Architecture Proposal for Constructing a Data Warehouse

IU International College

Project Report	10.08.2022
Project: Business Intelligence	DLBCSEBI02
Tutor	Sebastian Werning
Sandro Schweiss	3206640

Table of Contents

I. List of Abbreviations

Abbreviation	Meaning
SRD	Software Requirement Document
DWH	Data Warehouse
SDLC	Software development life cycle
ETL	Extract, Transform, Load
ERP	Enterprise Resource Planning
FAS	Financial Accounting System
FB	Facebook
SCM	Supply Chain Management
PPS	Production Planning System
PPC	Pay-Per-Click
DBMS	Database Management System
RDBMS	Relational Database Management System
C-DWH	Core Data Data Warehouse
ODS	Operational Data Store
KPI	Key Performance Indicators
ETL	Extracting, Transforming, and Loading
FTP	File Transfer Protocol

II. List of Figures

Figure Description **Page**

III. List of Tables

4. Introduction

4.1 Project Overview

Data is an invaluable asset for every organization. It can be used to make better decisions, improve customer experience, and increase revenue. The use of data in the industry has been common for many years. The first information systems, for example, could be traced back to as early as the 1960s. Over these decades, data has become a crucial and exponentially growing aspect of the business world. Through this increasing importance and exponential growth, many companies face the challenge of correctly utilizing their data. Most organizations do not have the proper handling of data in mind when they first establish their business. This leads to an improper implementation of the architecture for their data, which presents a whole host of problems like data credibility and the inability to transform data into information. This is where the data warehouse becomes crucial.

A DWH is a repository of data from one or more sources to support business intelligence and analytics. Data warehouses can be implemented in different ways, but they all share the same goal: to provide a single point of the truth for all stakeholders (Dedić & Stanier, 2016). The following project report concerns itself with the implementation of a data warehouse in the organization “Balustrade.” Balustrade is a company that manufactures and sells staircases. The goal of my work was to implement a data warehouse in the organization. This meant that operational data could be adequately transformed, integrated, and stored in the data warehouse, which is then distributed to the individual departments using the data for analysis and reporting.

4.2 Project Approach

A data warehouse project can be challenging and complex. Many different structures, options, and procedures must be taken into consideration. The goal is to ensure the final data warehouse solution will meet the business needs and the company's strategic objectives. The first step, therefore, was to review related literature. In this approach, relevant information from multiple literature sources was extracted. The current state of the art was explored, and possible solutions were outlined. In addition to literature, different real data warehouse use cases were inspected to identify strengths, weaknesses, and possible implementations.

4.3 Project Methodology

After evaluating the material and use cases, the actual project methodology was chosen. As previously elaborated, many different components must be considered when constructing a DWH. But not only is it vital to determine what to build but also how. Many project methodologies can be used to build a data warehouse. Changing the method can alter the outcome of the end product significantly. Usually, software products are developed with the classical SDLC, which starts with a set of requirements and then proceeds with the design and development stages. The development of DWH, however, operates under a different lifecycle called the CLDS, which is the reverse of the

SDLC. The CLDS methodology starts with a basic set of requirements and data, working backward again towards the requirements. Once the requirements are understood, adjustments are made, and the cycle starts over again (Inmon, 2005). This data-driven development life cycle ensures that the organization's data is the DWH implementation's primary concern.

4.4 Project Planning and Organization

After determining the project methodology, it was necessary to organize and gather information about the company and plan accordingly. The first steps of a DWH lay the foundation for all the succeeding phases. Therefore, understanding the business process, the risks, impact, and the organization as a whole is necessary for any DWH project to be successful. The data collected included various information to support the project methodology, including company history, company structure, business processes, risk assessment, and key stakeholders. The primary data sources were interviews, secondary research, and stakeholder surveys. The result of the planning process was a summary of the current state of the business, which is a pivotal foundation for the DWH implementation.

