ More than 15,000

18. To what extend would you say your data warehouse is using cloud services (7 = most use of cloud services)

1	2	3	4	5	6	**7**

19. Comments on Cloud Usage

Had on-prem before now moved completely to cloud

20. To what extend would you say your data warehouse implementation used outsourcing services (7 = most use of outsourcing)

1	2	3	4	5	**6**	7

21. Comments on Outsourcing

biaas needs data warehouse to begin with -> bi solutions are then built by provider based on data that is already there. for smaller companies makes more sense to outsource instead of getting a specialist to built reports

22. Based on the previously introduced factors, what else was driving your decision on implementing the data warehouse in that way

scalability, flexibility, costs of running dw in cloud vs infrastructure cost. when building machine learning solutions it is always cheaper to go to the cloud, because a lot of GPU power is used at once. GDPR was not considered for implementation.

23. Management support - Overall management has encouraged the use of DW

1	2	3	4	5	6	**7**

24. Management support - User satisfaction has been a major concern of management

1	2	3	4	5	**6**	7

25. Champion - A high level champion for DW came from Information systems

1	2	3	4	5	6	**7**

26. A high level champion for DW came from a functional area

1	2	3	4	5	6	**7**

27. Resources - The DW project was adequately funded

1	2	3	4	5	6	**7**

28. Resources - The DW project had enough team members to get to work

1	2	3	4	**5**	6	7

29. Resources - The DW project was given enough time for completion

1	**2**	3	4	5	6	7

30. User Participation - Information System and users worked together as a team on the DW project

1	2	3	4	5	6	**7**

31. User Participation - Users were assigned full-time to parts of the DW project

1	2	3	4	5	6	7

32. User Participation - Users performed hands-on activities (e.g. data modeling) during the DW project

1	2	3	4	5	6	7

33. Team skills - Members of the DW team (including consultants) had the right technical skills for DW

1	2	3	4	5	6	**7**

34. Team skills - Members of the DW team had good interpersonal skills

1	2	3	4	5	**6**	7

35. Source systems - Common definitions for key data items were implemented across the source systems

1	2	3	4	5	6	**7**

36. Source systems - The data sources used for DW were diverse and disparate applications/systems

1	2	3	4	5	**6**	7

37. Source systems - a significant number of source systems had to be modified to provide data for DW

1	2	3	4	5	6	7

38. Development Technology - The DW technology that the project team used worked well with technology already in place in the organization

1	2	3	4	5	6	**7**

39. Development Technology - Appropriate technology was available to implement the DW

1	2	3	4	5	6	**7**

40. NPV - The DW NPV was above (1-3) or below (5-7) expectations

1	2	3	4	**5**	6	7

41. NPV - The DW implementation costs were above (1-3) or below (5-7) expectations

1	2	3	4	**5**	6	7

42. NPV - The DW ongoing costs were above (1-3) or below (5-7) expectations

1	**2**	3	4	5	6	7

43. NPV - The DW benefits were above (1-3) or below (5-7) expectations

1	2	3	**4**	5	6	7

44. NPV - the implementation time horizon was shorter (1-3) or longer (5-7) than expected

1	2	3	4	**5**	6	7

45. Was there an internal rate of return that the project was discounted with

internal rate of return high because of pharma industry therefore need to compare to other projects

View results

Respondent

2	Anonymous	91:53 Time to complete

1. What is your current professional title (e.g. Date Warehousing Specialist) / describe your DW expertise briefly

Data Warehouse Architekt (since more than 20 years)

2. Which Factors do you deem important for a data warehousing implementation? (open question first, then later particular ones)

Business Case must be there, e.g. Sales reporting, Overview on sales and revenue, problems: people that sit on their data and that dont want to make it public. Need to speak with departments to elaborate exactly what should be the goal of the DW. Focus on certain usecase. Business case, Technology must be decided on (e.g. customer says they dont want microsoft or no cloud).

3. Do you consider Firm size an important factor on the success of a DW implementation?

Big company: wants to implement dw but never did it before. Managing this huge project very difficult with inmon approach. For SME buy finished solutions. With only 10-20 people in the organization wont do data warehouse implementation oneself. Better use something pre-built.

4. How do you perceive the impact of the Firm Size on the NPV of the DW implementation?

Costs are more planable and lower when firm is smaller. Smaller time horizon, smaller focus, more planable, risks are also lower. For big company: if too many participants -> risk of failure is bigger the bigger the company. Firm size is like a leverage for costs and risks. The bigger the company the more sure it has to be what exactly they want to achieve with the implementation. Benefits reasonable? Not dependent on size. DW is built to have better more valid decisions. Very dependent on focus area.

5. Do you consider Outsourcing an important factor on the success of a DW implementation?

For smaller companies outsourcing is a very important factor since they do not have the expertise in-house. For bigger companies: if people dont have time to do DW, outsourcing is not a solutions because users are not behind the project and will be very dependent on the solution provider. Internal know-how to use the DW is not there. Risk is also giving data to externals.

6. How do you perceive the impact of Outsourcing on the NPV of the DW implementation?

Costs rise, so investment is bigger when we only use externals. Internal experts are cheaper. Company size lever is always there. The more money we work with, the higher is also the risk. e.g. VW implements DW and finds market gap -> revenue with new product.

7. Do you consider Cloud Usage an important factor on the success of a DW implementation?

Depends on business. If company is global player, then they need to go with the cloud. Because otherwise they would need to start multiple data warehouses on-premise in every country. So the bigger the company, the more important it is to move to the cloud. Scalability is big issue. If business case is very much growing. SME can use cloud because it does not have to worry about many aspects of DW technology. Small and cheap server for DW would alsowork.

8. How do you perceive the impact of Cloud Usage on the NPV of the DW implementation?

> Trend of Cloud Usage is that it will get more expensive. To begin with cloud is cheap but will become more expensive over time (energy crisis, workforce that maintains). For now cloud is cheap because not yet so many people are using it but once it is the main thing it will become more expensive because they need very high availabilities. e.g. Microsoft already increased prices for Azure. Expectation is that cloud computing costs will rise. For technology side: for now cheap because of cheap chinese production but will also increase in the future. For now cloud is cheaper because it is politically trying to get market share.

9. Do you consider GDPR an important factor on the success of a DW implementation?

> Indeed an important factor, maybe not priority 1. Maybe not for success but it can be a risk to implement data warehouse because employees have access to data and could potentially spread it. With a great DW the risk of spreading information unlawfully is higher.

10. How do you perceive the impact of GDPR on the NPV of the DW implementation?

> Costs rise for DW implementation. Users need to be trained, who may see which data, which data am I allowed to have. Benefits too abstract to estimate. DW can be seen as an enabler. Enables better decision making that is then used to create value.

11. Do you consider the implementation approach acc. to Inmon vs. Kimball an important factor on the success of a DW implementation?

> Technicians need to evaluate how to implement the data warehouse based on the goal of the business.

12. How do you perceive the impact of the implementation approach on the NPV of the DW implementation?

> Say a company decides for approach that company is not familiar with -> needs more training, other tools -> IT wisdom "keep it simple and stupid". Inmon approach was always bad from experience. Need to build modularly. Always used star/ snow flake architectures.

13. When did you start implementing your (last) data warehouse

02/01/2019

14. When did you end with data warehouse implementation (or expected)

01/07/2020

15. How far along are you in the implementation of a data warehouse (7 = done)

1	2	3	4	**5**	6	7

16. Which industry does the company operate in (between Manufacturing, Healthcare, Retail/Wholesale, Telecommunications, Financial Services/Banking, Insurance, Government, Utilities, Education/Publishing, Petrochemical, Transportation, Market Research, Reseller, Travel, Defense, Distribution, Consumer Products, Other)

Insurance

17. What size is your firm in terms of employees (according to ramamurthy2008empirical)

- ○ 1-400
- ○ 401-800
- ○ 801-1,500
- ○ 1,501-3,000
- ◉ 3,001-6,000
- ○ 6,001-15,000
- ○ More than 15,000

18. To what extend would you say your data warehouse is using cloud services (7 = most use of cloud services)

1	2	**3**	4	5	6	7

19. Comments on Cloud Usage

Proof of Concept with AWS cloud computing. One step after the other moving to cloud. 80% on premise, 20% on cloud (for Germany).

20. To what extend would you say your data warehouse implementation used outsourcing services (7 = most use of outsourcing)

1	2	**3**	4	5	6	7

21. Comments on Outsourcing

Many things done internally (80% self-done, 20% externals)

22. Based on the previously introduced factors, what else was driving your decision on implementing the data warehouse in that way

Cloud Hype comes from Management level. Maybe not needed. GDPR no issue.

