



UNIVERSITETI I EVROPËS JUGLINDORE  
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА  
SOUTH EAST EUROPEAN UNIVERSITY

POSTGRADUATE STUDIES-SECOND CYCLE

THESIS:

DATA WAREHOUSE DEVELOPMENT

CANDIDATE:

ARTAN VELIU

MENTOR:

PROF. VISAR SHEHU

TETOVO, FEBRUARY 2022

## **ABSTRACT**

The way of reporting and sharing information was always linked with databases from the moment that they started to exist. With the time the technology evolved and ways of doing business changed also, users expect to have information immediate, secured and efficient. However, with the sheer number of databases in a company, getting the data in the best possible way was never so easy due of the coordination between different databases that has to be done. Therefore, Neptun as a big company needs a single place to storage and gather data in which the users can easily use and based on that make decisions which will help the company to make better decisions and grow market share in Kosovo. One single database is much better than many places with different data and this will be done by implementing the data warehouse.

Business Intelligence (BI) Business intelligence is a collection of activities to understand business situations by performing various types of analysis on the company data as well as on external data from third parties to help make strategic, tactical, and operational business decisions and take necessary actions for improving business performance (Building a data warehouse, Vincent Rinardi, p:12). BI includes gathering, managing, analysing the data about suppliers, sales and other operations of the company in order to help the management improve the decisions they take for strategic reasons.

A data warehouse is a system that retrieves and consolidates data periodically from the source systems into a dimensional or normalized data store (Building a data warehouse, Vincent Rinardi, p:1). It usually keeps years of history and is queried for business intelligence or other analytical activities. It is typically updated in batches, not every time a transaction happens in the source system (Building a data warehouse, Vincent Rinardi, p:1).

In this thesis, I will describe data warehouse techniques and designs also challenges about cleaning and transforming the data and other challenges that we will face during this implementation. This discussion will also include topics about how data from different databases will integrate in one data warehouse. A data warehouse prototype will also be developed to show how databases undergo Extract, Transform and Load (ETL) process and those will be loaded in a database stare schema to make the

reporting easier. Therefore, this thesis seeks to develop BI & DW system to support decision makers in Neptun to make better decisions using historical that we already have in database.

*Keywords: Business Intelligence (BI), Data Warehouse (DW), Extract, Transform, Load (ETL).*

## ABSTRAKTI

Mënyra e raportimit dhe shkëmbimit të informacionit ishte gjithnjë e lidhur me bazat e të dhënave që nga momenti kur ato filluan të ekzistojnë. Me kohën teknologjia evoluoi dhe mënyrat e të bërit biznes ndryshuan gjithashtu, përdoruesit presin të kenë informacion të menjëhershëm, të siguar dhe në mënyre efikas. Sidoqoftë, me një numër të madh të bazave të të dhënave në një kompani, marrja e të dhënave në mënyrën më të mirë të mundshme nuk ishte kurrë aq e lehtë për shkak të koordinimit midis bazave të të dhënave të ndryshme që duhet të bëhen. Prandaj, Neptun si një kompani e madhe ka nevojë për një vend të vetëm për të ruajtur dhe mbledhur të dhëna në të cilat përdoruesit mund të përdorin me lehtësi dhe bazuar në atë të marrin vendime të cilat do t'i ndihmojnë kompanisë të marrë vendime më të mira dhe të rrisë shtrirjen në tregun në Kosovë. Një bazë e të dhënave e vetme është shumë më e mirë se shumë vende me të dhëna të ndryshme dhe kjo do të bëhet duke implementuar depon e të dhënave apo data warehouse.

Biznes Intelejjenca (BI) është një koleksion i aktiviteteve për të kuptuar situatat e biznesit duke kryer lloje të ndryshme të analizave në të dhënat e kompanisë, si dhe në të dhënat e jashtme nga palët e treta për të ndihmuar në marrjen e vendimeve strategjike, taktike dhe operative të biznesit dhe ndërmarrjen e veprimeve të nevojshme për përmirësimin e performancës së biznesit (Ndërtimi i një data warehouse, Vincent Rinardi, f: 12). BI përfshin mbledhjen, menaxhimin, analizimin e të dhënave në lidhje me furnitorët, shitjet dhe operacionet e tjera të kompanisë në mënyrë që të ndihmojnë menaxhmentin të përmirësojë vendimet që marrin për arsye strategjike.