4.5 Project Report Structure

Section 5 presents the main body of the project report, containing the steps of the preparation and implementation of the DWH. The beginning of the main body is Section 5.1, which identifies the current situation and lays the project's foundation. Section 5.1.1 analyses the company's structure, history, and future. Followed by the identification, description, and classification of stakeholders in section 5.1.2. Section 5.1.3 contains the requirements engineering process of the project. In this segment, the requirements of the different stakeholders are identified, and project goals are laid out. Section 5.1.4 identifies the source systems and describes their details. The overview ends with segment 5.1.5, where the different architecture variants are discussed and evaluated. Section 5.2 contains the implementation, starting with understanding and identifying the business process in segment 5.2.1. The business process is the foundation for the fact table in the following section 5.2.2, where the "grain" is identified, which entails one row in the fact table. Subsequently, section 5.2.3 selects the data model, and corresponding dimensions to the fact table are chosen. The main body ends with segment 5.3, which describes the ETL process. The ETL process is the actual conversion from operational data to management relevant information. Section 5.3.1 defines KPIs, and an exemplary KPI is chosen to demonstrate the ETL process. Segment 5.3.2 then describes the actual ETL process in detail. The final part of the project report is section 4, which contains a conclusion discussing the project results and achievements.

5. Architecture Proposal for Constructing a Data Warehouse (DWH)

5.1 Overview

The overview section of this project report describes the overall situation of the DWH implementation and lays the groundwork for the project. The overview commences with the company analysis. This section entails the state of the business and an analysis of the company's structure as well as history. This stage is the first step to getting a rough idea of the project and narrowing down possible architectures and implementations. Initially, all the stakeholders related to the project were identified and subsequently categorized in terms of their importance, namely critical, major and minor. After this phase concluded, the different perspectives of the stakeholders were established, and the requirements were gathered for the implementation. The requirements came from the business owner and managers in the form of the business perspective and the departments which make up the user perspective. After all the requirements had been gathered and shaped into detailed specifications, the relevant source systems were identified and analyzed. The systems were categorized into different departments. Further information like the DMBS and their functionality inside the organization was denoted. The section concludes with the choice of an architecture variant. The most common architecture variants are elaborated, and their respective pros and cons for this specific business case were evaluated. The section that follows entails the implementation steps.

5.1.1 Company Analysis

The following project report concerns itself with the implementation of a data warehouse in the organization "Balustrade." Balustrade is a company that manufactures and sells staircases. It can be inferred from this that the company operates in the construction sector as a producer and retailer. The organization currently has only one establishment, which is located in Austria, in the city of Eisenstadt. The business was founded in the year 2017. During this timeframe, the organization established itself quite rapidly in the niche of staircase production, which allowed it substantial growth. Presently, the firm has 40 employees, which according to the OECD, classifies the organization as a small enterprise (theOECD, 2022). Through the company's previous successes and rapid growth, the organization plans to expand into other areas of Austria to grow its customer base. These factors needed to be accounted for during the development of the DWH.

5.1.2 Stakeholders

When implemented, a DWH becomes the organization's central repository of enterprise data. Consequently, the DWH turns into the focal point of all data-driven decisions for all stakeholders. A DWH can be used by many different stakeholders, which have their own needs and requirements. Therefore, it is crucial that all stakeholders and their requirements are identified and adequately represented during the development. According to studies, 40% of DWH projects never develop, and 85% fail to meet business objectives (Josef Schiefer, Beate List, & Robert M. Bruckner, 2002). There are many ways to categorize all the different stakeholders and determine their importance. According to Wiegers, the first two subsets of a stakeholder in software projects consist of "customers" and "non-customers." A customer is an individual or firm that benefits directly or indirectly from the product. The "non-customer" subset consists of, for example, suppliers or contractors, while the customer subset can be further broken down into other subsets: The "direct and indirect user" and the "other customer" subset. The "other customer" subset contains, for example, the acquirer and sponsor of the project. The "user" subset, also called "end users," are the individuals who will actually use the product either directly or indirectly. Direct users operate the product hands-on, while indirect users might only receive outputs by directly interacting with the system. The users can be further classified into favored, disfavored, ignored, and other. These classifications determine objectives like priority, security, and importance for project implementation (Wiegers & Beatty, 2013). The stakeholders can then be further classified into their significance. These classifications are critical, major, and minor.