23. Management support - Overall management has encouraged the use of DW

1	2	3	**4**	5	6	7

24. Management support - User satisfaction has been a major concern of management

1	**2**	3	4	5	6	7

25. Champion - A high level champion for DW came from Information systems

1	2	3	4	**5**	6	7

26. A high level champion for DW came from a functional area

1	2	3	4	5	6	**7**

27. Resources - The DW project was adequately funded

1	2	3	4	5	6	**7**

28. Resources - The DW project had enough team members to get to work

1	2	3	4	5	6	**7**

29. Resources - The DW project was given enough time for completion

1	2	3	4	5	6	**7**

30. User Participation - Information System and users worked together as a team on the DW project

1	2	3	4	5	**6**	7

31. User Participation - Users were assigned full-time to parts of the DW project

1	2	3	4	5	6	7

32. User Participation - Users performed hands-on activities (e.g. data modeling) during the DW project

1	2	3	4	5	6	7

33. Team skills - Members of the DW team (including consultants) had the right technical skills for DW

1	2	3	4	5	**6**	7

34. Team skills - Members of the DW team had good interpersonal skills

1	2	3	4	**5**	6	7

35. Source systems - Common definitions for key data items were implemented across the source systems

1	2	3	4	5	6	7

36. Source systems - The data sources used for DW were diverse and disparate applications/systems

1	2	3	4	5	6	**7**

37. Source systems - a significant number of source systems had to be modified to provide data for DW

1	2	3	4	5	6	7

38. Development Technology - The DW technology that the project team used worked well with technology already in place in the organization

1	2	3	4	5	6	**7**

39. Development Technology - Appropriate technology was available to implement the DW

1	2	3	4	5	**6**	7

40. NPV - The DW NPV was above (1-3) or below (5-7) expectations

1	2	3	**4**	5	6	7

41. NPV - The DW implementation costs were above (1-3) or below (5-7) expectations

1	2	3	4	**5**	6	7

42. NPV - The DW ongoing costs were above (1-3) or below (5-7) expectations

1	2	**3**	4	5	6	7

43. NPV - The DW benefits were above (1-3) or below (5-7) expectations

1	**2**	3	4	5	6	7

44. NPV - the implementation time horizon was shorter (1-3) or longer (5-7) than expected

1	2	3	4	**5**	6	7

45. Was there an internal rate of return that the project was discounted with

> Dividends on Stock is planned forward. This is discount rate, so this project must also be used for this. Was not done for this project, as long as users were happy all good.

End of Master Thesis I

Master Thesis II

Case Study: Evaluating the Net Present Value of a Data Warehouse at Transalb, a Small German Trucking Company

Ferdinand Netsch
Student Number: 42195
Master's Thesis (M.Sc. Finance)

Abstract

The continuous advent of data management technologies has enabled small and medium-sized enterprises (SME)s like Transalb Kühl- und Express-GmbH, a small German trucking company, to explore the strategic and financial benefits of data warehousing. This case study evaluates the data warehouse investment at Transalb by employing a Net Present Value (NPV) analysis to determine its financial viability. By collecting qualitative and quantitative data, engaging extensively with a data warehousing service provider, and incorporating management consultations, the study presents an unambiguous financial forecast for the project. With a focus on cloud-based solutions and leveraging outsourcing, the NPV calculation balances tangible benefits against implementation and ongoing costs. Despite a cautious risk adjustment to benefits, the investment emerges as profitable. Improvements encompass better pricing accuracy, time savings, and fuel efficiencies, among others, that contribute to the firm's competitive edge and growth potential. Through this empirical examination, the study underscores the accessibility of data warehousing for SMEs and potentially significant returns, emphasizing the strategic importance of data-driven decision-making.

Stockholm School of Economics
Department of Finance

November 2023

Contents

Abbreviations

MRP Market Risk Premium. 99, 100

NPV Net Present Value. 92, 94, 96, 101

SME Small and medium-sized enterprises. 94, 95, 98, 101

TV Terminal Value. 100

WACC Weighted Cost of Capital. 99, 100

1 Introduction

In today's data-driven business environment, firms recognize the strategic importance of effectively organizing and analyzing their data assets. A commonly adopted approach is to implement a data warehouse, a centralized repository that consolidates data from diverse sources for analysis and decision-making purposes.

The thesis by Ferdinand Netsch (2023) provides a comprehensive theoretical framework for understanding the implications of investing in data warehousing from both a strategic and financial perspective. Drawing from seminal literature and two in-depth case studies, Netsch's research delineates key factors influencing the success and NPV of data warehousing projects. Specifically, the study explores how emergent trends such as firm size, cloud usage, outsourcing, and data privacy regulations can impact implementation outcomes.

Building on Netsch's foundational work, this empirical case study aims to apply the developed theoretical framework to understand real-world data warehousing decisions, particularly for Small and medium-sized enterprises (SME). The purpose is to gain further insights into how the various conceptual determinants of success and value manifest in practice. Drawing from Transalb Kühl- und Express-GmbH, a German SME firm in the logistics sector planning to integrate a data warehouse, the study examines how a calculation of the feasibility of such investment is determined by applying the NPV as decisive measure.

By interacting with and gathering information from experts for data warehousing implementation and the staff of Transalb, the case could be analyzed comprehensively. The findings offer pragmatic implications for managers navigating similar transformations, highlighting trade-offs between benefits, costs, and risks of such undertaking. By grounding academic theories in lived experiences, the case provides a textured account of contemporary investments in data warehousing.

2 Methodology

The study employed a qualitative case study approach, involving multiple steps to gather data, conduct analyses, and derive insights pertinent to the research question. The methodological pathway as follows:

2.1 Company Selection and Engagement

The initial step was to search for and identify companies that would be willing to participate in the case study. Three companies were approached, and Transalb agreed to participate. This involvement offered a unique opportunity to observe the process of a data warehouse integration within an SME operating in the logistics sector.

Transalb is a trucking company specializing in both heavy transportation across Europe and express deliveries for various businesses such as garages, discounters, and flower shops. The company's office staff included the senior leader (the original founder), a junior leader handling the technological aspects, one staff member responsible for truck disposition, two accountants, and one personnel expenses accountant. Additionally, there are approximately 45 drivers for trucks and delivery vehicles.

2.2 Data Warehouse Solution Provider Inquiry

Collaborating with Transalb, six data warehousing solution providers were contacted to request proposals. Of these, two providers deemed Transalb too small for their services, one offered a proposal that was too expensive (€100,000) without adequate context or requirements understanding, and another provider insisted on a full-day workshop costing €3,000 for requirements elicitation, which was not pursued due to the cost.

2.3 Proposal Evaluation and Vendor Interaction

The remaining two providers issued questionnaires and engaged with the junior leader through virtual meetings. Their responsiveness and proactive engagement were documented. The answers to the questionnaires provided essential information for proposal assessment and can be found in Appendix F. Both companies then submitted proposals based on the interaction with the technical expert from Transalb and the questionnaire. The proposal from CINTELLIC GmbH was deemed superior due to its thoroughness, detail, and cost-effectiveness, making it the chosen provider. CINTELLIC GmbH consented to partake in the case study, allowing for the inclusion of their proposal in this document. It can be found in Appendix G.

2.4 NPV Calculation

Further discussions and refinements concerning the costs and planned benefits of the data warehouse implementation led to the development of an NPV calculation. The NPV model serves as a financial evaluation tool to forecast the potential monetary benefits of the investment against its costs over time. The data collected for this case study were primarily qualitative, derived from discussions with the data warehousing service provider CINTELLIC and internal conversations with Transalb staff, complemented by quantitative data from discussions with the Transalb management team and internal reports. The NPV calculation is provided in Appendix A and will now be evaluated in detail in the Discussion part of this study.

3 Discussion of Empirical Results

Since the feasibility of the investment is evaluated based on its NPV, we will have a look at the constituents based on the spreadsheet calculation provided in Appendix A. This spreadsheet is showing the benefits and costs of the data warehousing implementation and the weeks at which they are expected to occur. The project is divided into four phases, which can be examined in detail in the proposal by CINTELLIC in Appendix G. Phase 0 is concerned with the project setup and includes a workshop including the management team of Transalb, a Management Consultant and a Data Warehousing Architecture Consultant from CINTELLIC. In Phase 1a, the concept is created and documented to align all systems with a plan that is then implemented in phase 2. Phase 3 is the phase in which the data warehouse is already fully operational and used.

3.1 Monetary Benefits

Starting from the top of the spreadsheet, we can see weekly benefits of €420 starting at phase 3 that come from new products and demand insights. At Transalb, proposals to customers currently include positions for "getting to" and "leaving from" the pick-up point and destination of a truck load. Based on insights from the data warehouse, these values can be adjusted more accurately. E.g. it could be argued that a delivery to far-north Germany is less attractive, because it is more difficult to find a connection load. So instead of quoting the usual average 150km "leaving from", one could refer to the insights from the data warehouse, showing that on average, a connection load takes generally 250km from this location. With this more accurate insight, the pricing - which is done on a manual basis, since information in this market is not considered efficient -

can be adjusted. The €420 are based on the assumption that 10km per day per truck can be charged in addition, with the general price of €2.10 per kilometer as it is already done today. All assumptions can be found in Appendix B.

Secondly, on two positions, cost can be saved. Starting in phase 3, €18 per week are the estimated savings on research time if it is estimated that one colleague spends two hours per week on research and 40% of this time can be saved due to better availability of information. Since this company does not outsource any research activity, this position is comparably small, but could be considerable in other cases. The manual reporting labor has a higher impact with €288 weekly savings. It is assumed that 20% of the working time can be automated. The personnel creating the reports could then be staffed on other opportunities or the working contract could be adjusted reflecting the reduced hours.