Një data warehouse është një sistem që tërheq dhe konsolidon të dhënat në mënyrë periodike nga sistemet e burimit në një depo dimensionale ose të normalizuar të të dhënave. Zakonisht përmban në vete të dhëna për shume vite dhe kërkohet për biznes intelligence ose aktivitete të tjera analitike. (Ndërtimi i një data warehouse, Vincent Rinardi, f: 1).

Në këtë tezë, unë do të përshkruaj teknikat dhe mënyrat e planifikimit të data warehouse gjithashtu sfidat për pastrimin dhe transformimin e të dhënave dhe sfidave të tjera me të cilat do të përballemi gjatë këtij implementimi. Ky diskutim do të

përfshijë gjithashtu tema se si të dhënat nga bazat e të dhënave të ndryshme do të integrohen në një data warehouse. Do të zhvillohet gjithashtu një prototip i data warehouse për të treguar se si bazat e të dhënave i nënshtrohen procesit të Ekstraktimit, Transformimit dhe Ngarkesës (ETL) dhe ato do të ngarkohen në skemat e shikimit të të dhënave për të bërë më të lehtë raportimin. Prandaj, kjo tezë kërkon të zhvillojë sistemin BI & DW për të mbështetur vendimmarrësit në Neptun për të marrë vendime më të mira duke përdorur historin e të dhënave që ne kemi tashmë në bazën e të dhënave.

*Fjalë kyçe: Inteligjencë e Biznesit (BI), Magazina e të Dhënave (DW), Ekstrakt, Transformim, ngarkesë (ETL).*

## TABLE OF CONTENTS

1.0	RESEARCH FIELD .....	8
1.2	AIMS OF THE RESEARCH .....	9
1.2.1	General Objective .....	9
1.3	HYPOTHESIS.....	10
1.4	RESEARCH METHODOLOGY .....	10
1.5	IMPORTANCE OF THE THESIS .....	11
2.0	LITERATURE REVIEW.....	12
2.1	Definition Data Warehouse & Business Intelligence.....	12
2.2	Data Warehouse.....	13
2.3	Data modelling techniques.....	15
2.4	Data Warehouse development .....	17
3.0	System Analysis .....	20
3.1	Collab - RDBMS.....	20
3.2	Reporting tools .....	23
3.3	Microsoft Access & Excel.....	25
3.3.1	Sales.....	25
3.3.2	Other reports.....	30
4.0	Data Warehouse Modelling.....	35
4.1	System Design .....	35
4.2	System Development & Verification .....	36
5.0	Design and Implementation .....	37
5.1	Data .....	37
5.2	Neptun Kosovo Data.....	38
5.3	Requirements .....	39
5.5	Data Marts.....	42
5.7	ETL Process .....	50
5.8	ETL and SSIS (SQL Server integration Services) .....	53
5.8.1	Dimensional Tables.....	53
5.8.2	Fact tables.....	60
5.9	Creating views and reports.....	63
6.0	CONCLUSION .....	73
7.0	BIBLIOGRAPHY OF THESES.....	74

## TABLE OF FIGURES

Figure 1 Differences between Inmon and Kimball .....	15
Figure 2 Data Warehouse development life cycle .....	18
Figure 3 Collab .....	21
Figure 4 Report generated in Excel from Collab.....	24
Figure 5 Sales per store in Collab .....	24
Figure 6 Sales per product in Collab.....	25
Figure 7 Data in Microsoft Access .....	28
Figure 8 Import data to Access.....	29
Figure 9 Connection of Excel with Access Base.....	29
Figure 10 Connections for a pivot table .....	30
Figure 11 Stock Age in Collab .....	31
Figure 12 Stock Age in Access.....	32
Figure 13 Import & Export Documents .....	32
Figure 14 Import-Export data from Collab .....	33
Figure 15 Import & Export in Access .....	33
Figure 16 Stock Report from Collab.....	34
Figure 17 Stock Report in Access.....	34
Figure 18 Example of a star schema (Star Schema – Complete Reference, Christopher Adamson, p:10) .....	36
Figure 19 Star schema for Sales Data Mart .....	42
Figure 20 Logical Model for Sales Fact Table .....	44
Figure 21 Star schema for Stock Data Mart .....	44
Figure 22 Logical Model for Stock Fact Table .....	45
Figure 23 Star schema for Import & Export Data Mart .....	46
Figure 24 Logical Model for Import & Export Fact Table .....	47
Figure 25 Database in SQL Server Management Studio.....	48
Figure 26 Physical Design for Sales Data Mart .....	48
Figure 27 Physical Design for Stock Data Mart.....	49
Figure 28 Physical Design for Import & Export Data Mart .....	49
Figure 29 Loading a star .....	51
Figure 30 Loading a dimensional table.....	52
Figure 31 Loading a fact table .....	53
Figure 32 Dimensional Tables.....	53
Figure 33 Products table.....	54
Figure 34 Products table results.....	55
Figure 35 Date table .....	55
Figure 36 Date table results .....	56
Figure 37 Store table .....	56
Figure 38 Store table results .....	57
Figure 39 Employees table .....	57
Figure 40 Employees table results.....	58

