Designing a Data Platform Consolidation

A Case Study

Niklas Harjunpää

Degree Thesis

Information technology

2020

EXAMENSARBETE	
Arcada	
Utbildningsprogram:	Informationsteknik
Identifikationsnummer:	8012
Författare:	Niklas Harjunpää
Arbetets namn:	Planering av en konsolidering av dataplattformar: fallstudie
Handledare (Arcada):	Dennis Biström
Uppdragsgivare:	Fortum Oyj

Sammandrag:
Examensarbetet fokuserar på att samla in nödvändig information för att kunna planera en konsolidering av dataanvändningsfall mellan de olika dataplattformarna och föreslå en möjlig framtida lösning för dataplattformarna. Resultatet kommer att hjälpa företaget att öka sin förståelse för de olika dataplattformarna och att hitta potentiella överlappningar mellan dem. Examensarbetet tar inte ställning till teknologival. För att samla in användningsfallen användes olika kvalitativa metoder, såsom dokumentationsanalys, fördjupande intervjuer och enkäter. Fallföretagets dataplattformar består av data warehouses och data lakes. För att kunna planera konsolideringen, utfördes en djupgående jämförelse av verksamhetens målsättningar, dataändpunkter och visualiseringsändamål mellan dataanvändningsfallen. Resultatet visar att det finns potentiella överlappningar mellan fallföretagets olika dataplattformar. Förutom att samla in nödvändig information och visa potentiella överlappningar, ger examensarbetet två olika teoretiska lösningar för konsolidering. Båda lösningarna strävar efter att bevara alla användningsfall och deras funktionaliteter, men har sina egna för- och nackdelar. Ifall man skulle vilja implementera någondera av lösningarna, skulle både en djupare forskning för varje dataanvändningsfall och en teknisk genomförbarhetsstudie vara nödvändig.

Nyckelord:	data warehouse, data lake, dataplattform, dataanvändningsfall, platform konsolidering
Sidantal:	37
Språk:	Engelska
Datum för godkännande:	17.12.2020

DEGREE THESIS	
Arcada	
Degree Programme:	Information technology
Identification number:	8012
Author:	Niklas Harjunpää
Title:	Designing a Data Platform Consolidation: A Case Study
Supervisor (Arcada):	Dennis Biström
Commissioned by:	Fortum Oyj

Abstract:
This thesis focuses on gathering the necessary information to be able to design a consolidation of data use cases between the different data platforms, and also proposes a possible future solution for the data platforms. As a result, this will help the case company increase their understanding of these different data platforms and to find potential overlaps between them. This thesis does not take a position on technology choices. The methods used to collect the use cases is a combination of different qualitative methods, such as documentation analysis, in-depth interviews and single question surveys. The different data platforms at use at the case company consists of data warehouses and data lakes. To perform the consolidation, an in-depth data comparison was done on a use case level, comparing strategic business goals, data endpoints and visualization purposes. The results show that there are potential overlaps between the case company's different data platforms. In addition to gathering the data and showing potential overlaps, the thesis provides two different solutions for consolidation on a theoretical level. Both solutions strive to preserve all the use cases and their own functionalities, but have their own benefits and shortcomings. If one would want to implement either of the designed solutions, a deeper research for each for each data use case would be necessary, as well as a technical feasibility study.

Keywords:	data warehouse, data lake, data platform, data use cases, platform consolidation
Number of pages:	37
Language:	English
Date of acceptance:	17.12.2020

CONTENTS

Figures

Tables

1 INTRODUCTION

When companies extend their operations geographically, acquire other companies or otherwise just grow their operations enough, there commonly is a good reason to start investigating which IT systems could be consolidated. IT consolidation means to combine several systems or things into a single more coherent unit. There are several benefits of consolidating IT systems, such as lowering operational costs, creating more centralized and coherent systems, improving data management and much more. In a study conducted by Forrester Consulting, they found out that:

> Companies who have successfully consolidated to a single platform report shorter time-to-market with new apps, more business agility, better innovation, and increased security. Platform consolidation also improves employee productivity inside and outside of IT, increasing employee satisfaction and making the company more attractive for hard-to-hire talent. (Forrester Consulting 2018: 3)

Thus, companies have great reasons to investigate their different IT systems to possibly find consolidation opportunities.

1.1 The Case Company

The case company of this study is Fortum Oyj. Fortum is an energy company that develops and provides its customers with solutions related to electricity, heating, and cooling. Their business includes operation and maintenance of power plants, production and sales of electricity and heat, and other energy-related services. In addition, Fortum provides services such as recycling and waste services, environmental expert services, and recycling services. Curiosity, responsibility, honesty, and respect form the foundation of Fortum's corporate culture. These values have guided their different opportunities and decision-making. Fortum has for decades focused on reducing CO2 emissions. Their mission and strategy is to drive change towards a cleaner world by reshaping the energy system, improving resource efficiency and providing smart solutions. (Fortum 2020)

1.2 Background of the Thesis

Businesses often grow over time and create several different IT systems. When businesses focus on growth, IT consolidation often becomes imperative for the growth. Every day, data is generated and collected from different systems and other sources. This data is often aggregated into a single central storage and can be stored in several different ways, such as databases, data warehouses, data lakes, etc. This data is often for further purposes, such as machine learning, artificial intelligence, and business analytics. Therefore, the data storage is a central part of the data architecture and the data platform.

Currently the case company has several different data platforms within the Customer and Sales data domain. The company wants to increase the understanding of these different data platforms and to find out the potential overlaps between the different platforms on a data use case level. This study will provide the necessary data for this and a possible solution for the future, that will strive to combine the best elements of the existing data platforms.

1.3 Objective & Scope of the Study

The objective of the study is to gather the necessary information to be able to design a consolidation of the data use cases between the different data platforms and propose a possible future solution for the data platforms. The study focuses on examining the different data use cases within each data platform. The data collection will be done manually by reading documentation, as well as by interviewing Business and IT representatives. Once all the data has been collected, the target is to design a possible future solution for the data platforms. It is important to note at this point, that the objective of the study is not to decide on what technologies to use, but to help increase the understanding of the data platforms.

Within the scope of this study are the different Customer and Sales data platforms the case company has in use. The study will focus on gathering the correct level of data and formulating the correct questions related to these platforms. The study will also focus on

providing a solution based on flexible architecture designed with coherent use cases in mind.

2 DATA STORAGE SOLUTIONS

All data that is generated and collected needs to be stored somewhere. There are several data storage solutions for this purpose and some of these are better at certain tasks than other. That is why it is important to think about what aspects are important to the specific use cases of the data. These aspects can include factors, such as accessibility, data storage format, storage flexibility, security or some other aspects that are more important to the specific use cases. While some solutions are designed for querying and analyzing data, and others are designed to provide the data to other components. This section introduces the basic theory of different data storage solutions, data storage pipelines, as well as some basic techniques and data warehouse architecture.

2.1 Data Warehouses and Data Lakes

A data warehouse is a system that structures all the best sources of data into a single, central, consistent data storage system that supports data analysis, data mining, artificial intelligence (AI), and machine learning (IBM 2020a). Data warehouses are designed for easy querying, reporting, and analysis. Older data warehouses are for the most part hosted on-premise, but more recently they have been hosted in the cloud with added functionality such as, extensive analytic capabilities and data visualization tools.

Figure 1 illustrates the architecture of a basic data warehouse. The data warehouse gathers raw data from different sources. The sources can be operational databases, external sources or even something other like files or data generated by Internet of Things (IoT) devices. The staging area is used to extract the data from the different sources, minimizing the number of operations on the source systems and to reduce the time to extract the data from them. After data has been loaded into the data staging area, it is used for data cleansing, data transformation and data combination. Data staging areas are often split by individual business units. After processing, the data is loaded to the warehouse where all the different summary, detailed, and raw data is stored with the relevant metadata. Finally, the data is accessed by the end-users to let them perform analysis, reporting, queries, data mining, machine learning, and artificial intelligence related tasks.

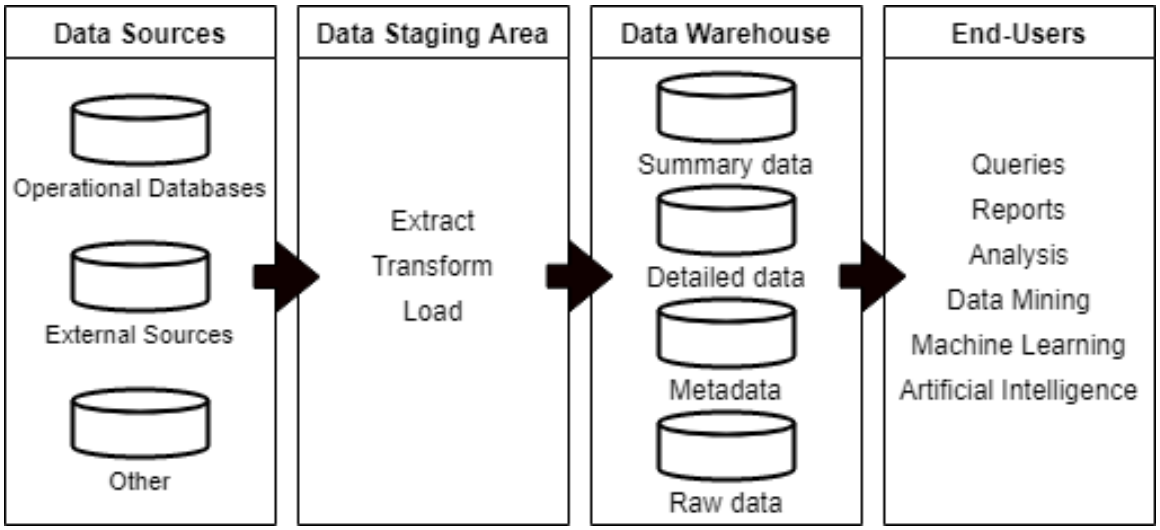

Figure 1. Architecture of a basic data warehouse

Even though both data warehouses and data lakes are used for storing data, they do have some differences. The main difference between data warehouses and data lakes is that a data warehouse stores processed and structured data, while a data lake stores raw data. Data lakes are great for storing vast amounts of transactional data in its original form. Data lakes can store both unstructured and structured data. As shown in figure 2, data lakes are a single centralized repository that can be scaled up and support all data formats. Because the data is left in a raw format, to be able to provide data exploration and visualization for end users, data that is extracted from the data lake often needs to be transformed. However, some of the data can be used in its raw format depending on the use. A data lake layer can be implemented before the data staging area in a data warehouse, to be able to provide the raw data to endpoints that might need it and to improve the flexibility of a data warehouse.