Another benefit of the implementation of the data warehouse would be savings on petrol. Currently, truck drivers will find the nearest petrol station available when trucking long distances. An early warning system derived from the insights of the usual range of a truck compared with daily kilometers could give insights in which area the truck would be when a refueling is needed. Then, cheaper petrol stations in the area could be chosen, with an assumed saving per liter of one Euro cent. With an average daily consumption of 504 liters of diesel per day, this amounts to €5 per week.

Furthermore, the route planning and insights on time needed as well as kilometers driven to complete shipments will give better insights into which routes are more optimal. It is assumed that this way, 10km can be saved on every route. With 3 trips per week and truck, this amounts to a weekly saving of €55.

Lastly, a one time benefit of €2,000 is assumed in the beginning of phase 3 for the time saved that it would take a digital due diligence consultant to evaluate the systems in place and possible issues when merging the company into another. The documentation done by the data warehousing consultants would cover this topic thoroughly. Since this benefit is only realized in case it comes to a sale of the company, it was added to the non-tangible benefit section and can be toggled off in the spreadsheet.

3.2 Risk adjustment of the benefits

As we learned in the thesis by Netsch (2023), data warehousing implementation projects are prone to to fail, especially due to cultural reasons or non-alignment with overall goals of the implementing company. To account for these risks, the cash-flows of the calculated benefits were each reduced by 50% in each period. Research in this field showed that failure rates range from 41% to 90%. Since the studies with failure rates above 50% were looking at perceived benefits and not the failure of the overall project, a 50% adjustment of the cash-flows seems reasonable.

3.3 Implementation Costs

Due to the fact that the data warehousing implementation at transalb was calculated on a 100% cloud-basis, there is no hardware to be invested. Because the spreadsheet should be useable as a template for other data warehouse implementations, those costs were also included in the calculation but values were left out. The same is true for software components, which are fully paid for on an ongoing, monthly basis.

The main cost in the implementation phase, and overall, is therefore the personnel. As can be seen in the proposal by CINTELLIC in Appendix G, all sub phases of the implementation phase include a project leader as well as a tech consultant. The former comes in mainly in the concept phase since more interaction with the customer is needed during this time. The latter then starts with the implementation from the technical side of things, mainly contributing towards the project in the later phases of the overall implementation. The project leader is paid a daily rate of €1,400 and the tech consultant €1,200. Including charges for food and drink that are amounting to 28€ per day per person according to German law, this amounts to on average €8,500 per week until phase 3. In total, €67,840 are spent on the implementation personnel over 8 weeks until the data warehouse is completed.

It is assumed that €3,000 should be spent on training Transalb staff every 3 years, which completes the expenditure for implementation. As the current plan only covers 11 weeks, the amount is divided into weekly installments, similar to an amortization of the final cost.

3.4 Ongoing Costs

Regarding the cloud environment, it was assumed that the setup is based entirely on Microsoft Azure. This way, the cloud cost could be estimated using the Azure Price Calculator by Microsoft (2023). All inputs, that were closely aligned with the technical consultant, can be found in Appendix C. For database system, storage accounts, and analysis tools (Azure Data Factory), weekly costs of €34, €2, and €32 are assumed, respectively. This gives us a total weekly cost of €68 per week. Here we can see which great opportunity the cloud offers to SMEs, since no up-front hardware cost is required and the monthly cost is still kept reasonably low. Even the cost increase by cloud providers and the cost of expanding the data warehouse was taken into account, however, their impacts are insignificant and therefore won't be discussed in detail here.

To keep the setup secure and compliant, cost of €1,500 and €300 were assumed to occur every two years for cyber-security and GDPR audits, respectively. This accounts for €17 weekly cost in total.

Starting from the second week of phase two, two Microsoft PowerBI licences are maintained. With these licences, the senior and junior manager can both access a dashboard that gives them insights into the data from the data warehouse. The weekly cost amounts to €8.

3.5 Discount Rates and Terminal Value

The inputs for the calculation of the Weighted Cost of Capital (WACC) can be found in Appendix B. The formula for the applied weekly WACC is as follows:

$$WACC_{\text{weekly}} = (1 + WACC_{\text{annual}})^{\frac{1}{52}} - 1 \tag{1}$$

where the general WACC formula for annual rates is:

$$WACC_{\text{annual}} = \frac{E}{V} \times ke + \frac{D}{V} \times kd \times (1 - T) \tag{2}$$

The market value of equity E, market value of long-term debt D, and enterprise value V can be found in the 2022 financial report of Transalb in Appendix E. By information from the management, the annual cost of debt kd for Transalb is 4.5% per annum and the corporate tax rate T is 35%.

The cost of equity ke was calculated using the yield of 10-year German government bonds as risk-free rate r_f, that is as of the beginning of November 2023 2.27% per annum according to Deutsche Bundesbank (2023). Added to that is the product of the Market Risk Premium (MRP) and a blended, levered beta. The MRP for Germany is according to Statista (2023) 5.70% in September 2023.

$$ke = r_f + \beta_{\text{levered}} \times MRP \tag{3}$$

The blended, levered beta β_{levered} is the levered beta applicable for a software project in a trucking company, levering the blended beta β_{blended}:

$$\beta_{\text{levered}} = \beta_{\text{blended}} \times \left(1 + (1 - T) \times \frac{D}{E}\right) \tag{4}$$

where the blended beta β_{blended} is calculated as weighted-average of β_{truck} (unlevered beta for trucking companies) and β_{software} (unlevered beta for EU software) using the weights w_{truck} and w_{software}:

$$\beta_{\text{blended}} = \beta_{\text{truck}} \times w_{\text{truck}} + \beta_{\text{software}} \times w_{\text{software}} \tag{5}$$

Due to the heavier weight of the project towards the software sector, w_{software} was assumed with 60% and w_{truck} with 40%. The unlevered betas β_{truck} (1.83) and β_{software} (5.13) were found on the website of Damodaran (2023), a table showing the betas in an overview can be found in Appendix D. These betas are total betas, meaning that they expand the unlevered beta risk measure by the firm-specific risk, not just the market risk. Damodaran defines total beta as the unlevered beta divided by the correlation with the market. It makes sense to use total unlevered betas in this case, since we are looking at the investment from the perspective of the firm's management, which is not diversified across multiple companies.

From these assumptions, the blended beta β_{blended} was calculated to be 3.81 and the levered, blended beta to be 4.30. Combined with the MRP and risk-free rate r_f this resulted in a cost of equity ke of 27.23%. Given the capital structure of Transalb, the resulting WACC is 23.21% annually and therefore 0.40% weekly. A continuous growth rate g of 2.00% was assumed, according to the long-term inflation goal of the European Central Bank (2023).

With these values, the Terminal Value (TV) of each position can be calculated as:

$$TV = \frac{CF_{11} \times (1 + g)}{WACC - g} \tag{6}$$

where CF_{11} is in this case the cash-flow in the last planned week 11.

4 Conclusion

This case study illuminates the viability of data warehousing investments for SMEs through the lens of a German trucking company, Transalb Kühl- und Express-GmbH. By dissecting the project's financials via a comprehensive NPV calculation, it is evident that Transalb stands to gain significantly from implementing a data warehouse tailored to its operational needs. The calculated overall NPV is a robust €7,800, which contrasts the ongoing costs of €25,354 and the execution costs of €71,866 against the tangible benefits totaling €210,040. Even after applying a prudent 50% risk adjustment to the benefits, acknowledging the intrinsic risks of data warehousing projects, the project remains profitable.

The benefits of the data warehouse include enhanced accuracy in demand insights and pricing, savings on research time, reductions in manual reporting, optimized operational logistics such as route planning, and possible fuel cost efficiencies. These improvements, though some appear minor in isolation, cumulatively embody substantial value over time.

The key to realizing the data warehouse for SMEs is cloud-based solutions and outsourcing, which eliminate the need for large upfront investments in hardware and in-house expertise, opening up value creation and scaling opportunities that were previously reserved for larger companies. Beyond the apparent financial benefits, the project also offers strategic value by positioning the company more competitively and preparing it for scalable growth. Moreover, the data warehouse could enhance Transalb's attractiveness as a potential acquisition target, a non-tangible benefit that, though not factored into the core NPV, heralds potential future value.

References

Bank, European Central (2023). *Monetary policy.* URL: `https://www.ecb.europa.eu/mopo/html/index.en.html` (visited on 11/17/2023).

Bundesbank, Deutsche (2023). *Daily yields of current Federal securities.* URL: `https://www.bundesbank.de/en/statistics/money-and-capital-markets/interest-rates-and-yields/daily-yields-of-current-federal-securities-772220` (visited on 11/03/2023).

Damodaran, Aswath (2023). *Total Beta (beta for completely undiversified investor): Europe.* URL: `https://pages.stern.nyu.edu/~adamodar/pc/datasets/totalbetaEurope.xls` (visited on 01/05/2023).