Figure 41 Clients table.....	58
Figure 42 Clients tables results.....	59
Figure 43 DocType table.....	59
Figure 44 DocType table results.....	60
Figure 45 Sales fact table.....	61
Figure 46 Sales fact table results.....	61
Figure 47 Stock fact table.....	62
Figure 48 Stock fact table results.....	62
Figure 49 ImpExp fact table.....	63
Figure 50 ImpExp fact table results.....	63
Figure 51 View of sales of products in categories.....	65
Figure 52 Sales in categories over the years.....	65
Figure 53 View of Sales by Stores.....	66
Figure 54 Sales in Stores per month.....	66
Figure 55 View of Sales by channels.....	67
Figure 56 Sales by channels.....	67
Figure 57 View of Imported products.....	68
Figure 58 Imports by Supplier.....	68
Figure 59 Top products imported.....	69
Figure 60 View of Stock by Main Brands.....	69
Figure 61 Stock by Main Brands.....	70
Figure 62 View PowerBi.....	71
Figure 63 Reports created in PowerBi.....	71



## 1.0 RESEARCH FIELD

One of the most important things in a company it is for sure the informatics. Informatics is used for many purposes but two of them are the most important: operational record keeping and the other one is about decision-making. These informatics systems is where we as a company put the data that we get from different channels and based on that we generate reports using Data Warehouse and BI.

Users who work on these operational systems usually are the leaders of the company. These users take orders, do sales, register new customers and monitor all the other activities by using these operational systems. However, these operational systems do not maintain history, they rather update the data to show the new trending and reflect the reals state or the company.

Users of the DW/BI in the other hand, watch the wheels of the organization turn to evaluate performance. They count the new orders and compare them with the last week's orders. They worry about whether operational processes are working correctly. Although they need, detailed data to support their constantly changing questions. (The Microsoft Data Warehouse Toolkit 2nd Edition, p: 2)

Data warehousing and business intelligence are techniques to provide business people with the information and tools they need to make both operational and strategic business decisions. (The Microsoft Data Warehouse Toolkit 2nd Edition, p: 21). During this thesis I will try to break this down a bit so you can really understand the nature of the discussion that we are about to start.

Business intelligence is a collection of activities to understand business situations by performing various types of analysis on the company data as well as on external data from third parties to help make strategic, tactical, and operational business decisions and take necessary actions for improving business performance. This includes gathering, analysing, understanding, and managing data about operation performance, customer and supplier activities, financial performance, market movements, competition, regulatory compliance, and quality controls. (Building a Data Warehouse, Vincent Rainardi, p: 21).

## 1.2 AIMS OF THE RESEARCH

### 1.2.1 General Objective

Collecting the data in retail business is very important and there have been significant number of investments made in order to develop data warehouses for that purpose. One of the reasons why I choose this to be my thesis is because Neptun Kosovo the company for which I work for needs a data warehouse in order to help the management and the others in making better decisions with the data that we already have in different databases and those are available for some years in the past. Our management lack the possibility to have the data when they are most needed. Different departments in one retail company as ours have different data from different databases. Sometimes these data are the same but they are allocated in different places for different purposes. In our case, those data are extracted from the same source, which our database is called Collab. Those data are extracted on excel files and then modified from different departments in order to get the information that they need but this is done without any strategy. With the time passing, the management are realizing that having different databases for different departments is leading the company to incorrect results because the data are not matching with each other and there is lack of integration with the information. Due of the negative effect that having different databases is causing to the company the management of the company decided that it is time to work on a centralized data warehouse which will help us on enhancing data quality, save time, provide historical intelligence, generate more revenue and save more money.