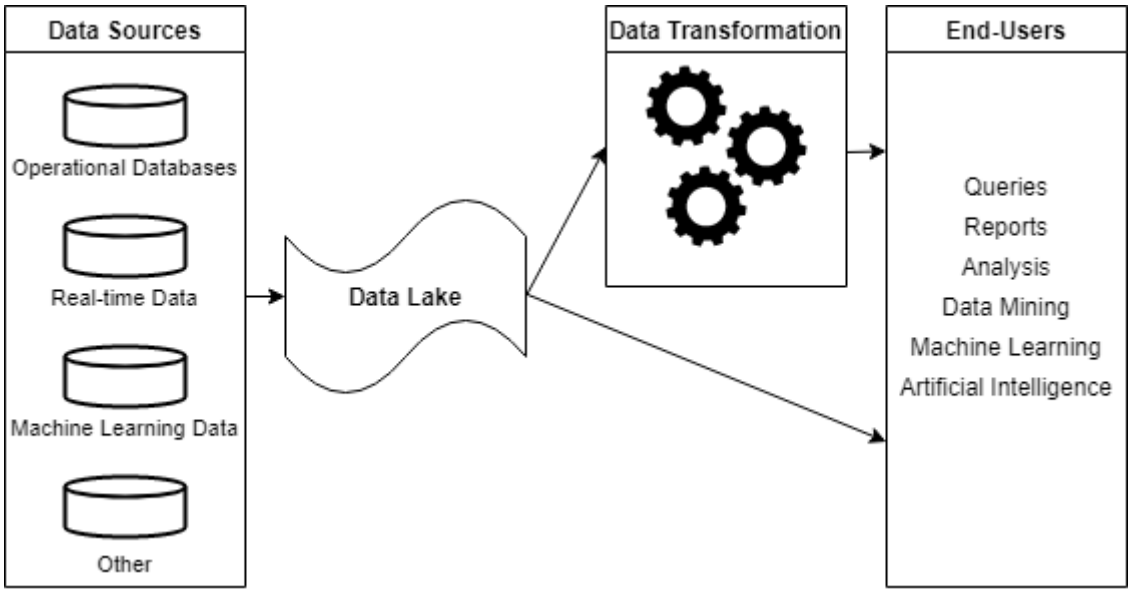

Figure 2. Architecture of a basic data lake

2.1.1 ETL and ELT

ETL (Extract, Transform, Load) has been the standard data integration process for data warehouses (IBM 2020b). The alternative to this is an ELT (Extract, Load, Transform) approach which changes position of the transform and load step. As shown in figure 3, the ETL pipeline extracts raw data to a staging area from where it is then transformed and loaded to the data warehouse. This way, raw data is prepared to data that can be used with less transformations before transferring it to Business Intelligence tools.

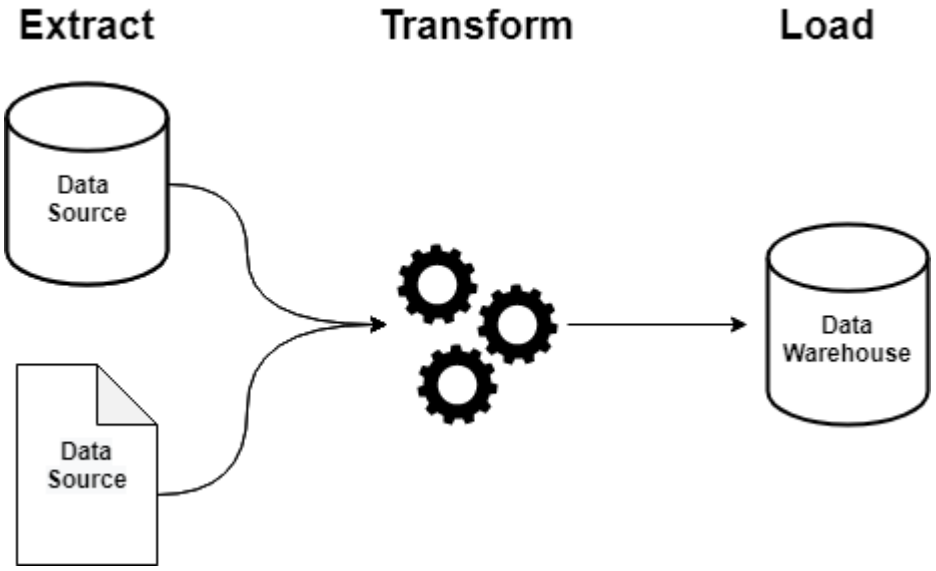

Figure 3. A basic representation of a ETL pipeline

Figure 4 presents the basic ELT process. By comparing Figures 3 and 4, you can see the difference between ETL and ELT. ELT changes the approach from ETL so that the data cleansing and transformation happens after the loading process. The change of when business logic is applied enables both untransformed and transformed data to be stored in the data warehouse. The raw data can be used in data visualizations if needed and the transformation can more easily be changed when necessary. The data is transformed on platform side instead of in-flight as in the ETL pipeline.

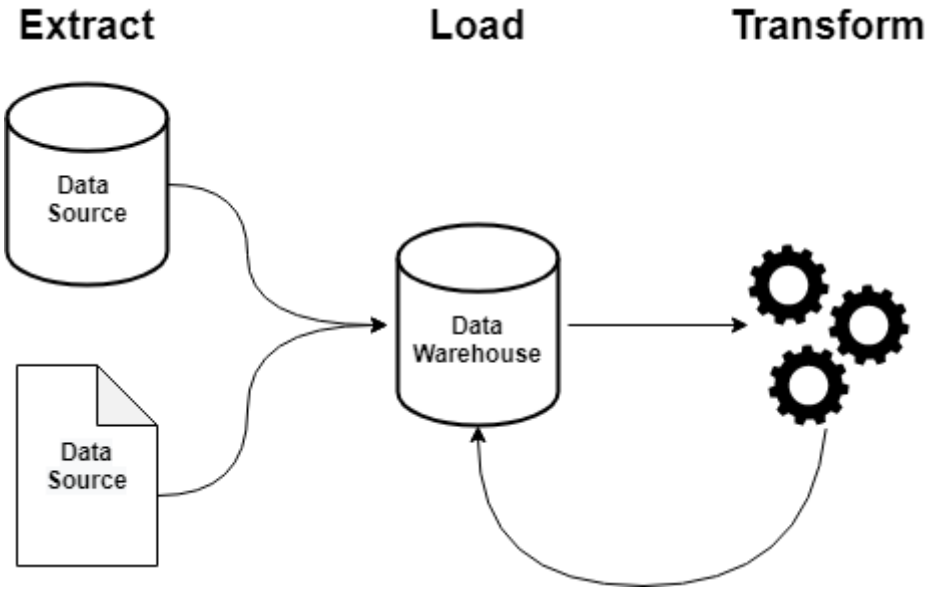

Figure 4. A basic representation of a ELT pipeline

2.2 Data Vault 2.0 Methodology

Data Vault 2.0 is an open standard created by Dan Linstedt, which is based on the Data Vault architecture, which was also invented by him. (Linstedt & Olschmike 2016: 8). The original Data Vault was mostly a modeling architecture for data warehouses. Contrary to popular belief, Data Vault 2.0 is not only a modeling architecture like its predecessor, but also an entire methodology for building data warehouses. Data Vault 2.0 is relatively new and is being continuously developed. “It is derived from core software engineering standards and adapts these standards for use in data warehousing” (Linstedt & Olschmike 2016: 55).

Each data storage method has their own benefits and drawbacks. Data Vault 2.0 has been designed to address the problems of agile development, supporting larger data sets, data protection, real-time processing, and to increase use of data across whole organizations. Data Vault 2.0 modeling is great for when a project needs to be developed in an agile fashion and when data scalability is a major factor for operations. The visualization is built sequentially based on business demands. The data warehouses are designed to support multiple different sources that can handle different data structure types (structured, unstructured, and semi-structured data). (Linstedt & Olschmike 2016: 23-31)

There are also downsides to using Data Vault 2.0 methodology when building warehouses. While Data Vault 2.0 is scalable, it can also mean that it can produce a lot of tables. In most cases, this is not a problem since business rules are often built on the data extraction end of a data platform. Developing a data warehouse with Data Vault 2.0 can require a wide range of technologies. Maintenance can become problematic for less experienced teams. Data Vault 2.0 is great for historical data, but implementing it on a project that requires a simple database is excessive. Another downside to this methodology is that it is still relatively new and it might be hard to find resources about it. (Linstedt & Olschmike 2016: 23-31)

2.3 Data Platform

A data platform is a solution for processing and analyzing the data that is in a data storage system. A data platform is built to provide end-to-end data management. Data platforms are used to improve the ability to learn and act on all and any of your data. (Splunk 2020) Figure 5 illustrates the difference between the data architecture and a data platform. The difference is that data architecture consists of the data sources, data processing, and the data warehouse, meanwhile the data platform consists of the data warehouse and the different endpoints. Notice that the data warehouse is a part of both the data architecture and the data platform simultaneously. Data architecture is solely responsible for extracting, storing, and delivering the data to different endpoints. Data platforms in turn access, analyze and validate the data for end-users.