Microsoft (2023). *Azure Pricing Calculator*. URL: https://azure.microsoft.com/pricing/calculator/ (visited on 11/16/2023).

Netsch, Ferdinand (2023). "The Shift in Data Organization: Uncovering Influences on the NPV of Investment in Data Warehousing and Differences across Firm Sizes". In: *Master's Thesis in Finance*.

Statista (2023). *Average market risk premium in Germany (MRP) from 2011 to 2023*. URL: https://www.statista.com/statistics/664825 (visited on 09/04/2023).

A Appendix: NPV Calculation for Data Warehouse Implementation

Data Warehouse Implementation Business Plan

NTB 1	Phase	
	Calendar Week	**NPV**
Benefits:	New Products Introduced (Due to Demand Insights)	111.205 €
	Cross-Selling	-
	Up-Selling	-
	Quicker Time to Market	-
Increased Revenue		**111.205 €**
	Increased Reponse Rate to Mailings	-
	Reduced Customer Churn	-
	Decreased Material Cost	-
	Data Democratization	-
	Reduced Research Time	4.707 €
	Improved Inventory Management	-
	Reduction in Cost of Information for Report Creation	76.377 €
	Reduced Customer Acquisition Costs	-
Cost Saving		**81.084 €**
	Fuel Cost Saving	1.334 €
	Optimized Routing	14.488 €
Firm-Specific Benefits		**15.823 €**
	Culture of innovation with data-driven insights	-
	Enhanced decision-making through insights	-
	Reduction in data, analysis and decision latency	-
	Enhanced System Quality	-
	Ability to compete on a global market	-
	Reduced Dependency on Suppliers	-
	Reduced Dependency on Customers	-
	Reduction in Challenges of Data Sharing (M&A)	1.929 €
Non-Tangible Benefits (NTB)		**1.929 €**
Risk-adjustment for probability of failure of Benefits (prob. of 50%)		**-105.020 €**
Benefits		105.020 €

		0	1a	1a	1a	2	2
NPV		**1**	**2**	**3**	**4**	**5**	**6**
	111.205 €	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	111.205 €	**-**	**-**	**-**	**-**	**-**	**-**
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	4.707 €	-	-	-	-	-	-
	-	-	-	-	-	-	-
	76.377 €	-	-	-	-	-	-
	-	-	-	-	-	-	-
	81.084 €	**-**	**-**	**-**	**-**	**-**	**-**
	1.334 €	-	-	-	-	-	-
	14.488 €	-	-	-	-	-	-
	15.823 €	**-**	**-**	**-**	**-**	**-**	**-**
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
	1.929 €	-	-	-	-	-	-
	1.929 €	**-**	**-**	**-**	**-**	**-**	**-**
	-105.020 €	**-**	**-**	**-**	**-**	**-**	**-**
	105.020 €	-	-	-	-	-	-

	2	2	3	3	3	3
NPV	**7**	**8**	**9**	**10**	**11**	**TV**
111.205 €	-	-	420 €	420 €	420 €	115.421 €
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
111.205 €	**-**	**-**	**420 €**	**420 €**	**420 €**	**115.421 €**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
4.707 €	-	-	18 €	18 €	18 €	4.886 €
-	-	-	-	-	-	-
76.377 €	-	-	288 €	288 €	288 €	79.272 €
-	-	-	-	-	-	-
81.084 €	**-**	**-**	**306 €**	**306 €**	**306 €**	**84.158 €**
1.334 €	-	-	5 €	5 €	5 €	1.385 €
14.488 €	-	-	55 €	55 €	55 €	15.038 €
15.823 €	**-**	**-**	**60 €**	**60 €**	**60 €**	**16.423 €**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
1.929 €	-	-	2.000 €	-	-	-
1.929 €	**-**	**-**	**2.000 €**	**-**	**-**	**-**
-105.020 €	**-**	**-**	**-1.393 €**	**-393 €**	**-393 €**	**-108.001 €**
105.020 €	-	-	1.393 €	393 €	393 €	108.001 €

Data Warehouse Implementation Business Plan

NTB 1	Phase	
	Calendar Week	**NPV**
Costs:	Servers	-
	Storage Devices	-
	Networking Equipment	-
	Backup Systems	-
	Redundant Hardware (Power Supply, Cooling, Networking)	-
	Hardware Cost	**-**
	Data Integration Tools	-
	ETL Processes	-
	Software	**-**
	Project Leader	-22.061 €
	Database Analyst	-
	GUI Programmer	-
	Mainframe/Host programmer	-42.362 €
	Charges for food and drink	-2.200 €
	Personell Cost (External)	**-66.623 €**
	Education and Training (includes ongoing Costs)	**-5.243 €**
Implementation Costs		**-71.866 €**
	Azure SQL Database	-9.141 €
	Storage Accounts	-607 €
	Azure Data Factory	-8.859 €
	Cloud Cost	**-18.607 €**
	Cloud Cost Increase	**-17 €**
	Loading Additional Data	-11 €
	New Release Upgrades	-
	Expanded User Populations	-
	Scalability costs	**-11 €**
	Electricity	-
	Cooling	-
	Space for Housing	-
	Utility Cost	**-**
	Disaster Recovery Systems (already included in cloud cost)	**-**
	Backup Systems (already included in cloud cost)	**-**
	Security	-
	Helpdesk Service	-
	Maintanance	-
	Staff	**-**
	Software Licence Renewals	-
	Periodic Security Audits	-3.932 €
	Periodic GDPR Audits	-786 €
	Updates and Audits	**-4.719 €**
	Hardware Depreciation	**-**
	Software Licences (2x PowerBI)	**-2.000 €**
Ongoing Costs		**-25.354 €**
Total		**7.800 €**

	0	1a	1a	1a	2	2
NPV	**1**	**2**	**3**	**4**	**5**	**6**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-22.061 €	-1.400 €	-7.000 €	-3.500 €	-3.500 €	-1.750 €	-1.750 €
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-42.362 €	-1.200 €	-6.000 €	-6.000 €	-6.000 €	-6.000 €	-6.000 €
-2.200 €	-280 €	-280 €	-280 €	-280 €	-280 €	-280 €
-66.623 €	**-2.880 €**	**-13.280 €**	**-9.780 €**	**-9.780 €**	**-8.030 €**	**-8.030 €**
-5.243 €	**-19 €**	**-19 €**	**-19 €**	**-19 €**	**-19 €**	**-19 €**
-71.866 €	**-2.899 €**	**-13.299 €**	**-9.799 €**	**-9.799 €**	**-8.049 €**	**-8.049 €**
-9.141 €	-34 €	-34 €	-34 €	-34 €	-34 €	-34 €
-607 €	-2 €	-2 €	-2 €	-2 €	-2 €	-2 €
-8.859 €	-32 €	-32 €	-32 €	-32 €	-32 €	-32 €
-18.607 €	**-68 €**	**-68 €**	**-68 €**	**-68 €**	**-68 €**	**-68 €**
-17 €	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-0 €**
-11 €	-0 €	-0 €	-0 €	-0 €	-0 €	-0 €
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-11 €	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-0 €**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-	**-**	**-**	**-**	**-**	**-**	**-**
-	**-**	**-**	**-**	**-**	**-**	**-**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-	-	-	-	-	-	-
-3.932 €	-14 €	-14 €	-14 €	-14 €	-14 €	-14 €
-786 €	-3 €	-3 €	-3 €	-3 €	-3 €	-3 €
-4.719 €	**-17 €**	**-17 €**	**-17 €**	**-17 €**	**-17 €**	**-17 €**
-	**-**	**-**	**-**	**-**	**-**	**-**
-2.000 €	**-**	**-**	**-**	**-**	**-**	**-**
-25.354 €	**-86 €**	**-86 €**	**-86 €**	**-86 €**	**-86 €**	**-86 €**
7.800 €	**-2.985 €**	**-13.385 €**	**-9.885 €**	**-9.885 €**	**-8.135 €**	**-8.135 €**

	2	2	3	3	3	3
NPV	7	8	9	10	11	TV
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-22.061 €	-1.750 €	-1.750 €	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-42.362 €	-6.000 €	-6.000 €	-	-	-	-
-2.200 €	-280 €	-280 €	-	-	-	-
-66.623 €	**-8.030 €**	**-8.030 €**	**-**	**-**	**-**	**-**
-5.243 €	**-19 €**	**-19 €**	**-19 €**	**-19 €**	**-19 €**	**-5.285 €**
-71.866 €	**-8.049 €**	**-8.049 €**	**-19 €**	**-19 €**	**-19 €**	**-5.285 €**
-9.141 €	-34 €	-34 €	-34 €	-34 €	-34 €	-9.214 €
-607 €	-2 €	-2 €	-2 €	-2 €	-2 €	-612 €
-8.859 €	-32 €	-32 €	-32 €	-32 €	-32 €	-8.930 €
-18.607 €	**-68 €**	**-68 €**	**-68 €**	**-68 €**	**-68 €**	**-18.756 €**
-17 €	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-0 €**	-18 €
-11 €	-0 €	-0 €	-0 €	-0 €	-0 €	-11 €
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-11 €	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-0 €**	**-11 €**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-	**-**	**-**	**-**	**-**	**-**	**-**
-	**-**	**-**	**-**	**-**	**-**	**-**
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	**-**	**-**	**-**	**-**	**-**	**-**
-	-	-	-	-	-	-
-3.932 €	-14 €	-14 €	-14 €	-14 €	-14 €	-3.964 €
-786 €	-3 €	-3 €	-3 €	-3 €	-3 €	-793 €
-4.719 €	**-17 €**	**-17 €**	**-17 €**	**-17 €**	**-17 €**	**-4.756 €**
-	**-**	**-**	**-**	**-**	**-**	**-**
-2.000 €	**-8 €**	**-8 €**	**-8 €**	**-8 €**	**-8 €**	**-2.061 €**
-25.354 €	**-93 €**	**-93 €**	**-93 €**	**-93 €**	**-93 €**	**-25.602 €**
7.800 €	**-8.142 €**	**-8.142 €**	**1.281 €**	**281 €**	**281 €**	**77.114 €**