In an organization, the employees always need decisions supported by data for their daily tasks in the form of a report or other media. They are either the ones procuring the data or, with management, the ones receiving them, which makes them the direct and indirect users of the DWH. The procuring of data without a DWH is often a resource-intensive procedure that repeats itself over and over again. Hence, in a DWH implementation, it is generally the case that the different departments have requested the need for decision support. Therefore, the employees are the key stakeholders of the project, as they are the ones who have identified the requirements and are the primary users of the DWH. They must be actively involved in developing the DWH and are with the project sponsor critical stakeholders. Not meeting the requirements of the critical stakeholders would result in project failure. Since the DWH concerns itself almost exclusively with the internal workings of an organization, the external stakeholders, i.e., non-customers like the supplier, government, and legal staff, in this case, all fall under the minor stakeholder category. They have only a negligible influence on the development of the DWH and are only marginally affected by the project's outcome. A major stakeholder is a stakeholder who has a significant impact on the DWH implementation and is considerably affected by the project's outcome. Through the organization's size, no parties, like shareholders, could be identified as major stakeholders. The stakeholders were either actively involved in the development, i.e., internal stakeholders, which are simultaneously the customers, or external stakeholders, which have no real impact on the project and are only marginally affected by it. To summarize all previously mentioned information, the following table gives a quick overview of the identified stakeholders and their importance for the project.

Critical Stakeholders	Major Stakeholders	Minor Stakeholders
Company owner/acquirer		Suppliers
General Management		Legal staff
Marketing Department		Government
Sales Department		Clients
Purchase Department		
IT Department		
Production		
Fiance Department		
Project Team		

Table 1.: Overview table of project stakeholders

5.1.3 Requirements

Subsequently, now that the stakeholders and their corresponding importance had been identified, it was necessary to gather the requirements from the critical stakeholders. The requirements enable stakeholders to communicate their expectations, goals, and purpose. In addition, they directly affect the technical and design aspect of the DWH system. The requirements determine, for example, what data must be accessible, how it is transformed, stored, aggregated, and how it is further used.

The stakeholders often express their needs in the form of general expectations of the DWH to improve their business. This formulation of needs encompasses the goals, expectations, and scope of the project, also described as “business perspective.” The business perspective is the foundation of the DWH requirements and comes from the business owner, manager, and sponsors of the project (Karl E. Wiegers, 1999). The formulation of business needs, however, is vague and does not describe any specifications of the system in detail. Therefore, the business perspective is further refined into the user perspective. The user perspective describes a task that the individual users must accomplish with the DWH. The collection of this data comes from the actual users of the system and must align with the business requirements. The last step is the requirements from an implementation perspective, which concerns itself with the DWH developer team's point of view. The higher-level and lower-level specifications are then considered and then shaped into a very detailed level for the actual DWH implementation (Josef Schiefer et al., 2002).

The business requirements for this project were gathered from the business owner and management. These requirements included the organization's overall goals, objectives, needs, vision, features, constraints, priorities, scope, and overall primary benefits of the DWH. The DWH needed to provide a high degree of flexibility, scalability, and ease of use. This meant that it should be able to grow as more information is stored and processed in it, should the company expand, while still being practical. The user requirements were gathered from all the different departments and concerned themselves with their individual tasks. The business owner and general management department primarily require highly aggregated data that helps make decisions and inspect positive or negative business trends. The data only needed to contain the most crucial KPIs to give a birds-eye view of the organization. The individual departments required more granular information for analysis. Here it was essential to converge this data in a central location for easy access and analysis without complex processing. The data was then to be used to support different types of queries or ad-hoc requests, like tracking marketing campaigns' performance and viewing manufacturing progress in real-time.

5.1.4 Source Systems

With all the necessary stakeholder requirements and expectations gathered, the source systems were identified. The source systems are the operating systems where the organization's data is created initially and then integrated into the DWH via the ETL process. The source systems can include internal company data and externally procured data. To ensure that the DWH represents "a single point of truth," it is crucial that all data sources inside an organization are identified. The following table describes the source systems gathered for the DWH implementation:

Department	Source Systems	Application Name	DBMS
Sales Department	ERP	Orlando	MySQL
Purchase Department	ERP/Material Prices	Orlando	MySQL
Production	ERP	Orlando	MySQL
Fiance Department	FAS	SAP	MySQL
Marketing Department	FB	FB Business Manager	Excel

Table 2.: Overview table of source system

The main source system of the organization is the ERP system called Orlando. This system is used in the sales department to manage customer information and contact details, in the purchasing department to handle supplier information and their material prices, and in production to administrate the manufacturing processes, like inventory and progress of product assembly. This ERP combines the functionalities of a PPS and SCM into one application, which significantly reduces the implementation time of the DWH. The finance and marketing department, however, use independent applications. The finance department uses a FAS, which records the financial transactions of a business. This is an important tool for managing the business and making impactful decisions about future actions. The marketing department uses the FB business manager to operate their Social Media PPC campaigns to drive traffic to their business. The data from this software can only be exported in the form of excel sheets.