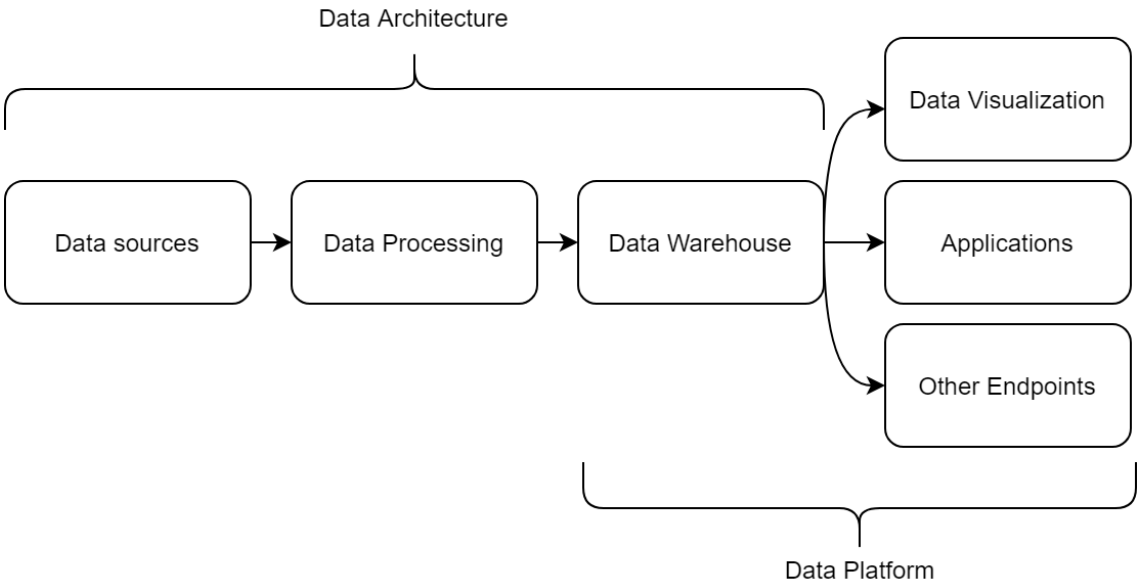

Figure 5. A basic representation of a data platform

2.3.1 Data Visualization

Data visualization is the graphical representation of data. All the data that is collected is useless if no one can understand it. That is why data visualization is used as a tool for Business Intelligence to understand and visualize data. One value that is often measured is a Key Performance Indicator (KPI). Data visualization and KPIs makes the data more accessible. There are many ways to visualize data, such as programming language libraries Plotly (Plotly 2020) and D3 (D3 2020), analytics platforms Qlik Sense (Qlik 2020) and Tableau (Tableau 2020), or even software such as Microsoft Excel (Microsoft 2020). Some of the platforms or software will require specific input formats like data bricks or cubes to take full advantage of them. These solutions often create reports that give insights to business or fulfill their other needs for reporting.

2.3.2 Applications

Data can also be fed to other endpoints, such as Business Intelligence analytics platforms or the data could be directly accessed from the data warehouse. Applications that do not fall into the data visualization category are for example different Business Intelligence

platforms that use data for automated marketing campaigns and customized email marketing letters, such as Hubspot (Hubspot 2020).

2.3.3 Other Endpoints

The same data that has come into a data warehouse could be fed back to its source after cleaning and filtering. Other endpoints might also consist of external systems that need specific data sent to them. The data is often transferred in a file or with direct access to the data warehouse. Servers and IoT devices are examples of these endpoints, just to name a few.

3 DATA COLLECTION

This section provides an overview of the methodology that was used to collect data. The focus is on collection methods and how questions, use cases and data use case information columns were defined.

3.1 Data Collection Methodology

This study analyzes the current state of the platforms. The underlying research questions behind this analysis are: “Are there overlaps between the different data platforms?” and “What would a future hybrid-solution of the platforms look like?”. These questions arise from the fact that the case company has wanted to increase the understanding of these different data platforms, find potential overlaps between them, and wanted to see how a possible future solution could look like. With these research questions defined, we still need to decide on:

- Where should data be collected from?
- What type of data is needed and what can be discarded?
- How to be able to use the data for consolidation purposes?

To gather data about the different platform data use cases, a combination of different qualitative research methods were used. Qualitative methods can be used to gain knowledge about a problem or to help develop ideas. The qualitative research methods used to collect the data were in-depth interviews, documentation analysis and single question surveys sent by email. Figure 6 illustrates the workflow for collecting the data. Since the platforms had much documentation and personnel knowledgeable about the platforms, the first step for each platform was to gather a rough overview of the data use cases per platform followed by in-depth interviews to validate the findings and fill in missing information. After the interviews, the use cases were adjusted and consolidated to be on a high enough level. Even after the interviews, there were missing information of certain use cases. To gather the missing data, single question surveys were sent by email. After all data was collected, the data use cases were unified across the platforms to be able to plan a consolidation more easily. The same data collection process was repeated for each

platform. Because the data use cases of Platform 3 were recently documented and validated, the use cases have been copied over and unified to fit with the rest of the use cases.

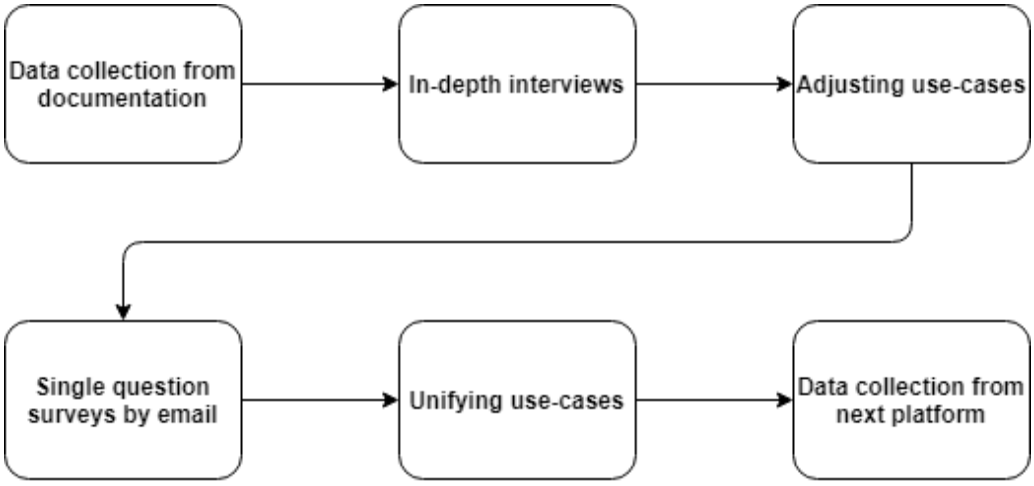

Figure 6. Data collection workflow

3.2 Data Column Definitions

To be able to find possible overlaps between the different data platforms, the data that is collected needs to be on a high-level. The idea of a consolidation is to find general overlaps between two or more systems that could be combined. Documenting each report and use case would provide a too detailed view and could not provide sufficient information for platform consolidation. Several of the use cases in this study include more than one report or use case for the data. Each use case has been aggregated to fit a core strategic business goal. To identify the different data use cases, Cockburn's guide to "Writing Effective Use Cases" (Cockburn 2000) has been used as a basis for the use case collection. Cockburn's guide is one of the first and most thorough guides on writing use cases and has therefore been chosen as a basis for this thesis. According to the guide, it is best to work top-down first. Figuring out the scope, actors, goals and main story is the most important steps to writing use cases. Since the main actors for most of the data use cases are business users, the use cases have been designed with them as the actors. For this study, the main goal is to identify what kind of data is used and for what purposes. This differs from Cockburn's guide, where his definition is to find all the possible scenarios between the system and its various actors. In this study we want to find all the current

goals from the point of an actor on the platforms. To be able to design the consolidation, the following information was collected for each data use case:

- Platform
- In Use
- Current Data visualization solution
- Data Endpoint
- Use case
- Visualization purpose
- Strategic business goal
- Use case impact

An example of the data that was collected can be found in Table 1. The “Platform” field identifies to which platform the data use case belongs to. The “In Use” field explains if the use case is currently used in some endpoint. The “Current data visualization solution” shows if a specific Business Intelligence tool, database access or data transfer is used to later be able to visualize the data. The “Data Endpoint” field depicts what endpoint type data is provided to. The “Use Case” column is the general name for what the data is used. “Visualization purpose” illustrates in what format the data is mainly used for. This can be either reports, ad hoc reporting, files, or database tables. The “Strategic business goal” describes what the goal of the data use case is. This describes what the data use case aims to achieve. “Use case impact” describes the benefits of using a data use case. All collected data use cases can be found in Appendices 2,3 and 4 containing Platform 1, 2 and 3 data use cases respectively.

Table 1. Example of collected data

Platform	In Use	Current Data visualization solution	Data Endpoint	Use Case	Visualization purpose	Strategic business goal	Use case Impact
Platform 1	Yes	Data Transfer	Applications	Marketing	File	To provide customer and contract data for a Marketing BI-tool	Ability to provide personal marketing
Platform 2	Yes	Data Transfer	Other Endpoint	GDPR	Ad hoc	To be able to provide use GDPR requests	Increase effectiveness of GDPR workorders
Platform 3	Yes	BI Tool 3	Data Visualization	KPI Management	Reports	To provide the most important KPIs for all brands	Ability to compare figures, performance tracking and improved insights across business units

3.3 Data Collection from Interviews

Interviews were conducted to confirm and gather the missing data that was not present in documentation. Interviewees were representatives from IT and Business. As mentioned in the methodology, Platform 3 had recently documented their use cases. Since the documentation was sufficient, there was no need to have interviews with Business or IT representatives. Table 2 is a detailed list of all the interviews that were held. Some data has also been collected by emails, when specific persons knew more about a data use case. Interviewees were experienced in their respective areas, having several years of work experience in the case company and with several data use cases. All interviews followed the same structure. Each interview started with a brief introduction on why this data is collected, on what level the answers should be described and examples from previously collected data. All questions were based on Appendix 1. The questions where designed with the business goals in focus and the interviews were documented with notes and with the use case tables, Appendices 2,3 and 4. Swedish, Finnish, and English was used as languages for the interviews, based on which felt most comfortable for each representative.