Data warehouse provides businesses with powerful data analyses. It supports management of the company on making better business decisions and examine data to carry out analyses in the best way possible. Gathering data in a data warehouse and using business intelligence helps the company to measure the data using different variations like: sales, stock, budgeted values per store, sales per store, brands, clients, supply chain and trends. Neptun as a retail company can use this to optimize performance including: product prices, service prices, monitor the promotions. Making decisions based on the integrated data that we are going to generate from data warehouse will give us as a company benefits on monitoring marketing campaigns,

identifying account health, which will keep our customers away from our competitors, optimizing the operations and the last but not least increasing the product usage. Ultimately, producing better business decisions that lead to company success.

### 1.3 HYPOTHESIS

This thesis is going to be conducted on a retail company here in Kosovo, which is Neptun Kosovo SH.P.K. This topic is not something new on the technology because many companies around the world and in our country are operating and getting reports using the data warehouse that they have developed for their needs. Questions that need to be answered in this project are as below:

1. To see the importance of BI and DW in a retail company.
2. To design and develop a DW for Neptun.
3. To see the importance of the reports that are going to be generated from the DW on the decision making for the company.
4. To validate the design and development of data warehouse using the case study.

### 1.4 RESEARCH METHODOLOGY

This paper of dissertation for its own characteristic is based on quantitative method of analyses based on the issues that we are going to cover on this thesis. The quantitative methodology includes in itself what research is used to quantify the problem by way of generating numerical data that can be transformed into useable statistics. It is used to quantify attitudes, opinions, behaviours, and other defined variables, and generalize results from a larger sample population. Quantitative research uses measurable data to formulate facts and uncover patterns in research.

Data warehouse and business intelligence have become very important components on supporting the decisions that are taken by the business managers on the future of the organizations. Data warehouses after implementation will offer to our company access to integrated and historical data during the years and these data will support

managers on planning and decision making for the future months or years. The value of this data warehouse is going to be on the usage of it not on the implementation.

Building a data warehouse is going to be a very challenging work for me based on the education that I have and in the work that I have done in the past and what I am doing now but the most important thing is that data warehouse is a young discipline that is not still well-established on the strategies of the development. The current methods of developing a data warehouse are separated on three groups: data-drive, goal-driven and user-driven. All these three development approaches will be used on this decision support system that we are going to develop. In this thesis we are going to use all these development methodologies by different criteria that the users are going to give to us.

Quantitative research methodology that we are going to use usually includes various forms of surveys – online surveys, paper surveys, mobile surveys and kiosk surveys, face-to-face interviews, telephone interviews, longitudinal studies, website interceptors, online polls, and systematic observations. Face-to-face interviews is the most important method on collecting the functional and non-functional requirements from the users. Based on this interview we will be able to collect all the requires that we need in order to develop data warehouse.

## 1.5 IMPORTANCE OF THE THESIS

This master thesis paper will attempt to solve some of the issues that Neptun is having with the data quality over the years now. A data warehouse is a system that retrieves and consolidates data periodically from the source systems into a dimensional or normalized data store (Building a data warehouse – Vincent Rainardi, p:1). It usually keeps years of history and is queried for business intelligence or other analytical activities. It is typically updated in batches, not every time a transaction happens in the source system. (Building a data warehouse – Vincent Rainardi, p:1). From the data that we are going to get on the servers/files in Neptun we will consolidate those data into a dimensional model in order to have a better picture about the plans that the management will take for the future of the company.

The contribution of the research on developing data warehouse will enhance business intelligence to the company which will bring strategic, tactical, and operational insights. Data warehouse will contain a copy of the analytical data and this is going to expand the decision making. The data cleansing will give us the data quality that we always needed before using the reports for different purposes: sales, purchase, stock, budgeting etc. Doing this will allow us to have all the data in one place using different sources but generating reports from the same reporting tool. Another important thing is the large volume of the historical data which will help on analysing different time periods and trends that will help the managements make better decisions for the future.

## **2.0 LITERATURE REVIEW**

In this chapter, we are going to provide the necessary information about the data warehouse and business intelligence as topics that are going to be covered in this master thesis project.