B Appendix: Assumptions for NPV Calculation

Assumptions

Type	1	Unit
Number of weeks per year	52	weeks
Inflation (yearly)	3,00%	
Inflation (weekly)	0,06%	
GDPR Audit cost	300,00 €	
Audits every # years	2	# years
Security Audit cost	1.500,00 €	
DW and BI Training cost	3.000,00 €	
DW and BI Training every # years	3	# years
Cloud Cost Increase (yearly)	5,00%	
Cloud Cost Increase (weekly)	0,09%	

Benefits Estimates

Type	Value	Unit
<u>Manual Labor Saving:</u>		
Weekly saving on manual labor salary	288,46 €	
Long-term reduction in manual labor FTE	20%	
Manual Labor Annual Salary	75.000,00 €	
<u>Addional Revenue due to refined offering</u>		
Weekly Fuel saving	420,00 €	
Price per km in offering	2,10 €	€/km
Additional pricing per day	10	km/day/truck
Operating days of heavy trucks per week	5,0	days/week
Number of heavy trucks	4	trucks
<u>Fuel Saving:</u>		
Weekly Fuel saving	5,04 €	
Average saving per liter on fuel	0,01 €	
Average weekly fuel consumption (big trucks)	504	liters
Hours driven per day	7	hours/day
Average truck speed	60	km/h
Diesel usage per km of heavy truck	0,30	liters/km/truck

Optimized Routing		
Weekly fuel saving by route optimizing	54,72 €	
Saved km per route	10	km
Trips per week per truck	3	trips/week/truck
Average Diesel price per liter	1,52 €	€/liter
Reduced Research Time		
In money terms	17,78 €	
Average time spent on research each week	2,00	hours
% of research time reduction	40%	
Researcher yearly salary	40.000,00 €	
Researcher weekly salary	888,89 €	
Researcher contract hours weekly	40,00	hours
Researcher working weeks per year	45,00	weeks/year
Reduction in Data Sharing Challenges (M&A)		
In money terms	2.000,00 €	
Average digital due diligence consultant hourly rate	200,00 €	
Hours needed to create connectivity study	10,00	hours
Risk-adjustment for failure of DW Project		
Risk adjustment by probability of failure	50,00%	Prob. of failure

WACC Calculation

Type	Value	Unit
Cost of Debt (annual)	4,50%	
Continuous growth rate (yearly)	2,00%	
Continuous growth rate (weekly)	0,04%	
Market Value of Equity (2022)	886.308,90 €	
Market Value of long-term Debt (2022)	176.036,53 €	
Market Value of enterprise (2022)	1.062.345,43 €	
Cost of Equity (annual)	27,23%	
WACC (annual)	23,21%	
WACC (weekly)	0,40%	

CAPM Calculation for Return on Equity

Type	Value	Unit
Unlevered Beta for Trucking Companies	1,83	beta
Unlevered Beta for EU Software (System and Application) Project	5,13	beta
% Trucking of Project	40,00%	
% Software (System and Application) of Project	60,00%	
Check weights sum to 100%	TRUE	
Blended Unlevered Beta for Software Project in Trucking Company	3,81	beta
Corporate Tax Rate Transalb GmbH	35,00%	
Blended Levered Beta for Software Project in Trucking Company	4,30	beta
Risk-free rate: 10Y German Government Bond (03/11/2023)	2,72%	
Average Market Risk Premium 2023	5,70%	
CAPM for Cost of Equity (annual)	27,23%	

C Appendix: Azure Cloud Cost Estimation for NPV Calculation

Microsoft Azure Estimate

Service category	Service type	Description	Estimated monthly cost	Estimated weekly cost	Estimated upfront cost
Database	Azure SQL Database	Elastic pool, DTU purchase model, standard tariff, 50 eDTUs: 50 GB included storage per pool, 100 databases per pool, 1 pool(s) x 730 hours, 50 GB storage, 15 GB point-in-time restore, ZRS redundancy for backup storage, 10 x 5 GB long-term retention	€134,12	€33,53	€0,00
Storage	Storage Accounts	Block Blob Storage, General v2, Hierarchical Namespace, ZRS Redundancy, Access Level: Hot, Capacity: 300 GB - Usage-based payment, 10 x 10,000 writes, 10 x 10,000 reads, 10 x 10,000 iterative reads, 10 x 100 iterative writes, 1,000GB data retrieval, 1,000GB data write, 10GB index, 1 x 10,000 other operations	€8,91	€2,23	€0,00
Analysis	Azure Data Factory	Azure Data Factory V2 type, Data Pipeline service type, Azure Integration Runtime: 30 activity execution(s), 149 data movement unit(s), 1,000 pipeline activities, 1,000 pipeline activities - external, Azure Integration Runtime for VNET: 10 activity execution(s), 60 data movement unit(s), 30 pipeline activities, 10 pipeline activities - external	€129,97	€32,49	€0,00
	Total		**€273,00**	**€273,00**	**€0,00**

Disclaimer

All prices shown are in Euro Zone – Euro (€) EUR. This is a summary estimate, not a quote. For up to date pricing
This estimate was created at 11/6/2023 10:11:32 AM UTC.

D Appendix: Industry Beta Table by Damodaran

Date updated:	*05-Jan-23*	
Created by:	Aswath Damodaran, adamodar@stern.nyu.edu	
What is this data?	Total Beta (beta for completely undiversified	Europe
Home Page:	http://www.damodaran.com	
Data website:	https://pages.stern.nyu.edu/~adamodar/New_Home_Page/data.html	
Companies in each industry:	https://pages.stern.nyu.edu/~adamodar/pc/datasets/indname.xls	
Variable definitions:	https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/variable.h	

Industry Name	***Number of firms***	***Average Unlevered Beta***	***Average Levered Beta***	***Average correlation with the market***	***Total Unlevered Beta***	***Total Levered Beta***
Reinsurance	4	1,04	1,09	40,44%	2,58	2,70
Software (Entertainment)	64	1,26	1,33	25,08%	5,04	5,31
Software (Internet)	30	1,04	1,16	23,78%	4,37	4,87
Software (System & Application)	352	1,12	1,14	21,86%	**5,13**	5,23
Steel	57	1,08	1,13	26,68%	4,05	4,25
Telecom (Wireless)	13	0,30	0,67	23,89%	1,27	2,81
Telecom. Equipment	54	0,96	0,96	22,15%	4,34	4,31
Telecom. Services	68	0,44	0,82	24,40%	1,82	3,36
Tobacco	6	0,29	0,39	24,03%	1,22	1,63
Transportation	35	0,85	1,05	28,25%	3,01	3,73
Transportation (Railroads)	5	0,81	0,82	25,27%	3,21	3,26
Trucking	30	0,44	1,03	24,19%	**1,83**	4,26
Utility (General)	17	0,47	0,82	32,91%	1,43	2,50
Utility (Water)	10	0,35	0,59	25,46%	1,36	2,32
Total Market	7126	0,78	1,05	25,44%	3,05	4,14
Total Market (without financials)	6279	0,88	1,07	25,22%	3,50	4,23

E Appendix: Transalb Kühl- und Express-GmbH: 2022 Financial Report

Transalb Kühl- und Express-GmbH

Stetten am kalten Markt

Jahresabschluss zum Geschäftsjahr vom 01.01.2022 bis zum 31.12.2022

Bilanz

Aktiva	31.12.2022 EUR	31.12.2021 EUR
A. Anlagevermögen	479.615,00	488.320,00
I. Immaterielle Vermögensgegenstände	1,00	128,00
II. Sachanlagen	479.364,00	487.942,00
III. Finanzanlagen	250,00	250,00
B. Umlaufvermögen	575.867,43	413.875,52
I. Vorräte	10.380,00	12.650,00
II. Forderungen und sonstige Vermögensgegenstände	242.965,55	225.802,93
III. Kassenbestand, Bundesbankguthaben, Guthaben bei Kreditinstituten und Schecks	322.521,88	175.422,59
C. Rechnungsabgrenzungsposten	6.863,00	6.000,00
Aktiva	1.062.345,43	908.195,52

Passiva	31.12.2022 EUR	31.12.2021 EUR
A. Eigenkapital	488.175,82	344.704,07
I. Gezeichnetes Kapital	26.000,00	26.000,00
II. Gewinnvortrag	318.704,07	222.247,20
III. Jahresüberschuss	143.471,75	96.456,87
B. Rückstellungen	60.997,36	55.812,66
C. Verbindlichkeiten	513.172,25	507.678,79
Passiva	1.062.345,43	908.195,52

F Appendix: Questions and Answers

Question: How many departments are involved or work with data?