Table 2. Detailed table of the interview schedule

Date	Duration	Platform	Topic	Interviewee position	Style of documentation
28.9.2020	30 min	Platform 1	Data visualization Endpoints	Business owner	Appendix 1. was used for questions. Notes were taken during interview
30.9.2020	30 min	Platform 1	All Endpoints	Data Engineer	Appendix 1. was used for questions. Notes were taken during interview
12.10.2020	45 min	Platform 1	All Endpoints	Business owner	Appendix 1. was used for questions. Notes were taken during interview
19.10.2020	45 min	Platform 2	Application Endpoints	Data Engineer	Appendix 1. was used for questions. Notes were taken during interview
22.10.2020	45 min	Platform 2	Application Endpoints	Data Engineer	Appendix 1. was used for questions. Notes were taken during interview
29.10.2020	30 min	Platform 2	Data visualization Endpoints	Data Engineer	Appendix 1. was used for questions. Notes were taken during interview

4 CURRENT STATE

This section provides the overview of the current architecture and state of the different data platforms. This section also gives a brief overview of the current data use cases and a short description of each platform.

4.1 Architecture of Current Data Platforms

The current state of the Customer & Sales data platforms includes three different data platforms that were originally built for different brands. The different brands have their own goals and data use cases they want to utilize, but the majority of the use cases should be comparable, considering they all are related to customer and sales data. All platforms are built to support a 360-degree customer-oriented (Digital Marketing Institute 2020) data-driven approach. The biggest difference between the platforms are their architecture and purpose they were originally built for. Platform 1 is a cloud-based data warehouse and analytics platform. Platform 2 is also a data warehouse and analytics platform, mostly hosted on premise. Platforms 1 and 2 are both data warehouses while Platform 3 is a cloud-based data lake and analytics platform. Platforms 1 and 2 can in theory be easily consolidated as they are both data warehouses, but to be able to consolidate the data from Platform 3, which is a data lake, some modifications need to be done. The data needs either to be transformed, to have a data lake before the data warehouse or use an data warehouse with an ELT pipeline.

4.2 Analysis of Current Use Cases

All the current data use cases for each data platform have been collected to a table. These tables of the different data use cases can be found in the Appendices 2, 3 and 4 containing the data use cases of Platforms 1, 2 and 3 respectively. All the platforms use some sort of data visualization tool to be able to gain insights from the collected data, but they do also provide data to other endpoints as well.

Some of the data use cases did not have information available at the time of the study. Most of these unknown data use cases were old and there were not people around who knew what they were used for. These redundant data use cases will not be included in the possible future solutions. Even though these data use cases will not be included in the possible future solution, that does not mean that the future solution would not be able to provide the data for the excluded use cases in case someone would want to implement them later.

4.2.1 Platform 1

In general, Platform 1 is a typical data warehouse and analytics platform where most of the data use cases are located in the data visualization endpoints. Platform 1 consists mostly of data visualization endpoints with a few application and other endpoints, as seen in figure 7. For data visualization, Platform 1 utilize two different Business Intelligence tools. Since several of the use cases from Business Intelligence tool 2 have been moved to Business Intelligence tool 1, those particular use cases are no longer in use or there were no people who knew if they were in use and what they exactly were used for. Since the data use cases were unclear and the reports had not been run for a long time, these use cases will not be included in the possible future solutions. Figure 8 illustrates the data use case endpoints that are currently used and needed. The application endpoint data use cases that exist in Platform 1 are used to either provide a Marketing Business Intelligence -tool data for personal marketing or provide electricity related data to end user application. Both endpoints provide a file with data, that is transferred to their respective systems. The other endpoints for Platform 1 include a file of data transferred to Platform 3 to be able to provide global reporting and a file with customers to be excluded from call lists.

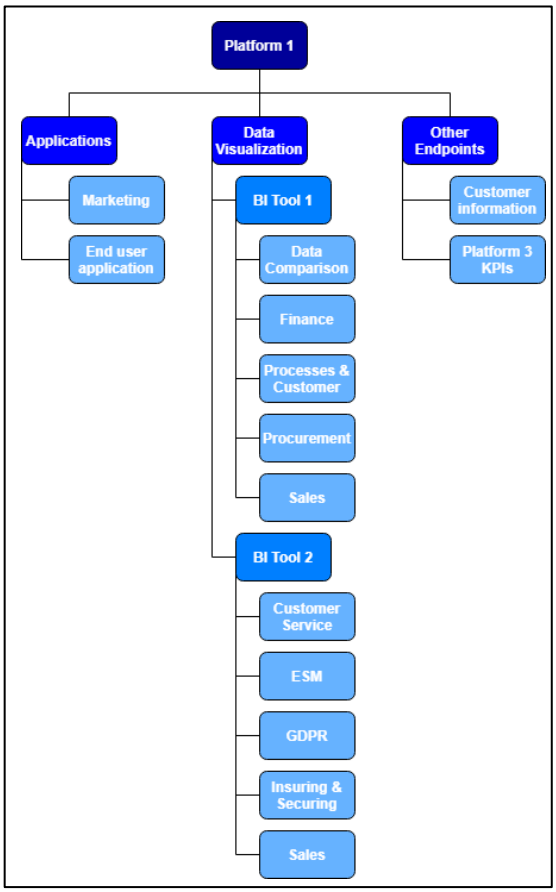

Figure 7. Current state structure of Data Platform 1

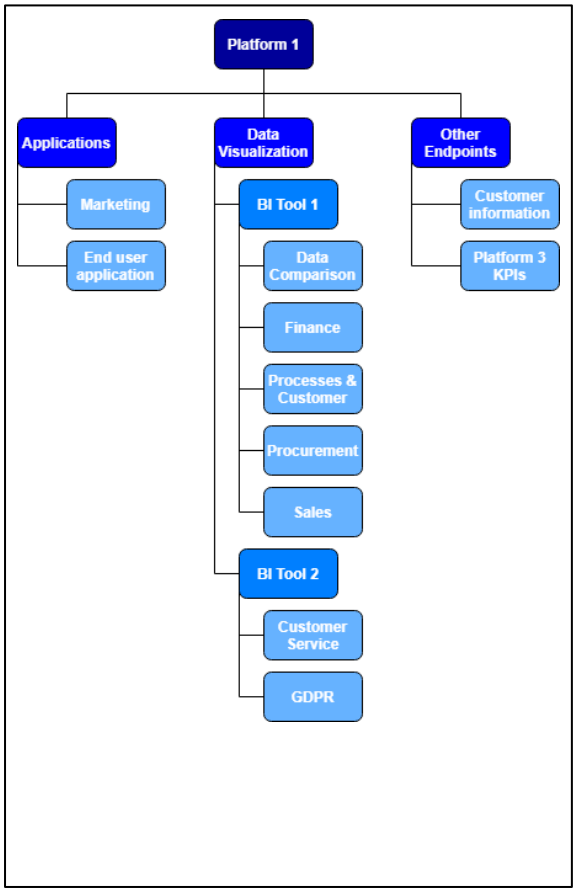

Figure 8. Current state structure of Data Platform 1 with redundant use cases removed

4.2.2 Platform 2

Like Platform 1, Data Platform 2 is a data warehouse and analytics platform where most of the data use cases are in the data visualization endpoints. Along with the data visualization endpoints, Platform 2 provides many other data use cases to several different Partners, as shown in figure 9. As a difference from Platform 1, most of the data visualization data use cases are for Business Intelligence tool 2. Some of the data use cases for Business Intelligence tool 1 are proof of concepts or did not have sufficient information available, and thus these use cases will not be included in the possible future solutions. Because the need for providing data to Partner 2 is expiring during 2021, this use case will not be used either. Figure 10 illustrates the data use case endpoints without the redundancies. Even though the “customer journey” data is not currently used on this platform, there is a use case for it in Platform 1 and this could easily be consolidated with it. Most of the data

visualization endpoint data use cases have the same strategic business goal as in Platform 1 and have great potential to be consolidated. The major difference is that Platform 2 provides data to many different partners, which could not be consolidated, but would need to be moved to the new solution. As for the application endpoint data use cases, Platform 2 has opted to use database access and file transfer as a means for the data use cases. One use case that stands out right away is end user application data use case, which is also found in Platform 1. This is a great sign to possibly find several other use cases for consolidation. Most of the other endpoints provide data to partners for analysis or operation purposes. The endpoints provide data transfer either by file transfer or database access.