### **2.1 Definition Data Warehouse & Business Intelligence**

Business Intelligence comprises a collection of methodologies, processes, architectures and technologies that transform raw data into meaningful and useful information for decision making (Data Warehouse Systems, Data Warehouse Systems p:3)

This data transformation comprises a set of tasks that take the data from the sources and, through extraction, transformation, integration, and cleansing processes, store the data in a common repository called a data warehouse. Data warehouses have been developed and deployed as an integral part of decision support systems to provide an infrastructure that enables users to obtain efficient and accurate responses to complex queries. (Data Warehouse Systems, Data Warehouse Systems p:3)

The fact table in a dimensional model stores the performance measurements resulting from an organization's business process events. (The Data Warehouse Toolkit, Ralph Kimball, p:10)

Dimension tables are integral companions to a fact table. The dimension tables contain the textual context associated with a business process measurement event. They describe the "who, what, where, when, how, and why" associated with the event. (The Data Warehouse Toolkit, Ralph Kimball, p:13)

This collection of a fact table and its dimension tables is called a data mart. There are more complex forms of data marts, such as when you have more than one fact table, but for now let's consider only one fact table. Bear in mind that this concept of a data mart is applicable only if the data warehouse is in a dimensional model. (Building a Data Warehouse, Vincent Rainardi, p:74)

Dimensional models implemented in relational database management systems are referred to as star schemas because of their resemblance to a star-like structure. Dimensional models implemented in multidimensional database environments are referred to as online analytical processing (OLAP) cube. (The Data Warehouse Toolkit, Ralph Kimball, p:8).

The extract, transform, and load (ETL) system brings data from various source systems into a staging area. ETL is a system that has the capability to connect to the source systems, read the data, transform the data, and load it into a target system (the target system does not have to be a data warehouse). (Building a Data Warehouse, Vincent Rainardi, p:2). ETL system integrates, transforms and loads data into a dimensional data store (DDS). ETL process structures the data into a data warehouse.

## 2.2 Data Warehouse

A data warehouse contains massive amounts of highly detailed, time-series data used for decision support. Data warehouses often contain terabytes of data that can be readily queried by end users. The sources of most of the data in a data warehouse are internal transaction processing systems (also known as operational systems). (Data Warehousing Battle of the Giants, Mary Breslin, p:7).

There are two main authors that wrote about the data warehouse and the forms on developing a successful data warehouse: William Inmon and Ralph Kimball. There are differences on the way that these two authors approach to this subject: the approach by Inmon is top-down design while that of Kimball is bottom-up design. Inmon describes a logical architecture that extracts detailed, time-stamped data from disparate operational databases. The data is then transformed and stored in a single database (the data warehouse). Data extracts from this monolithic data warehouse create smaller, departmental databases. Decision support users query and create reports from the departmental databases. To create both the data warehouse and the departmental databases, Inmon proposes a top-down variation of the spiral system development methodology. (Data Warehousing Battle of the Giants, Mary Breslin, p:7). In the other hand Kimball, he recommends an architecture of multiple databases, called data marts, organized by business process. The sum of the data marts comprises the data warehouse. He recommends a development methodology that is unique to data warehousing. It involves a bottom-up approach that must adhere to an enterprise wide standard “data bus”. (Data Warehousing Battle of the Giants, Mary Breslin, p:7)

One of the most important differences between the two theories from Kimball and Inmon is that Kimball's data warehouse is de-normalized in the other hand Inmon uses a highly normalized central data warehouse.

The most prominent similarities between Inmon's and Kimball's models are the use of time-stamped data, and the extract, transform, and load (ETL) process. Although the execution of these two elements differs between the two models, the data attributes and query results are very similar. (Data Warehousing Battle of the Giants, Mary Breslin, p:15)

In the image below, you can check some of the most important differences between two authors:

	Inmon	Kimball
<b>Methodology and architecture</b>		
Overall approach	Top-down	Bottom-up
Architectural structure	Enterprisewide (atomic) data warehouse "feeds" departmental databases	Data marts model a single business process; enterprise consistency achieved through data bus and conformed dimensions
Complexity of the method	Quite complex	Fairly simple
Comparison with established development methodologies	Derived from the spiral methodology	Four-step process; a departure from RDBMS methods
Discussion of physical design	Fairly thorough	Fairly light
<b>Data modeling</b>		
Data orientation	Subject- or data-driven	Process oriented
Tools	Traditional (ERDs, DISs)	Dimensional modeling; a departure from relational modeling
End-user accessibility	Low	High
<b>Philosophy</b>		
Primary audience	IT professionals	End users
Place in the organization	Integral part of the Corporate Information Factory (CIF)	Transformer and retainer of operational data
Objective	Deliver a sound technical solution based on proven database methods and technologies	Deliver a solution that makes it easy for end users to directly query the data and still get reasonable response times