- Accounting
- Billing
- Disposition
- Management
- Tracking of vehicles (any tractor unit, no smaller cars or trailers)

Question: Are there any dependencies

- Management: dependent on all others
- Billing is dependent on data from scheduling. There is a controlling function here, as the billing department checks the data from the scheduling department against
- Accounting is dependent on billing, controlling again via totals and balances list
- Tracking of vehicles is independent

Question: Responsibilities for reports, key figures, data entities:

- Accounting generates reports (monthly BWA, cost overview)
- Cost accounting is prepared once a year by the management.
- Each area is responsible for the data created there

Question: Which roles/employees work in the area of Business Intelligence?

- Management analyses financial reports
- Dispatching monitors vehicle tracking; downtimes are planned if the driver is not already experienced enough himself

Question: What is the current collaboration model for business <> IT?

- IT is involved in all business areas except accounting, accounting is provided externally
- IT is internal, departments call IT specialists directly, no need to go through detours

Question: Which dispositive use cases already exist or are being planned (e.g. financial reporting, process monitoring, bad case analyses, campaign reporting)?

- Financial reporting: Financial reports are created using accounting software, order data is recorded in MRP and automatically transmitted to the accounting department for invoicing.
- There is no process monitoring; since it is not a manufacturing company, there is also no bad case analysis.
- Actual case analyses are carried out. E.g. toll increase: the financial impact of this increase is analysed
- There are no marketing campaigns
- <u>In the planning stage:</u>
    - Dashboard for monitoring costs and sales per vehicle. Would work because cost centres are specified per vehicle. E.g. also external service costs by subcontractors
    - Automating manual work between accounting and profitability analyses for each lorry in the management. Currently, lists are exported in PDF from the accounting programme and inserted into the Excel template.

Question: How are use cases currently played out (e.g. Excel reports, standard reports, dashboards, data mining, raw data access):

- Generally, the accounting system delivers reports in Excel format

- No dashboards
- No data mining
- Raw data access would be possible
- Partly manual effort, see above

Question: Are there non-functional requirements from a functional point of view (e.g. performance)?

- None. User-friendliness ok, reliability of the systems ok, performance, e.g. real-time analyses are not required. Errors are controlled by double checks.

Question: How many reports and standard analyses already exist?

- Cost accounting and BWA every month

Question: What functionalities do they provide?

- Parameterisation (changing certain values in a report or query without changing the underlying code or the query itself) works
- Dynamic filter (select only France shows only costs for France): not in dashboard functionality, but there are filters by time, customer, etc.

Question: Approximately how many key figures and dimensions are there?

- Margins and operating results are calculated, personnel costs, fuel costs, kilometres driven, ... Estimate on 20 key figures
- Dimensions are e.g. profit and loss account per vehicle (thus also per driver, because drivers keep their vehicles and are usually not changed)
- Key figures and dimensions are documented

Question: What does the current BI system landscape / DWH architecture look like?

None.

Question: Source systems:

- Log-IT: Logistics software for planning and viewing customer orders and tours (recurring orders that are travelled to four times a day, for example)
- TimoCom: Freight exchange for finding (short-term) transport orders
- DATEV accounting software
- Tracking system: Webfleet
- Reporting: Microsoft Office Suite
- Communication via Whatsapp and email
- Route planning via Map&Guide from PTV
- Technical drawings of vehicles: Brics-CAD

Question: Are there plans to rebuild existing source connections?

- Yes, DWH should be connected to the sources mentioned above.

Question: Do data quality checks exist?

- Billing is dependent on data from scheduling. There is a controlling function here, as the billing department checks the data from the scheduling department against
- Accounting is dependent on billing, controlling again via totals and balances list

Question: Which data models exist?

- No DWH available yet.

Question: Industry standard?

- No DWH available yet.

Question: Need for customisation/expansion?

- No DWH available yet.

Question: Tools in use?

- None so far, all manual.

Question: Favourite software solutions?

- No.

Question: Which functionalities must, should or could be used?

- Analyses such as Existing customer: If we do business with customer X, we generally have to travel x kilometres. Or, for example, if we have an order beforehand and then do business with customer X, we have to pick up another trailer in the workshop beforehand, which makes the journey longer again.
- Interesting developments between years and postcodes for connecting loads. For example, vehicles that unloaded in location X always had to travel a long way to find a connecting load.
- Plan refuelling and fuel consumption more accurately with tracking software. Until now, consumption has been planned by drivers entering the mileage at the petrol pump, which is often inaccurate. Or software saves location if close to a petrol station via geofencing. Then you can use the fuel card to check whether the driver has filled up there. Or just check the driver's entry with this system, if the deviation is too large, take the value from the tracking system. Fuel consumption can then be averaged via billing and litres filled. The problem is that drivers often enter incorrect information and fuel consumption cannot be determined.
- **Central dashboard**: A central dashboard on which management can quickly view all relevant key figures and analyses. This could include turnover, margins, fuel costs, kilometres driven per vehicle and other important metrics
- **Automated reporting**: Instead of manually exporting data from the accounting programme and inserting it into Excel templates, a solution could be implemented that automates this process and reduces errors caused by manual intervention -> example of cost accounting from above.
- **Advanced planning tools**: A system that allows dispatchers to better monitor vehicle tracking and plan more efficient routes

G Appendix: Data Warehouse Implementation Proposal

Offer

Between

CINTELLIC GmbH
Remigiusstrasse
16 D-53111 Bonn
(hereinafter referred to as CINTELLIC)

and

Transalb Kühl- und Express GmbH
Ebinger Str. 50
72510 Stetten am kalten Markt Germany
(hereinafter referred to as Transalb)

about

"Conception and structure of a data warehouse"

Status:	03.11.2023
Contact person:	Alexander Faber
Offer number:	CCG- 002382

Table of contents

1 Provider presentation CINTELLIC Consulting Group

1.1 CINTELLIC Consulting Group

Innovative ideas and constant further development are the key to success. Our consultants create sustainable solutions and provide new impetus for the success of our clients. Our range of services includes business consulting and the management of complex projects. CINTELLIC provides holistic advice in the areas of customer experience management, customer relationship management and business intelligence.

Our consulting services include strategy development, business analysis and conceptualisation as well as the professional and technical implementation of the measures. In addition, we support our clients with the transfer to the line and are also available for further support. Our teams always focus on the needs and goals of our clients.

CINTELLIC offers particular advantages through expertise in all three areas relevant to the client:

- Industry expertise
- Expertise
- Technology expertise

The teams contribute expertise from all three areas and proactively draw on the know-how of the respective practices within the projects. This ideal consulting approach for the customer is rounded off by the quality management of the partners, who contribute their many years of expertise to the projects.

CINTELLIC clearly differentiates itself from the competition in the areas of customer experience management, CRM, business intelligence and campaign and sales management. By combining different consulting expertise, CINTELLIC is able to offer innovative solutions for its clients.

Thanks to the broad experience of CINTELLIC employees in different industries and CRM maturity levels of your clients, we know what advice our clients need at what point in their development.

1.2 Our added value for your project

CINTELLIC is one of the leading management consultancies for the digital transformation of customer management and CRM.

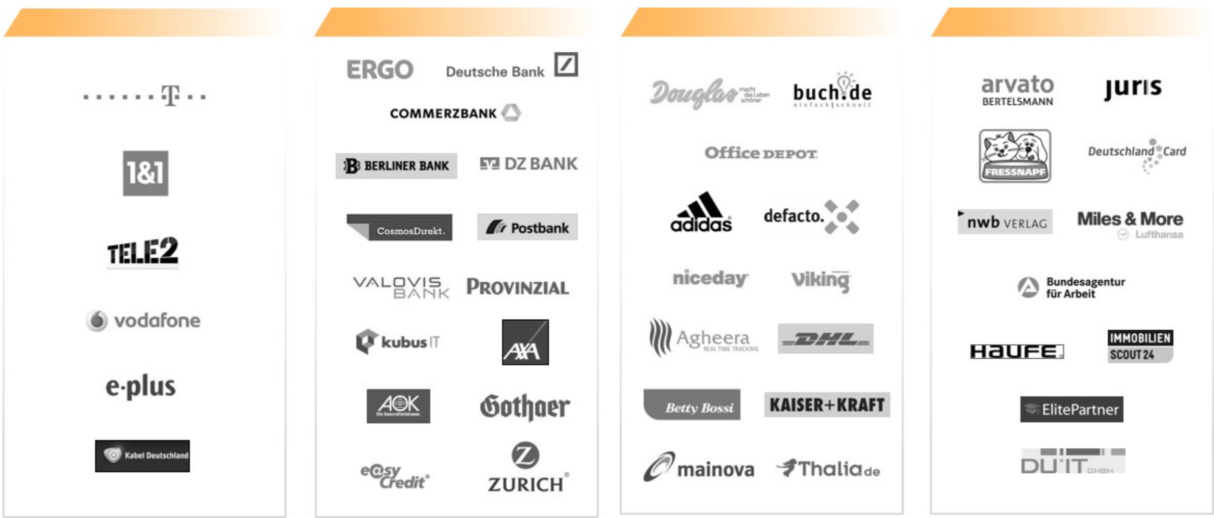
Due to the specialisation of its consulting expertise in customer management and CRM, CINTELLIC has a very high level of expertise in all aspects of customer management. Our expertise ranges from CRM strategies, processes and solution selection to customer analytics and DWH / business intelligence.