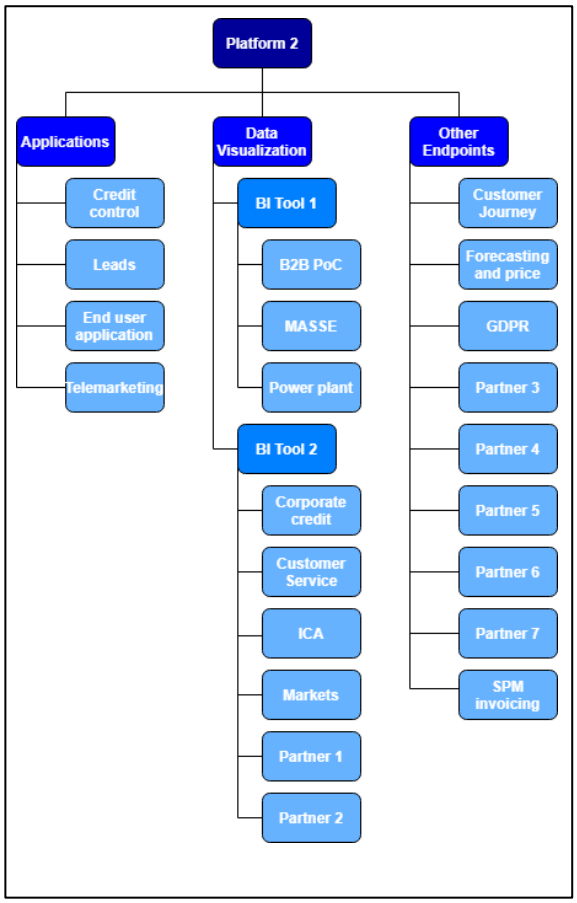

Figure 9. Current state structure of Data Platform 2

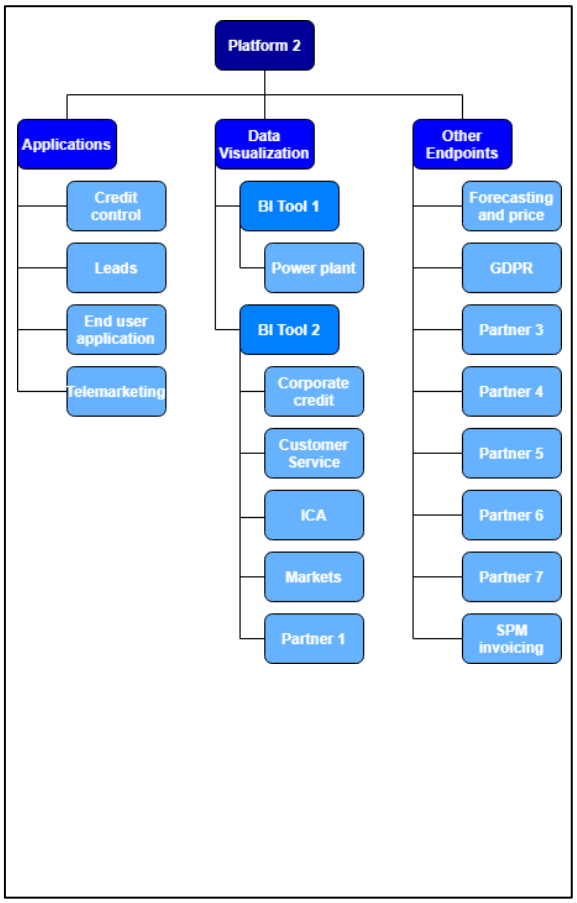

Figure 10. Current state structure of Data Platform 2 with redundant use cases removed

4.2.3 Platform 3

This platform is the odd one out and has a slightly different purpose than Platform 1 or 2. Since it is a data lake, it can store both structured and unstructured data. This platform was built to harmonize KPIs, business logic and definitions, and to provide comparable results across all existing brands. There might be some parts that can be consolidated, because at its core, the platform was created to reduce time spent on reporting. Figure 11 shows the structure of the platform. Platform 3 uses a different Business Intelligence Tool for visualization than the previous two platforms. Most of the data use cases are provided to other endpoints. These endpoints provide data for analysis purposes to many different visualization solutions.

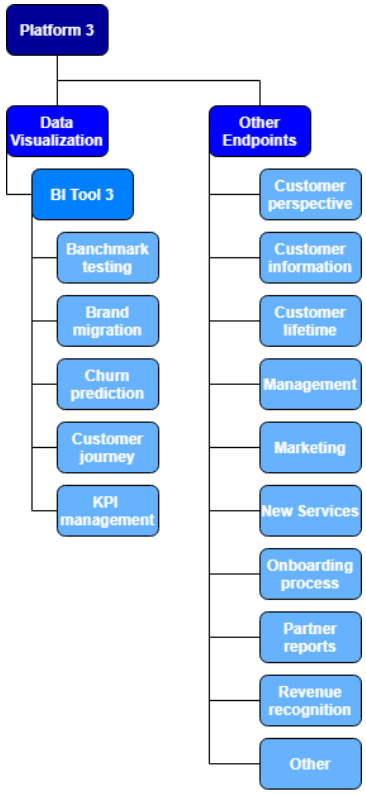

Figure 11. Current state structure of Data Platform 3

5 DESIGNING A SOLUTION

Designing a consolidation of platforms is often a lengthy and complex process that has many risks and is often difficult to design. Every solution has its pros and cons. This study presents two different solutions, one solution of a single platform based on Forrester Consulting's findings (Forrester Consulting 2018: 3) and another solution based on the strategic business goals of the original data platforms. Data Vault 2.0 methodology is used in the design process due to the benefits it offers. It is great for iterative implementations of moving data use cases one at a time, great for supporting large data sets, and great for increasing the use of data across the whole organization. Because this study focuses on finding possible consolidation opportunities and not choosing technologies, no specific applications or tools are mentioned. Several of the applications available on the market can be used interchangeably if they offer the same functionalities.

5.1 Solution 1: One Common Platform

According to Forrester Consulting (Forrester Consulting 2018: 3), business agility can be improved through platform consolidation. Companies who had successfully consolidated to a single platform reported many benefits. One of these benefits of having a single platform allows companies to focus on company-wide transformation to have more success.

Figure 12 illustrates the structure of the single platform solution. Single colored boxes are directly moved from a previous platform, gradients are consolidated data use cases and gradients with a colored text are consolidations from all platforms. Platform 1 and 2 are both data warehouses analytics platforms and were built to support a customer-oriented data driven-approach. By examining appendices 2 and 3 we can find several data use cases with similar business goals.

Let us start by consolidating the application endpoints from Platform 1 and 2. A common data use case for both platforms is the "End use application" use case. Both platforms provide electricity related data to end user application and could be consolidated. The "Marketing" use case from Platform 1 and "Leads" use case from Platform 2 provide

customer data so that automated marketing campaigns can provide personal marketing. Whether these can be merged into a single use case depends on the functionalities of the marketing tools that is used, but theoretically they can be consolidated. The “Telemarketing” and “Credit control” use cases from Platform 2 do not have directly equivalent use cases in Platform 1 and thus will be kept as their own use cases.

Next let us consolidate the data visualization endpoints. The “Finance” use case from Platform 1 and “Corporate Credit” use case from Platform 2 both have the goal of monitoring financial results. The only difference between these two platforms is what Business Intelligence tool is used to visualize the data, but in this example solution let us consolidate them under Business Intelligence tool 1. There is a use case that is not currently used in Platform 2 and that is the “Customer journey”. There is a use for the same data in Platform 1 and could be consolidated under the “Processes & Customer” in Business Intelligence tool 1. The “Forecasting and Price” use case in Platform 2 has the same business goals as the “Procurement” use case in Platform 1. Because the “Procurement” use case has additional business goals, such as reporting electricity origin, the “Forecasting and Price” use case should be consolidated to it. “Markets” use case from Platform 2 and “Sales” use case from Platform 1 have the same business goal of monitoring operational KPIs. Once again, the only difference between these two platforms is what Business Intelligence tool is used to visualize the data, but in this example solution let us consolidate them under Business Intelligence tool 1. The “Customer Service” use cases from both Platform 1 and 2 can instantly be consolidated, since both have the same data visualization solution, visualization purpose and business goals. In Platform 2, the “GDPR” use case is located under other endpoints, meanwhile in Platform 1 it is in Data visualization endpoint. Both use cases serve the same business goal and should be consolidated into one use case. The rest of the use cases do not have directly equivalents in Platform 1 and will be included as their own use cases.

Lastly between Platform 1 and 2, let us consolidate the other endpoints. The “Platform 3 KPIs” use case will not be needed in this solution because Platform 3 will be consolidated to it. The rest of the use cases from Platform 2 do not have directly equivalents in Platform 1 and cannot be consolidated to anything.

As the last step, let us consolidate Platform 3. Platform 3 is slightly different than Platform 1 and 2 since it is a data lake and was built with a slightly different purpose. The only use cases that could directly be consolidated would be “Marketing” and “Customer information”. The biggest reason to why many of the current use cases in Platform 3 can’t be consolidated is because Platform 3 was built to harmonize KPIs, business logic and definitions and to provide comparable results across all existing brands. Platform 3 is mostly focused on providing company wide brand analytics while Platform 1 and 2 are more focused on their respective brands. When consolidating a data lake with a data warehouse, the future architecture will need either an ELT pipeline or a data lake before the data warehouse. This way it is ensured that if the raw data is needed, it can be fetched from either the data warehouse or the data lake, depending on the technical architecture.

The pros of this solution are reduced costs of production environments, increased companywide analytic capabilities, less user training with fewer software and easier data management across the whole organization. The cons are that the strategic business goal of the platform expands and loses its original purposes, larger systems are often harder to maintain, system failure might affect more systems, and are complicated to build. As mentioned in the second section, larger systems have more components and are often much more complex and harder to maintain due to the increase in technologies. Having one central place for all data makes data management and governance easier, because everything is located under the same system.