It is important to note that no NoSQL or cloud databases are used in the organization. All databases are RDBMS, which also simplifies the implementation process. In addition, the finance department uses the same DBMS technology as the ERP system. The data from the marketing department must be exported at regular intervals from the FB business manager but can easily be integrated due to the simplicity of the spreadsheet format.

5.1.5 Architecture Variants and Components

The last step before the implementation was designing a DWH architecture. In order to develop an architecture for a Data Warehousing project, many considerations need to be taken. This section aims to outline the different possible architectures for the data warehouse project and explain their advantages and disadvantages. In addition, the various components making up a DWH are presented, and their pros and cons are evaluated. The section concludes with a selection of a DWH architecture and its components with respect to the business case and stakeholder requirements.

There are a vast number of architecture variants that can be used to construct DWH. The following list describes the basic architectural variants:

- Independent data marts
- Data marts with coordinated data models
- C-DWH without data marts
- Multiple C-DWH
- C-DWH and dependent data marts
- DWH architecture mix

The "independent data marts" architecture is the employment of independent DWHs or also called data marts, for the different departments. In this solution, a central database is not required, which reduces the complexity and makes it more manageable. This results in the faster achievement of usable results. The drawback to this architecture is that it's more challenging to implement a company-wide DWH later on due to the isolation of the data marts. The "Data marts with coordinated data models" is practically the same as the previous architecture. The difference is that the data marts coordinate data with each other, which mitigates the downside of the "independent data marts architecture," namely the implementation of a company-wide DWH. The "C-DWH without data marts," as the name suggests, is the implementation of a company-wide DWH without additional data marts. This solution again achieves usable results in a shorter time frame and is more suitable for small BI solutions. The architecture, however, has increased performance and administration issues the larger the organization evolves. The "Multiple C-DWH" architecture is the employment of multiple company-wide data warehouses. This is beneficial in business conditions where an organization has vastly different products and market structures. This is mainly implemented in large organizations. The "C-DWH and dependent data marts" architecture is the implementation of a C-DWH, which supplies data to the subsequent data marts for different departments. The upside is that only small extracts of data are stored in the data marts, resulting in faster response times and a smaller storage volume. The last architecture is the "DWH architecture mix," which consists of C-

DWHs, dependent and independent data marts, and direct data access. The benefit of this variant is that it can fit to the specific needs of a business and that the architecture evolves with the organization (“IU Learn,” 2022).

The architecture was selected after carefully considering the advantages and disadvantages of the architecture variants in combination with the previously established requirements. The reasoning behind this selection is as follows. As previously elaborated, the requirements specified the need for a high degree of flexibility, scalability, and ease of use. This would result in friction with the "Independent data marts," "Data marts with coordinated data models," and "C-DWH without data marts architecture" variants. Both data mart architecture variants would fulfill the requirement for ease of use and result in a fast and manageable implementation but fall short in scalability. Implementing one of those architectures would be short-sighted considering the company's rapid expansion. This would only result in more issues down the line. The "C-DWH without data marts architecture" faces a similar problem since the larger the company grows, the increasingly complex and resource-intensive it becomes to manage this solution. This variant would be a marginally better fit since the DWH can be further expanded, but considering the need for fast queries, this architecture also falls short. The "Multiple C-DWH" is also not a good fit for this case since the organization is smaller and focuses on only one market structure.

In conclusion, this leaves only two variants, the "C-DWH and dependent data marts," also known as "hub-spoke" architecture and "DWH architecture mix." Deciding between these two architectures depends on the case if a specialized solution, i.e., the "DWH architecture mix," is necessary. Utilizing the architecture mix can ensure optimal support for the organization but at the cost of added complexity in the structure, which would interfere with the requirements of ease of use and maintenance.