By combining business, data and technologies, CINTELLIC creates high added value for you, as the introduction of new technologies in your company is always carried out with a focus on the added value in business.

1.3 General references from CINTELLIC

As a leading management consultancy in holistic omnichannel CRM, CINTELLIC is in a position to manage corresponding projects from the conception and selection process through to the final implementation. We carry out your projects with great passion and expertise. Our holistic know-how - from strategy and approach concepts to business and process analyses, tool selection and implementation, customer insight and data warehouse / business intelligence - enables us to secure the future and competitiveness of our clients and increase sales, profits and customer satisfaction for you.

Our customers include the following companies:



CINTELLIC has many years of in-depth experience in the holistic realisation of CRM and customer management projects: from strategy to analysis and conception to implementation. Our expertise and experience in the following areas form the basis for effective and successful project implementation:

- Extensive experience in the profitable introduction of scalable CRM processes and solutions or optimisation of existing systems
- Linking best practice business (sales / profit increase) with best practice process expertise (automation, introduction, implementation)
- In-depth knowledge of current specialised CRM approaches and the underlying processes
- In-depth knowledge of current CRM solutions and their advantages and disadvantages
- Experience and expertise in data management and system integration
- Extensive knowledge in the field of data mining and statistical data analysis
- Standardised methodology for the implementation of CRM selection projects including a best practice CRM selection catalogue, which is adapted to the respective customer and at the same time provides a very good insight into current market requirements

1.4 Selected references specific to your project objective

Company	Detail
ITERGO	The aim of the project was to rebuild the data warehouse for sales management. New interfaces to all divisional systems and many other sales-relevant systems were set up and transferred to an integrated data model with data vault modelling. In this project, CINTELLIC supported the development of the architecture, business analysis and requirements engineering.
T	CINTELLIC integrated individual customer databases (mobile telephony, internet, TV, etc.) into an integrated customer database, which became one of the largest customer databases in Germany at the time. CINTELLIC developed the architecture, the data transformations including customer matching, data quality processes, testing and the establishment of a BI Competence Centre including the definition of the data steward role. In total, CINTELLIC developed over 7,000 analyses and models for holistic reporting.
TELE2	Due to the technical encapsulation of the parent company, the challenge was to rebuild the planning data landscape from scratch. CINTELLIC developed a three-layer DWH architecture to historicise all data from finance, contracts, customer master data, etc. and make it usable for CRM and BI.
unitymedia	CINTELLIC modernised the entire campaign management infrastructure, including all associated technical processes, organisation and planning processes. One part of the project was the creation of a data mart for campaign management, developed with Python / Apache Spark on HDFS.
DVV DUISBURGER VERSORGUNGS- UND VERKEHRSGESELLSCHAFT MBH	Following the de-monopolisation of the energy market, the challenge was to significantly expand the analytical CRM. CINTELLIC developed a customer data platform including historicised data from ERP, website and sales data. Based on this, a reporting and analytics tool was selected and introduced and a campaign management system implemented.

2 Initial situation and objectives

Transalb is planning to introduce a data warehouse (DWH) as a central element for the development of information management and reporting for the entire company.

The supply of information and the preparation of reports for corporate reporting and the reporting of operational processes are currently carried out by Transalb's organisational units in different ways and in most cases manually.

The aim of the project is to create a centralised data management system that calculates all relevant key figures and serves as the basis for automated reporting.

3 Procedure and consulting services

CINTELLIC is happy to support Transalb in the design and development of a comprehensive DWH.

We plan the structure of the DWH in 2 rough phases, a conception phase and an implementation phase. In the conception phase, the DWH is designed after a clean project setup, including the selection of the DWH software components. At the same time, the BI tool is selected.

In the implementation phase, we plan to implement the previously conceptualised DWH in an agile manner and enable users to work with the components that have been set up.

Figure 1 - CINTELLIC's planned process model

To begin with, we are planning a very lean setup phase in which we will achieve the following results as part of a kick-off workshop.

Phase 0: Project setup	
Observation object	**Description of the**
Scoping	Joint determination of the project scope and derivation of strategic objectives
Projectorga	Stakeholder definition and analysis as well as clarification of the project organisation and finalisation of the milestone planning

As a central data platform, the DWH should combine the data from all relevant productive systems in a common data model and thus form the data basis for use cases such as standard reporting, analytics and data provision processes. In addition to the existing productive data from all areas of the company, such as data from accounting and logistics software, it should be possible to integrate other file-based sources.

In our schematic target architecture, we consider it expedient to set up several suitable data stores (DS) in order to be able to process the planned use cases for specific purposes. The Integrated Data Store (IDS), in which the data is merged in a suitable data model in a historicised form, will play a core role.

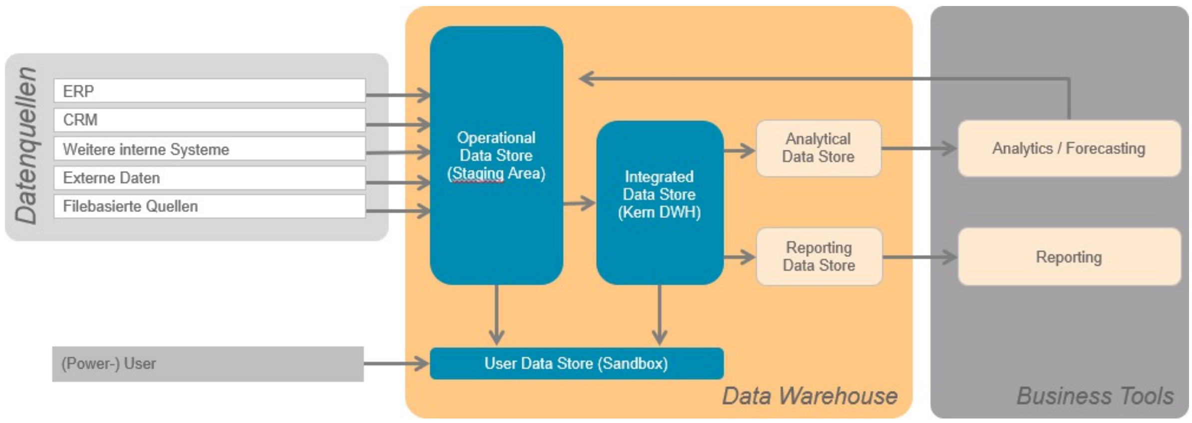


Figure 2 - Schematic CINTELLIC target architecture

In order to protect the productive systems, the first step is to copy the data from the operational systems to the Operational Data Store (ODS) or staging area for further processing.

An analytical data store (ADS) for processing DWH data in analytical models and evaluations and a reporting data store (RDS), which holds all reporting-relevant data suitable for the BI tool still to be selected, as well as a user data store (UDS) or sandbox, which provides power users with the option of carrying out analyses and quality checks directly on the complete data, are planned as customer systems.

For a better technical understanding of the data, we recommend the creation of a metadata catalogue in order to provide business and technical users with added value for a better understanding of the data and thus minimise the misuse and misinterpretation of data.

In order to create a successful DWH, a conception phase is required in which the topics from the following table are developed.

As part of the design phase, the guidelines of the EU General Data Protection Regulation (GDPR) are observed and incorporated into the technical design.

Phase 1a: DWH concept	
Observation object	**Description of the**
Planning target infrastructure	Selection of the appropriate cloud or on-premise environment and (cloud) components (data storage, ETL, monitoring, maintenance / backup) and definition of initial sizing and planning of the infrastructure setup
Selection of BI tool	Joint selection of a suitable BI & analytics tool for automated report generation
Testing the source systems (technologies)	Review of the IT landscape of the source systems and organisation of workshops with the respective system managers to discuss access, existing documentation and metadata definitions
Development of metadata catalogue	Creation of catalogue and filling of attributes for the technical description of the inventory systems
	Workshops with technical contacts to clarify the technical interpretation of the attributes, including Contexts
ETL	Modelling the future loading of the marketing data hub by the operational systems
Data modelling	Structure of the data model of the Integrated Data Store (IDS) for the Marketing Data Hub
History concept	Development of history concept for the IDS incl. best practices
Authorisation concept	Development of a role-based authorisation concept
Technical conception	Technical design of the systems, mappings, definition of the development packages, definition of the technology

Phase 1a: DWH concept	
Observation object	**Description of the**
Operation concept	Definition of the operator model incl. service level and coordination of the operating processes

After the design and tool selection, we recommend an agile, iterative implementation of the DWH in line with the requirements of the functional use cases. CINTELLIC carries out the implementation in close consultation with Transalb.