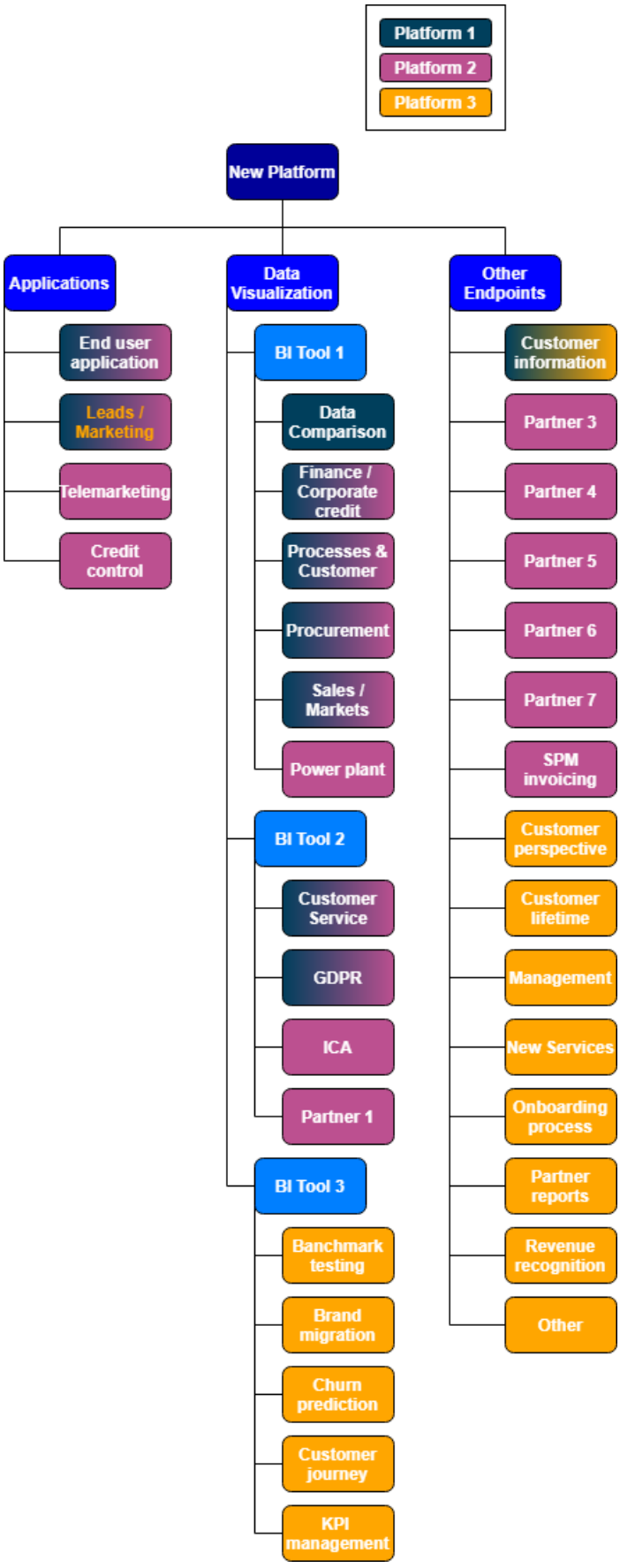

Figure 12. Structure of a single common platform solution

5.2 Solution 2: Consolidated Data Warehouses

Instead of designing a single platform, this solution is a design based on the original platforms' technical solution and business goals. Platforms 1 and 2 are both data warehouse and analytics platforms built to support a customer-oriented data-driven approach. Platform 3 is also built to provide reporting and analysis of customers but has a focus also on providing more unified definitions of KPIs, business logic, definitions, and to provide comparable results across all brands. This solution will consolidate Platforms 1 and 2 and keep Platform 3 as a separate system meant for company wide analytics. With this solution, the original technical architectures can be kept.

The consolidation of Platforms 1 and 2 follows the same idea as in solution 1. The only difference is that "Platform 3 KPIs" data use case is needed to be able to provide the necessary data. Platform 3 will be kept as a separate platform. The pros of this solution are less changes to architecture, scalability, strategic business goals stay the same and easier user management. The cons are data management needed across several platforms, need of external endpoints to Platform 3, and increased time for reporting. As mentioned in section two, less changes to the architecture will mean that the systems keep their benefits, such as scalability and original business goals.

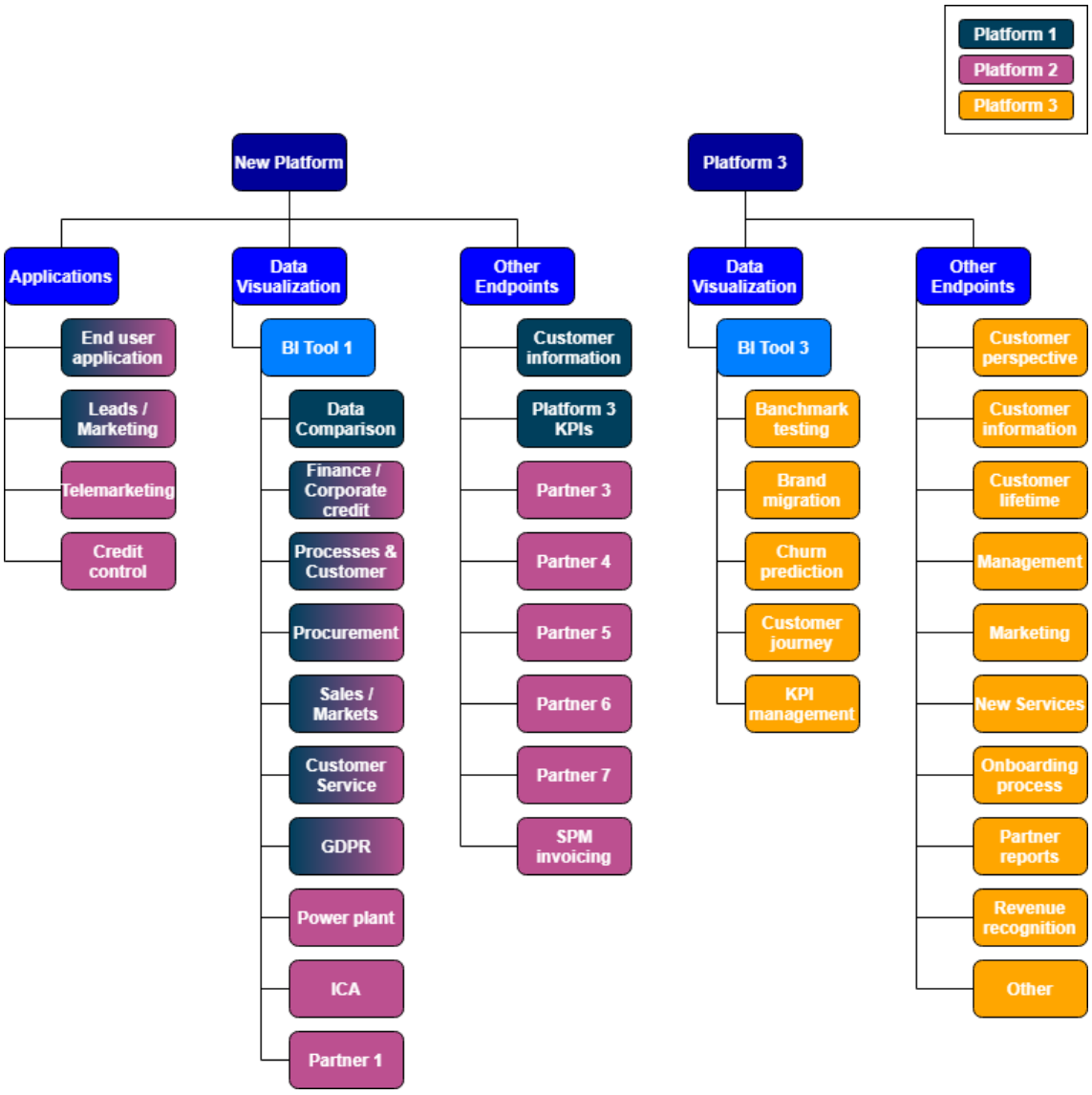

Figure 13. Structure of a consolidated data warehouse solution

5.3 Implementation of a solution

Before implementing the changes to the platforms, one should do a deeper examination on each data use case comparing reports and functionalities of the visualization tool. It is also important to examine each use case thoroughly to better understand if some of the reports should be moved under another data use case to provide a more coherent system or to be able to govern them more efficiently. A cost comparison of the systems would also be beneficial. Using the current costs of each platform and an estimate of work needed to consolidate the platforms would provide an estimated price for the project. It is

important to consider that cost savings might only occur further in the future (Investopedia 2020). This is often the case. Therefore, a business case should be done for say, 10 years, and count the savings gained vs costs for implementation, maintenance and phase out. Another thing to take into account is maintainability of the platforms. Each platform is crucial for daily operations and problems in the current system only affect the platforms own brands. Having a badly designed single platform might provide a worse service level than several functioning platforms. Even with thorough planning and back-ups, as well as enough on-call support to solve issues quickly would lead to increased costs as well. When implementing one of the solutions using Data vault 2.0 methodology, it is recommended to approach it with a gradual delivery approach rather than with a big bang release. This way reports and data use cases can be validated between the old and new platform. It is also a way of providing a proof of concept to ensure that it really is possible to consolidate a certain data use case.

6 CONCLUSION

This study has increased the understanding of the case company's different data platforms and the potential overlaps between them by providing the case company with a data use case mapping. The goal of the study is to find potential overlaps and how a future solution could look like. With the collected data, a future solution can be designed, but more work needs to be done to create the final implementation. The data use cases that were collected on a high level have proven that there are overlaps between the systems and there is a possibility to consolidate them. The study has shown that Platform 1 and 2 are very similar in terms of their functionality and could be consolidated. While the current data would support the decision making and fulfill the objectives of the study, it is lacking relatively much information to be able to plan the real solution.

For qualitative research, it is hard to measure reliability and validity. The reliability of this study is increased with the documented methods and workflow of data collection. The validity of the study is increased with using several different sources for the use cases, such as documentation, in-depth interviews and single question surveys sent by email. By involving representatives from different departments, the validity of the research is also strengthened.

Two different solutions have been designed as a part of the study. The solutions designed in this thesis strive to preserve all the use cases with their own functionalities. The solutions are theoretical and would need more research on a report or data use case level to be sure that a future implementation would be possible. If one would want to implement one of the designed solutions, a deeper research for each data use case would be necessary. This way it can be ensured that all needed functionalities are present in the consolidated solution. A cost comparison of the systems would also be beneficial, in order to get an estimate of the costs for the future solution. When implementing a solution, it is recommended to deliver it with a gradual delivery approach, as it is the go-to method for larger platform changes. An iterative approach enables time to validate the data use cases between the new and old platform. Both solution 1 and 2 are theoretically possible. Solution 2 is the solution I would recommend or iterate more on. Solution 2 retains the system

integrities, helps with user management by having separate platforms for company wide data and brand specific data.