*Figure 1 Differences between Inmon and Kimball*

None of these two authors are wrong or right about data warehousing. Ralph Kimball's idea about data warehousing is mostly used on the companies with different departments. These departmental data warehouses are called data marts. However, these data marts are not necessarily private to a department; they may be shared with other interested parts of the organization. (Data Warehouse Concepts, p:74)

### 2.3 Data modelling techniques

The type of analysis that will be done with the data warehouse can determine the type of model and the models' contents. Because query and reporting and multidimensional analysis require summarization and explicit metadata, it is important that the model contain these elements. (Data Modelling Techniques for Data Warehousing, Chuck Ballard, p:13). Data modelling is the technique that we choose in order for the data to be stored in the data warehouse. Data modelling helps on the visual representation of the data and it enforces business rules that we are going to set on the data warehouse, It usually takes care about what data are needed and how the data are going to be organized not the operations that are going to be performed with the data.

There are three levels of data modelling: high-level modelling (called the ERD, entity relationship level), midlevel modelling (called the data item set, or DIS), and low-



level modelling (called the physical model). (Building the Data Warehouse, W.Inmon, p:92)

The high-level called Entity Relationship Level features entities and relationships. The name of the entity is surrounded by an oval. Based on W. Inmon articles relationships among entities are depicted with arrows. The direction and number of the arrowheads indicate the cardinality of the relationship, and only direct relationships are indicated. In doing so, transitive dependencies are minimized.

ERL offers organizations coverage of the business processes and concepts. This data model is designed for business audience. The main focuses of the high level are to present the data as a user would see them in the real world.

The second level called the Logical Data model adds more information to the conceptual model. It defines the structure of the data and sets the relationships between them. At this model, no primary or secondary keys are defined. Here you just need to verify and adjust the connector details that were set earlier for relationships. Here are some of the characteristics of logical data model:

- Describes data needs for a single project but could integrate with other logical data models based on the scope of the project.
- Designed and developed independently from the DBMS.
- Data attributes will have datatypes with exact precisions and length.

Finally, physical model is created from the midlevel data model merely by extending the midlevel data model to include keys and physical characteristics of the model. At this point, the physical data model looks like a series of tables, sometimes called relational tables. (Building the Data Warehouse, W.Inmon, p:98). This model helps on generating schemas because of the Meta data offered by a Physical Data Model. It also helps to visualize the database structure. It helps us to model database column keys, triggers, indexes and other necessary features. Here are some of the characteristics of physical

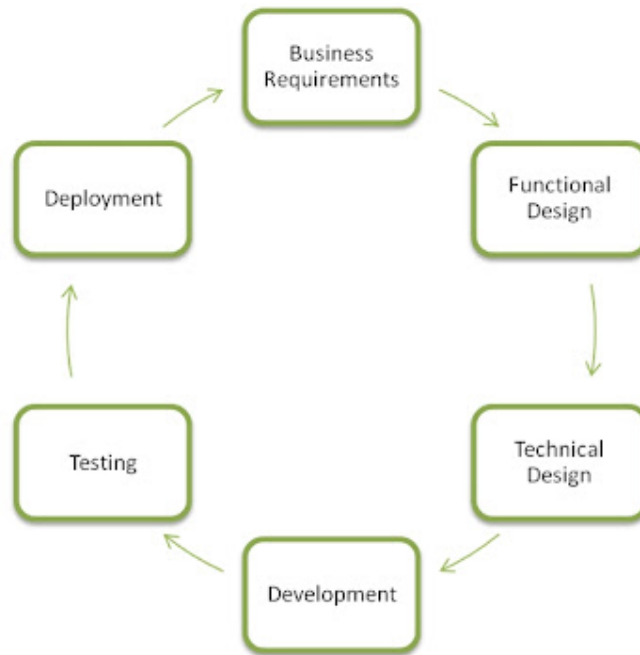
- The physical data model describes data need for a single project or application.