Phase 2: DWH implementation	
Observation object	**Description of the**
DWH development	Successful implementation of the technical concept
Testing & Documentation	Specialised Testing of development and documentation
Documentation	Documentation of the realisation
Roll-out planning and implementation	Planning the roll-out of the DWH
Structure of first reports	Joint creation of initial planned reports and empowerment of individual employees to create or modify existing reports
Technical training	IT training for the DWH components
Operation	Transfer to regular operation and handover of the application

4 Order processing

4.1 Period of service provision

The services will be provided by CINTELLIC in a period yet to be defined with Transalb. However, a start in 2021 is still desirable. CINTELLIC reserves a set-up time for the start of the project of ten working days after receipt of the order.

The assignment takes place on the premises of CINTELLIC or remotely in the home office of the assigned consultant. However, necessary on-site appointments can be guaranteed. Should the expenditure specified by CINTELLIC in the respective individual agreement increase, CINTELLIC shall only provide additional person days after written approval by the customer.

4.2 Obligations of the client to co-operate

Transalb provides a contact person for the respective project. The contact person must be communicated to CINTELLIC before the start of the project. On behalf of the customer, the contact person creates the framework conditions necessary for CINTELLIC to provide the service and gives CINTELLIC the technical and organisational possibility to provide the service described here within a reasonable period of time. The time period shall be agreed between the Customer and CINTELLIC and can also be replaced by an individual project plan.

Transalb undertakes to participate in all projects. The obligations to co-operate are defined and presented before the start of each project. These include

- Project management
- Detailed briefing on the general conditions and expectations of the Transalb
- Coordination of workshop dates with the individual contact persons
- Provision of documentation by Transalb (e.g. corporate strategy, current reports, necessary documents from the systems)
- Participation and collaboration in various workshops
- Participation in project team meetings

The project period must be agreed between Transalb and CINTELLIC and can also be replaced by an individual project plan.

Mr Alexander Faber is the contact person for CINTELLIC with regard to all aspects of project management and commercial issues.

4.3 The project team

CINTELLIC shall provide the service through qualified consultants. The final staffing shall be the responsibility of CINTELLIC and shall take place when the order is placed or the project begins. The level of qualification and area of responsibility will be taken into account.

The detailed profiles of the consultants on offer are enclosed as separate documents.

5 Commercial terms and conditions of CINTELLIC consulting services

5.1 Fee

The services to be provided shall be remunerated according to the time spent. CINTELLIC shall keep a record of the number of person-days used, which can be viewed on request. Invoicing shall take place on a monthly basis according to actual expenditure as documented.

The imputed fee serves as information and budget planning for the client and is non-binding. A working time of 8 hours per day is assumed. Additional or reduced services will be charged pro rata.

Phase 0: Project setup & kick-off

	Number of days estimated	Net amount per day	Total net
Management Consultant	1	€ 1.400,00	€ 1.400,00
BI & DWH Consultant	1	€ 1.200,00	€ 1.200,00
Total	**2**		**€ 2.600,00**

Phase 1a: DWH concept

	Number of days estimated	Net amount per day	Total net
Management Consultant	10	€ 1.400,00	€ 14.000,00
BI & DWH Consultant	15	€ 1.200,00	€ 18.000,00
Total	**25**		**€ 32.000,00**

Phase 2: DWH implementation

	Number of days estimated	Net amount per day	Total net
Management Consultant	5	€ 1.400,00	€ 7.000,00
BI & DWH Consultant	20	€ 1.200,00	€ 24.000,00
Total	**25**		**€ 31.000,00**

Total:

	Number of days estimated	Total net
Phase 0: Project setup & kick-off	2	€ 2.600,00
Phase 1a: DWH concept	25	€ 32.000,00
Phase 2: DWH implementation	25	€ 31.000,00
Total	**52**	**€ 65.600,00**

Travelling time, travel costs and expenses are not included in the daily rates and are paid separately. All fees are subject to VAT at the applicable rate.

5.2 Payment

In the case of project and individual services on a time and material basis, invoices shall be issued after completion of the service or at the latest monthly at the end of the month on the basis of the proof of performance provided by the project employee. Invoices are due for payment strictly net 14 days after invoicing.

5.3 Deviating special conditions

For activities that are invoiced on a time and material basis and are carried out at weekends, on public holidays or at night, the following invoicing method is agreed:

Work on Saturdays is subject to a 25% surcharge. On Sundays and public holidays, a surcharge of 50% will be charged.

Night work applies from 11 p.m. to 6 a.m. and is subject to a 25% surcharge.

Work at weekends, on public holidays and at night shall only be performed if urgently required and after prior agreement and corresponding approval. Otherwise no additional remuneration will be paid.

6 Supplementary provisions

6.1 Property rights and confidentiality

All information, documents and work results created or used within the scope of this project support are the exclusive property of the client.

CINTELLIC undertakes to maintain the strictest secrecy about all business and trade secrets of which it becomes aware during the planning and execution of the order, all business or operational events of which it becomes aware, even after termination of the agreement, and shall countersign a corresponding confidentiality obligation.

Notwithstanding the foregoing provisions, each party shall be entitled to disclose confidential information of the other party:

a. to the respective insurers or legal advisors, or
b. disclose to a third party if required to do so by a court of competent jurisdiction, governmental or regulatory authority, or if there is a right, obligation or requirement to disclose information

CINTELLIC is authorised to disclose confidential information to a third party if this is necessary for the provision of services. Such third party must agree in writing to comply with similar confidentiality terms. The same applies to the retention and use of working documents within the scope of this business relationship, which CINTELLIC may retain and use for CINTELLIC's internal use.

6.2 Fringe benefits

The customer agrees that CINTELLIC may name the service relationship with the customer as a reference and, in particular, refer to the provision of services to the customer in websites, print media and other advertising materials and use its logo. CINTELLIC reserves the right to prepare a project report after the successful completion of the project and to publish it after approval by the customer.

6.3 Liability

Both parties are mutually liable to ensure that the obligations defined in this contract are met.

With the exception of payment obligations, neither party shall be responsible for the non-fulfilment of obligations under this agreement for reasons of force majeure.

6.4 Non-solicitation clause

The client and all other companies within the Group are prohibited from poaching CINTELLIC employees.

If, within 12 months of completion of the final project, an employment contract is concluded with or the CINTELLIC employee is commissioned by another company, the relevant company shall be obliged to pay a contractual penalty amounting to 30% of the agreed annual salary (incl. bonuses) or the order value, but at least € 20,000.00.

6.5 Force majeure

With the exception of payment obligations, neither party shall be responsible for the non-fulfilment of obligations under this agreement for reasons of force majeure.

6.6 Notifications and changes

All notifications must be made in writing to the address of the other party stated in the order document or communicated at a later date. Amendments or additions to an agreement require the consent of both parties and the written form. This also applies to changes to the written form requirement. Additional or deviating conditions in a written communication from the customer (e.g. in an order) shall not become part of the contract.

6.7 Supplement

Postponements for which CINTELLIC is not responsible shall have no suspensive effect on the payment obligations.

The maximum liability limit is 100% of the commissioned value.

6.8 Duration of the contract and entry into force

The contract shall enter into force upon signature by both

parties. Cancellation must be made by registered letter.

Either party may terminate the contract early and with immediate effect if there is a serious breach of this contract and the defaulting party fails to remedy the breach of contract within 15 days of the written warning, or if one of the parties goes bankrupt.

Either party may terminate the contract at any time to the end of the following month. The prerequisite for this is that the contractual partner is immediately informed of the intention to terminate the contract in writing or by telephone as soon as the decision to terminate has been made.

6.9 Severability clause

Should individual provisions or parts of this contract prove to be void or ineffective, this shall not affect the validity of the remainder of the contract. In such a case, the contracting parties shall adapt the contract in such a way that the purpose intended by the invalid or ineffective part is achieved as far as possible.

6.10 Applicable law / place of jurisdiction

This agreement shall be governed by German law to the exclusion of its conflict of law provisions. The exclusive place of jurisdiction for all disputes arising in connection with this agreement is Bonn, Germany.

7 Offer status

We are committed to this offer until 31 December 2023.

Alexander Faber (mobile: +49 (0)176 21 70 67 99, e-mail: alexander.faber@CINTELLIC.com) will be happy to answer any questions you may have.

Yourconsent provided, please we you us the offer signed and returned to us.

In anticipation of a positive response, we remain, with

kind regards

Bonn, 03.11.2023

Dr Jörg Reinnarth Managing Director

We hereby confirm that we authorise CINTELLIC GmbH to provide the services under the conditions contained in this letter.

Click or tap here to enter text,

Place, date

Signature

Signature

Company stamp

End of Master Thesis II