Ultimately, it was decided on the "C-DWH and dependant data marts" architecture. The substantiation for this architecture variant is the fact that it was not necessary to implement the "DWH architecture mix." There are various components, which can be an addition to the basic DWH, like the ODS, but were not required for the organization's operations. The "hub-spoke" architecture was sufficient to fulfill all the previously specified requirements. The C-DWH allows storing large amounts of granular data, which can then be supplied to the different data marts of the various departments. This ensures no data loss and allows fast, efficient, and easy access to data, which the departments require. The architecture can then be extended with front-end components like the MIS, which is valuable for management to gain an overview of business processes, and EIS, which provides an overview of the internal and external KPIs of the organization. It also made more sense to start with a more basic architecture and then let the DWH naturally evolve with the business.

5.2 Implementation of the Data Warehouse

The previous section established the project's foundation, providing all the necessary information to start the implementation. The following section encompasses the exact steps which were required to construct the DWH.

5.2.1 Identifying the Business Process

A business process is a natural chain of activities sequentially performed to achieve the business objective of providing value to the customer. An organization has many different business processes occurring at any given time, which can also be further categorized into different types. One way of categorization is operational processes, management processes, and supporting processes. Operational processes are the core business activities that create the primary value stream. Management processes oversee the operational processes. Supporting processes support the core operational processes. All these processes are supported by some sort of data; hence the DWH implementation (Weske, 2012). The business process serves as a starting point for the implementation of the DWH. Therefore, it is crucial to determine the business process, providing the most value. It is essential to focus on the business process and not business departments to ensure consistent delivery across the organization. When determining a business process, it was crucial to focus on the process with the highest impact and lowest risk.

Impact and risks are meaningful indicators when evaluating a business process. The impact, in this case, refers to the importance a specific business indicator has on the organization. While the risk associated with the process are, e.g., data availability, difficulty, data quality problems, or complexity of business logic. When determining a business process as a starting point, it is critical for them to be highly impactful, to have a high return on investment, and to be low risk, so not choosing an area that's difficult to start with.

For these reasons, in the case of the company Balustrade, the business process with the highest impact and the lowest risk was deemed to be to product lifecycle process, starting with manufacturing to selling. The production of staircases is the company's primary driver of revenue and focus point; all their daily activities revolve around this process. Making this chain of activities the process with the highest impact on the organization. As a result of being the most critical procedure, the process also carries a relatively low risk in terms of data availability and difficulty. The risk of business logic complexity and data quality problems are unavoidable in this case and must be sufficiently dealt with during implementation (Kimball & Caserta, 2009).

5.2.2 Identify the grain

Identifying the grain is the most essential step in the whole DWH design process since this defines the level of data detail available for the part of the dimensional model. The fact table is the core of the dimensional model and contains all the business data measurements. If the "grain" is identified, it is showcased what one row of the fact table represents. It is crucial to select the most granular level so that data can not be subdivided anymore. The reason for this is that it can not be predicted how the users of the DWH will query the data. When data is integrated in the most granular way, it can be rolled up in every possible way, but if the granularity is too small, the data can no longer be accessed in greater detail (Kimball & Caserta, 2009). In the case of the organization Balustrade, the grain identifies what one item in the sales row represents, and it states the following:

“One row represents a product bought by a customer from an employee on a given day in a store.”

5.2.3 Choose Model and Dimensions

After the grain has been identified, the dimensions necessary for the fact table must be selected. When a fact table contains the measurements, the dimension tables contain the descriptive elements. They contain detailed information about what is stored in the fact table. However, the storage model must be determined before the actual selection of the dimensions. The two storage models are star schema and snowflake schema. The snowflake schema fully normalizes dimension tables and avoids data redundancy, while the star schema stores redundant data in the dimension tables. The result is that the star schema leads to simpler queries because the fact table is joined to only one level of dimension tables, while the snowflake schema has more complex queries since more joins are needed (Kimball & Caserta, 2009). For Balustrade, the star schema was chosen to emphasize simpler and quicker queries. The requirement for more disk space is negligible since the organization does not store that much data for this to be an issue. However, extra care must be taken during maintenance to avoid data integrity issues. The dimensions were then inferred from the grain, which led to the following dimensions: Products, Employee, Store, Date, and Customer. These dimensions contain all the necessary information of the business process, called attributes. The following graphic shows the physical data model of the DWH:

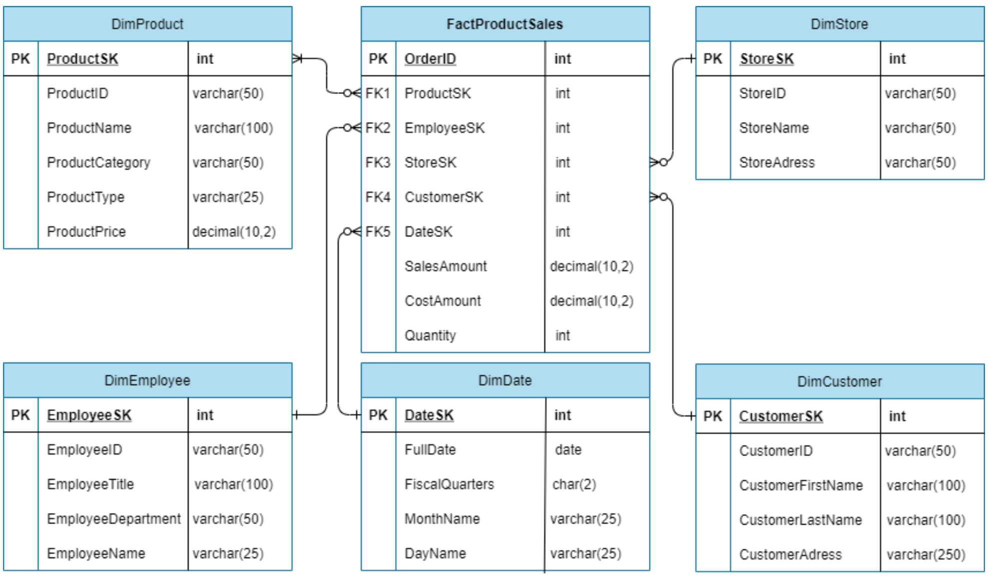

Fig. 1.: Physical data model of the DWH

5.3 ETL Process

The following section concerns itself with the arguable most challenging part of the DWH implementation, the ETL process. This process is also called targeted conversion and consists of three steps, extract, transform and load. While the previous section contains the procedures to set up the DWH infrastructure, the ETL process is the actual conversion from operational data to management-relevant information. The subsequent paragraphs are about the selection of one exemplary KPI and how the operational data is transformed into data interpretable by management via the ETL process.

5.3.1 KPI

A key performance indicator is a quantitative assessment of the performance of a specific strategic goal. They are an integral part of every successful organization since they represent milestones to achieve, provide insights for better decision making, and help every business department to move forward strategically. High-quality KPIs are the end result of a proper DWH implementation. (Roubtsova & Michell, 2014). As example, for the organization Balustrade, the “Gross Profit Margin” KPI is chosen to demonstrate the required ETL process.

5.3.2 ETL process

As previously elaborated, the ETL process consists of three distinct phases, beginning with the extraction step. In this step, the data are extracted from the heterogenous sources, established in section 5.1.3, into the staging area. This is done to avoid performance degradation in the source systems and to roll back in case of data corruption. In this step, the data was also transformed into a suitable format and validated (Kimball & Caserta, 2009). For the gross profit margin KPI it is necessary to collect all the data which constitute an order from the source systems. The ERP system contains all the necessary data: product, customer, employee, and store information. The data is collected from Orlando using an ERP adapter, which enables better metadata navigation. A crucial step that followed the extraction process was the creation of a logical data map. In this step, information about the original source systems was documented, namely, the source database, table name, and column name, as well as possible changes to the data during the extraction process. In addition, the target table name, column name, and type were also inferred from the previously established physical data model in section 5.2.3. The logical data map is a table that specifies the original state and the target state of the data and the necessary transformation process. The transformation phase is the next step in the ETL process, which happens in the staging area. The transformation process consists of four sub-processes: filtering, harmonization, aggregation, and enrichment. Filtering and harmonization are responsible for cleaning and preparing the data. The aggregation process compresses the data to the desired granularity, and the enrichment process summarizes frequently queried information. During the filtering process, the data is corrected for defects