While the study focused on gathering the data on a high level, the results could be improved by collecting the data on a lower level. For certain BI Tools it would be beneficial to gather the use cases by report level. This is because several of the BI Tools house several different types of reports within them. It is hard to create one category that fits all the different reports. This will leave some reports out of the consolidation solutions, but they would most likely be noticed during a deeper study or at the latest when planning the real solution. Similar studies in the future might consider adding the number of users, usage-frequency and query times to the data collection. This would confirm consolidation choices and implementation importance ratings based on the user demands and usage amounts. These consolidation possibility studies can be executed in several ways depending on the level of data needed and what the end result needs to prove. It is still more logical to do a light study to find out if there is even a benefit of consolidating IT systems before starting with planning the real future implementation.

REFERENCES

Cockburn, A., 2000, *Writing Effective Use Cases*, Accessible: https://people.inf.elte.hu/molnarba/Informaciorendszerek_ELTE/Writing_effective_Use_cases_Cockburn.pdf Accessed: 21.10.2020

Data Warehouse, IBM. Accessible: https://www.ibm.com/cloud/learn/data-warehouse Accessed: 15.10.2020a

Data-Driven Documents, D3.js. Accessible: https://d3js.org/ Accessed: 16.10.2020

ETL (Extract, Transform, Load), IBM. Accessible: https://www.ibm.com/cloud/learn/etl Accessed: 29.11.2020b

Forrester Consulting, 2018, *Improve Business Agility Through Platform Consolidation*, Accessible: https://www.servicenow.com/content/dam/servicenow-assets/public/en-us/doc-type/analyst-report/servicenow-forrester-platform-consolidation.pdf Accessed: 9.10.2020

Fortum, Fortum. Accessible: https://www.fortum.fi/ Accessed: 9.10.2020

Hubspot, Hubspot. Accessible: https://www.hubspot.com/Accessed: 16.10.2020

Linstedt, D. & Olschmike, M., 2016, *Building a scalable data warehouse with Data Vault 2.0*, Accessible: Adlibris e-books. Accessed: 19.10.2020

Microsoft Excel, Microsoft. Accessible: https://www.microsoft.com/en-gb/microsoft-365/excel Accessed: 16.10.2020

Plotly, Plotly. Accessible: https://plotly.com/ Accessed: 16.10.2020

Qlik, Qlik. Accessible: https://www.qlik.com/us/ Accessed: 16.10.2020

Tableau, Tableau. Accessible: https://www.tableau.com/ Accessed: 16.10.2020

The what, why & how of the 360-degree customer view, Digital Marketing Institute. Accessible: https://digitalmarketinginstitute.com/blog/the-what-why-and-how-of-360-degree-customer-view Accessed: 29.11.2020

Total Cost of Ownership - TCO, Investopedia. Accessible: https://www.investopedia.com/terms/t/totalcostofownership.asp Accessed: 29.11.2020

What is a data platform?, Splunk. Accessible: https://www.splunk.com/en_us/data-insider/what-is-a-data-platform.html Accessed: 15.10.2020

APPENDIX 1.

Business and IT representatives Interview Questions

The interviews for the data collection were done with 6 persons representing IT and Business. The questions in the interviews were designed to get detailed answers of the use cases. The data use case column has been defined with the strategic business objective in mind and that the end user for the different data use cases is business.

1. How is the data used for this specific data use case?
2. In what format is the data used? Is it used for reporting, ad-hoc reporting, provided to external systems or a combination of these?
    a. If the data is provided in another format than reporting or ad-hoc reporting, in what format is the data provided? CSV-files, DB tables or other text format?
    b. If the data is used for reporting or ad-hoc reporting, what application is used to visualize the data?
3. What is the strategic business goal with this specific data use case?
4. How does this data use case impact the business? Does it increase the efficiency, decrease errors or provide value in some other form?
5. Can you think of any other data use cases that are missing from this list?

APPENDIX 2.

Platform 1 Use Cases

Platform	In Use	Current Data visualization solution	Data Endpoint	Use-Case	Visualization purpose	Strategic business goal	Use case impact
Platform 1	Yes	BI Tool 1	Data Visualization	Finance	Reports	To monitor and calculate the financial results	Increase effectiveness of reporting and daily operations
Platform 1	Yes	BI Tool 1	Data Visualization	Processes & Customer	Reports	To improve customer satisfaction, monitor people performance & process performance	Increase customer satisfaction and people & process performance
Platform 1	Yes	BI Tool 1	Data Visualization	Procurement	Reports	To develop a hedging strategy and report electricity origin	Reduce the risk of losing money
Platform 1	Yes	BI Tool 1	Data Visualization	Sales	Reports	To follow all operational KPIs, such as customer lifetime value, customer sentiment analysis, churn prevention, etc	Increase sales
Platform 1	Yes	BI Tool 1	Data Visualization	Data Comparison	Reports	To compare measurements between Platform and source systems	Ability to validate hourly measurements between systems
Platform 1	Yes	BI Tool 2	Data Visualization	Customer Service	Reports & Ad hoc	To provide ad hoc reports to help with invoice and billing related daily operations	Increase effectiveness of daily operations
Platform 1	Unclear	BI Tool 2	Data Visualization	Electricity Sales Management	Reports & Ad hoc	N/A	N/A
Platform 1	Yes	BI Tool 2	Data Visualization	GDPR	Ad hoc	To be able to provide user GDPR requests	Increase effectiveness of GDPR workorders
Platform 1	Unclear	BI Tool 2	Data Visualization	Insuring & Securing analytics	Reports & Ad hoc	To use insurance contract information	N/A
Platform 1	Unclear	BI Tool 2	Data Visualization	Sales	Reports & Ad hoc	To use sales information	N/A
Platform 1	Yes	Data transfer	Other Endpoint	Customer information	File	To provide a list of customers to be excluded from a call list	Ability to call only correct people and follow laws/regulaitons
Platform 1	Yes	Data transfer	Other Endpoint	Platform 3 KPIs	File	To provide Platform 3 with top level KPIs related to contracts	Ability to provide global reporting
Platform 1	Yes	Data transfer	Applications	Marketing	File	To provide customer and contract data for a Marketing BI-tool	Ability to provide personal marketing
Platform 1	Yes	Data transfer	Applications	End user application	File	To provide electricity related data to end user application	Improve customer experience
Platform 2	Yes	BI Tool 2	Data Visualization	Markets	Reports & Ad hoc	To follow operational KPIs, such as customer flows, churn prevention, etc.	Increase sales, prevent churn

APPENDIX 3.

Platform 2 Use Cases

Platform	In Use	Current Data visualization solution	Data Endpoint	Use-Case	Visualization purpose	Strategic business goal	Use case Impact
Platform 2	Yes	BI Tool 2	Data Visualization	Improved Customer Analysis	Reports & Ad hoc	To provide more uniform customer analysis, define concepts within business	Ability to help with defining concepts
Platform 2	Yes	BI Tool 2	Data Visualization	Customer Service	Reports	To provide reports to help with invoice, billing, price model and partner related daily operations	increase effectivness of daily operations
Platform 2	Yes	BI Tool 2	Data Visualization	Corporate credit	Reports	To provide reports for financial results	Increase effectiveness of reporting and daily operations
Platform 2	Yes	BI Tool 2	Data Visualization	Partner 1	Reports	To provide Partner 1 with KPIs and financial results	Increase effectiveness of reporting data to partners
Platform 2	Expires	BI Tool 2	Data Visualization	Partner 2	Reports	To provide Partner 2 with KPIs and financial reports (will disappear 2021)	Increase effectiveness of reporting data to partners
Platform 2	Unclear	BI Tool 1	Data Visualization	MASSE	Reports	N/A	N/A
Platform 2	Unclear	BI Tool 1	Data Visualization	B2B PoC	Reports	N/A	N/A
Platform 2	Yes	Data transfer	Applications	End user application	DB tables	To provide electricity related data to end user application	Improve customer experience
Platform 2	Yes	Data transfer	Other Endpoint	Service price model invoicing	DB tables	To gather invoicing data from various sources	Increase effectiveness of internal invoicing
Platform 2	Yes	Database	Applications	Credit control	DB tables	To provide data for payment and debt collection related daily operations	Ability for payment and debt collection teams to perform needed checks
Platform 2	No	Database	N/A	Customer journey	DB tables	To imporive customer satisfaction through surveys	Increase customer satisfaction
Platform 2	Yes	Database	Other Endpoint	Partner 3	File	To provide list of customers and their points to Partner 3	Increase customer loyalty
Platform 2	Yes	Data transfer	Other Endpoint	Partner 4	File	To provide customer & contract information to Partner 4	Increase customer loyalty
Platform 2	Yes	Data transfer	Other Endpoint	Bull's eye segmentation	DB tables	To provide Improved Customer Analysis use-case with customer related KPIs	Ability to provide customer analysis in other applications
Platform 2	Yes	Data transfer	Applications	Leads	File	To provide list of customers for automated and ad hoc marketing campaigns	Ability to provide personal marketing
Platform 2	Yes	Data transfer	Other Endpoint	Forecasting and price	DB tables	To develop a hedging strategy and support with pricing	Increase effectiveness of forecasting and pricing analysis
Platform 2	Yes	Data transfer	Other Endpoint	Churn prediction	DB tables	To predict churn	Ability to prevent churn
Platform 2	Yes	Data transfer	Other Endpoint	GDPR	Ad hoc	To be able to provide user GDPR requests	Increase effectiveness of GDPR workorders
Platform 2	Yes	Data transfer	Other Endpoint	Partner 5	File	To provide customer, contract, product data to partner 5	Ability to provide automated marketing activities and analysis of customer responses
Platform 2	Yes	Data transfer	Other Endpoint	Partner 6	File	To provide filtered customer and prospect data to the marketing team	Increase effectiveness of telemarketing
Platform 2	Yes	Data transfer	Other Endpoint	Partner 7	File	To provide consumption data to Partner 7 for analysis	Ability to provide ad hoc marketing activities
Platform 2	Yes	BI Tool 1	Data Visualization	Power plant	Reports	To provide power plant management with KPIs	Ability to perform KPI related analysis
Platform 2	Yes	Data transfer	Applications	Telemarketing	DB tables	To revivece and store telemarketing results for later analysis	Ability to analyze marketing campaign results

APPENDIX 4.

Platform 3 Use Cases

Platform	In Use	Current Data visualization solution	Data Endpoint	Use Case	Visualization purpose	Strategic business goal	Use case impact
Platform 3	Yes	BI Tool 3	Data Visualization	KPI management	Reports	To provide the most important KPIs for all brands	Ability to compare figures, performance tracking and improved insights across business units
Platform 3	Yes	BI Tool 3	Data Visualization	Brand migration tracking	N/A	To track the customers changing brands within the Fortum group	Ability to avoid holdvack activities and unnecessary promotional offers
Platform 3	Yes	BI Tool 3	Data Visualization	Churn prediticition	N/A	To predict churn with ML models	Ability to provide targeted customer service activities
Platform 3	Yes	BI Tool 3	Data Visualization	Customer journey	Reports	To provide reports of customer journey from onboarding to churn	Ability to compare customer journey across brands
Platform 3	Yes	BI Tool 3	Data Visualization	Benchmark testing	Reports	To provide automatic benchmark testing of new processes, marketing campaigns, etc.	Ability to gain instant insights on campaigns and make changes based on insights
Platform 3	Yes	N/A	Other Endpoint	Marketing	Reports & Ad hoc	To provide partners with customer data	Improve customer relationships
Platform 3	Yes	Several	Other Endpoint	Management	Reports	To provide reports related to management	Increase effectiveness of daily operations
Platform 3	Yes	Several	Other Endpoint	Customer information	Reports & Ad hoc	To provide automatic and ad hoc list of customers for telemarketing and sales agencies	Increase sales
Platform 3	Yes	Several	Other Endpoint	Agile customer perspective	N/A	To provide agile customer view that can change as strategies change	Ability to provide desired customer perspective and makes changes
Platform 3	Yes	Several	Other Endpoint	Onboarding process	Process	To be able to add new brand or business unit to Platform 3	Ability to establish a complete picture of customers
Platform 3	Yes	Several	Other Endpoint	Customer lifetime	N/A	To provide customer analysis to evaluate what factors effect e.g. customer loyalty	Increase customer insights and input on strategic offers and future product development
Platform 3	Yes	Several	Other Endpoint	New services	N/A	To provide a platform to develop service strategies	Ability to develop and make strategical changes to services
Platform 3	Yes	BI Tool 4	Other Endpoint	Revenue recognition	N/A	To harmonize definitions and KPIs across brands and business areas	Ability to perform comparisons across whole Fortum group
Platform 3	Yes	Several	Other Endpoint	Partner reports	Reports & Ad hoc	To provide partner reports to large enterprise customers	Improves customer relations with largest customers
Platform 3	Yes	Several	Other Endpoint	Other	Reports & Ad hoc	To provide reports and ad hoc activities to surveys, customer insights, etc.	N/A

APPENDIX 5.

Swedish Summary

Planering av en konsolidering av dataplattformar: fallstudie

Då företag utvidgar sin verksamhet geografiskt, gör företagsförvärv eller på annat sätt utvidgar sin verksamhet tillräckligt, finns det ofta en god anledning att undersöka vilken nytta de kunde dra av IT-konsolidering. Det finns flera fördelar med att konsolidera IT-system. Konsolidering kan sänka de operativa kostnaderna, skapa mera centraliserade och sammanhängande system, förbättra datahantering och mycket mer. I undersökningen utförd av Forrester Consulting (Forrester Consulting 2018: 3) fick man reda på att företag som konsoliderat deras plattformar till en enda plattform, rapporterade kortare tid från idé till marknad för sina nya applikationer, bättre affärsflexibilitet, bättre innovation och ökad säkerhet. Plattformkonsolidering förbättrar också de anställdas produktivitet, vilket ökar medarbetarnas nöjdhet och gör företaget mer attraktivt för svåranställda talanger. Således har företag många goda orsaker att undersöka sina olika IT-system för att hitta potentiella konsolideringsmöjligheter. Fallföretaget för examensarbete är Fortum Oyj. Fortums verksamhet omfattar utveckling, produktion och försäljning av elektricitet- och värmerelaterade tjänster (Fortum 2020).

Företag växer ofta genom åren och implementerar flera olika IT-system. När företag fokuserar på tillväxt blir IT-konsolidering ofta nödvändigt för tillväxt. Varje dag skapas och samlas data från olika system och källor. Denna data samlas ofta till ett centralt data lager och kan lagras på flera olika sätt, såsom i databaser, data warehouses, data lakes etc. Denna data är ofta avsedd för att vidareanvändas, till exempel för maskininlärning, artificiell intelligens och affärsanalys. Därför är data lagret en central del av dataarkitekturen och data plattformen för att kunna använda data effektivt.

Fallföretaget har flera olika dataplattformar inom kund- och försäljningsdomänen. Inom fallföretaget har man velat öka förståelsen om dessa olika dataplattformar och ta reda på potentiella överlappningarna mellan dem på en dataanvändningsnivå. Examensarbetet kommer att ge företaget nödvändiga informationen och en möjlig framtida lösning, som strävar efter att kombinera de bästa elementen från de nuvarande dataplattformarna.

Målet med examensarbetet är att samla in nödvändig information för att kunna planera en konsolidering av dataanvändningsfall mellan de olika dataplattformarna och föreslå en möjlig framtida lösning för dataplattformarna. Examensarbetet tar inte ställning till teknologival. Datainsamlingen kommer att ske genom att läsa dokumentation, samt genom att intervjua representanter för företagsledningen och IT. När all data har samlats in, är målet att planera en möjlig framtida lösning för dataplattformarna. Det är viktigt att notera att målet med examensarbetet inte är att bestämma vilken teknik som skall användas, utan hjälpa att öka förståelsen för dataplattformarna.

För datainsamling användes en kombination av olika kvalitativa forskningsmetoder. De kvalitativa forskningsmetoderna som användes var dokumentationsanalys, fördjupande intervjuer och enkäter skickade via e-post. För att identifiera olika dataanvändningsfall har Cockburns guide ”Writing Effective Use Cases” (Cockburn 2000) använts som grund för insamlingen av användningsfall. Eftersom plattformarna hade mycket dokumentation var första steget för varje plattform att samla en grov översikt av alla dataanvändningsfall. Efter grova datainsamlingen intervjuades representanter för IT och företagsledning. Även efter intervjuerna saknades information om vissa användningsfall. För att samla in de uppgifter, som saknades, skickade man enkäter via e-post. När datainsamlingen var klar för en plattform, startade man datainsamlingen för nästa plattform enligt samma process. Eftersom dataanvändningsfallen för Plattform 3 nyligen hade dokumenterats och validerats, har användningsfallen kopierats och förenats så att de passar in med resten av användningsfallen.

Nuvarande läget för kund- och försäljningsdataplattformarna innehåller tre olika dataplattformar som ursprungligen byggdes för olika varumärken. Företagets olika varumärken har sin egen målsättning och sina egna dataanvändningsfall, vilka företaget vill utnyttja, men majoriteten av användningsfallen bör vara jämförbara. Den största skillnaden mellan de olika plattformarna är skillnaden i deras arkitektur och ursprungliga syfte. Plattform 1 och 2 är typiska molnbaserade data warehouses och analysplattformar. Plattform 3 däremot är en molnbaserad data lake och analysplattform. Plattformarna 1 och 2 kan i teorin enkelt konsolideras eftersom båda är data warehouses, men för att kunna konsolidera data från plattform 3, som är en data lake, måste vissa ändringar göras. Detta kan innebära att data behövs transformeras, ha en data lake före data warehousen eller använda en data warehouse med en ELT-pipeline.

Att planera en konsolidering av plattformar är ofta en lång och komplex process, som har många risker och är ofta svår att planera. Varje lösning har sina för- och nackdelar. Examensarbetet presenterar två olika lösningar. En lösning är baserad på Forrester Consultings resultat (Forrester Consulting 2018) och en annan lösning är baserad på dataplattformarnas ursprungliga verksamhetens målsättningar. Data Vault 2.0-metodik används i designprocessen på grund av de fördelar den erbjuder. Data Vault 2.0-metodik är utmärkt för iterativ implementering av dataanvändningsfall, bra för att stöda stora datamängder och bra för att öka användningen av data i hela organisationen.

Examensarbetet har ökat förståelsen för fallföretagets olika dataplattformar och de potentiella överlappningarna mellan dem genom att förse fallföretaget med en kartläggning av dataanvändningsfall. Med den insamlade informationen kan en framtida lösning planeras. Innan man implementerar förändringar på plattformarna, bör en djupare undersökning av varje dataanvändningsfall göras. En kostnadsuppskattning av den eventuella konsolideringen, vore nyttig. Genom att använda de aktuella kostnaderna för varje plattform och en uppskattning av det arbetet som behövs för att konsolidera plattformarna, skulle man kunna få en uppskattning av projektets kostnader. Det är viktigt att tänka på att kostnadsbesparingar kanske inte sker i den närmaste framtiden, utan senare och att detta ofta är fallet. För själva implementeringen, rekommenderas att leverera den gradvist istället för än ”big bang release”. Då kan rapporter och dataanvändningsfall valideras mellan de gamla och nya plattformarna. Både lösning 1 och 2 är teoretiskt möjliga. Själv skulle jag rekommendera lösning 2 eftersom den behåller system-integriteten och förenklar användarrättighetshantering genom att ha separata plattformar för företagsomfattande data och varumärkesspecifik data.