(Kimball & Caserta, 2009). Internal format control and special character recognition were implemented to cleanse first-class defects. These algorithms automatically detect errors like incorrect formats and missing data values, which are automatically corrected. For the gross profit margin, this step included revising and cleaning up orders with missing information and making plausibility checks by comparing balance sheet and inventory information. During the harmonization process, the data was normalized to ensure that the records use the same keys, codes, synonyms, and homonyms, as well as the same business terms. Due to the organization, only shipping in Austria semantic harmonization, like conversion into a uniform currency, was not necessary but must be considered in the future. The harmonization was done with a mapping table, where new artificial primary keys were created, as shown in section 5.2.3. The required granularity of the data was already established in section 5.2.2. Meaning, that the data was aggregated to a daily time frame. This allows for a slight aggregation while still being able to inspect the individual orders. This data is then written to the DWH in the loading phase. From the DWH, the data can then be further aggregated and enriched, and loaded into the individual data marts for the departments. For the gross profit margin, the data mart builds the foundation for the downstream analytical systems like the MIS and ad-hoc analysis systems. The management department can view sales in real-time, as well as under different aspects in the form of OLAP and generate reports of the gross profit margin KPI, e.g., which store produces the most sales, what is the best-selling product, etc. The EIS can be used for upper management to view gross profit margin in combination with other external information that might be relevant to leading the company.

6. Conclusion

The following chapter concludes the project report; here, the project results are summarized and discussed. The DWH has been effectively implemented and has become a vital part of the organization. The development team was able to meet the objectives that were established in the requirements. The business has seen a substantial increase in revenue since the implementation, which is attributed to the improved customer service and analysis capabilities offered by the data warehouse. The company was also able to reduce costs, and the employees are able to make better decisions that align with their objectives due to utilizing the KPIs. The DWH implementation set the foundation for a potential natural evolving architecture. Improvements can be made if the organization decides to expand its services and the DWH architecture needs to be revised to fit the new requirements. An example of this would be the implementation of an ODS if the organization grows to a considerable size and is in need of operational reporting due to an increased number of projects. The implementation of the DWH took longer than expected but was still able to meet project deadlines. In addition, the project resulted being more economical than anticipated. In conclusion, the overall project has been a success.

VII. Bibliography

Dedić, N., & Stanier, C. (2016). An Evaluation of the Challenges of Multilingualism in Data Warehouse Development. In S. Hammoudi, L. Maciaszek, M. M. Missikoff, O. Camp, & J. Cordeiro (Eds.), *Iceis 2016: Proceedings of the 18th International Conference on Enterprise Information Systems : Rome, Italy, April 25-28, 2016* (pp. 196–206). Setúbal, Portugal: SCITEPRESS - Science and Technology Publications Lda. https://doi.org/10.5220/0005858401960206

Inmon, W. H. (2005). *Building the data warehouse* (Fourth edition). *Wiley technology publishing Timely, practical, reliable*. Indianapolis, Indiana: Wiley. Retrieved from https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=239420

IU Learn (2022, July 6). Retrieved from https://learn.iu.org/courses/DLBCSEBI01/books/DLBCSEBI01/version/newest/CYCLE_e6ce70baa1fefd16fbf8027ab8fe98bc

Josef Schiefer, Beate List, & Robert M. Bruckner (2002). *A Holistic Approach For Managing Requirements Of Data Warehouse Systems*. Retrieved from https://www.researchgate.net/publication/2540188_A_Holistic_Approach_For_Managing_Requirements_Of_Data_Warehouse_Systems

Karl E. Wiegers (1999). First Things First: Prioritizing Requirements. Retrieved from https://www.processimpact.com/articles/prioritizing.pdf

Kimball, R., & Caserta, J. (2009). *The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data* ([Nachdr.]). *[Kimball's data warehouse toolkit classics]*. Indianapolis, Ind.: Wiley.

Roubtsova, E., & Michell, V. (2014). Kpis and Their Properties Defined with the EXTREME Method. In B. Shishkov (Ed.), *Springer eBook Collection Computer Science: Vol. 173. Business Modeling and Software Design: Third International Symposium, BMSD 2013, Noordwijkerhout, The Netherlands, July 8-10, 2013, Revised Selected Papers* (Vol. 173, pp. 128–149). Cham: Springer. https://doi.org/10.1007/978-3-319-06671-4_7

TheOECD (2022, February 26). Entrepreneurship - Enterprises by business size - OECD Data. Retrieved from https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm

Weske, M. (2012). *Business Process Management: Concepts, Languages, Architectures* (2. 2nd ed. 2012). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1493063

Wiegers, K., & Beatty, J. (2013). *Software requirements* (3. ed. [fully updated and expanded]). *Best practices*. Redmond, Wash.: Microsoft Press.