- Physical Data Model contains relationships between tables which addresses cardinality of the relationships.
- Columns must have set the exact datatypes, lengths assigned and default values.
- In the physical model, there are defined Primary and Foreign keys, views, indexes, access profiles, and authorizations, etc.

## 2.4 Data Warehouse development

The most important thing it's the gathering of all organization's information and data in one place and then using those data on making better decisions for the company all of this using the same data tools that we are going to choose based on our needs. However, to make this happen you must do a good planning of the development of the data warehouse. We as the developers or project managers must understand the requires of the system users. What do they like, what do they need or use on everyday work? The main thing why data warehouse system is used is to provide decision-makers the accurate, timely information they need to make the right choices for the future of the company in this case Neptun.

With the market changing quickly we as a company need to be faster also on making decisions, creating new business models or new practices on sealing products and these decisions with help our company grow. The decisions are going to be made based on the data information that we are going to gather from our sales, purchases, budgeted values that are achieved on the past years.



*Figure 2 Data Warehouse development life cycle*

First, we need to gather and analyse information from the users and the managers, the information that they used until now to make the decisions. We can start with all the reports and information that we can get from different departments and check them how they were generated. All departments or people have data that they use as reports to gather the information that they need on daily, monthly or yearly bases. Some of these data are try some not it depends on the way that those are created. A big challenge for a data warehouse developer is to find these data and understand why were these data or reports used for in order to have an idea of the needs of each one off the users that generated them.

Another part of this collection and analysis phase understands how people gather and process the information. Developing a data warehouse is going to automate must of the tasks used before but to do that first we must understand the requires that each one of the users has.

Second, after identifying the data processes that we are going to correlate we have to identify key performance indicators, like products sold, sales per store and the revenue generated from the sales. Data warehouse is a collection of data structures or data marts that we explained earlier. Each structure stores key performance indicators for a specific business process and correlates those indicators to the factors that

generated them. To design a data structure, we must first identify the entities that work together in a data mart.

After identifying the entities, we must identify the key performance indicators into fact tables. In addition, as it is said in the Kimball's books to include a set of facts, you must relate them to the dimensions.

After identifying the business processes, you can create a conceptual model of the data. We are going to choose the information that are going to show in the fact tables and the dimensions that are going to be connected with them. Each row in the fact table is generated by the interaction of specific entities on a data warehouse. To add data on a fact table you need to populate all the dimensions and correlate their activities. Before using the data on a data warehouse, you must first correct them then use them on dimension or fact tables. Maintaining and populating a data warehouse is not going to be easy for that reason we must do a careful planning.

Now that we know what we need we have to find the data sources and put those data on a data warehouse. Data that are going to go on a data warehouse must be consolidated and structured. We must be sure that the data are in good shape and completed before using them because incorrect data or data types will cause us problems in the future. The cost of fixing bad data can make the system cost-prohibitive, so we must correct the data in order to have less problems. The corrections must be done in the source data this is the best practice to follow in order to have better data.

Data transformations must be done before putting the data in the. Some might involve converting the data storage type. Some transformations are changing the unit measures. Some of the transformations are easier some more difficult or some others might require programming in order to apply an algorithm to do the transformation that is needed.

Finally, after we develop the plan, you can estimate the work and the time that is needed in order to develop the data warehouse. An effective strategy is to plan the entire data warehouse, then implement a part as a data mart to demonstrate what the system can do. The project is going to be done part by part and each part is going to

play its role on the structure by bringing something that is going to fulfil the needs of the users.

### 3.0 System Analysis

In this chapter, we are going to see the data, data formats, RDBMS, and other files used from Neptun over the years. We will try also to describe the benefits of DW for our company and what we will achieve after implementing it. Based on the requirements functional or non-functional requirements that we will get the overall goal is to bring to the company the data reports that they need in order to make better decisions that will affect the work off the clients in this case Neptun employees and managers and also it will have impact on the raising of the company in the total market share of the country.

After analysing the data files and formats, we will understand what the working user's requirements are and what are the reports that they need in order to finish their duties in the most efficient way possible. Based on the attributes that we will get we are going to build entities and relationships that will help us on generating the reports that are needed.

The interviews that we are going to do with the clients will help us on gathering the requirements needed for further developments of data warehouse.

#### 3.1 Collab - RDBMS

Before we start to describe the RDBSM that we use in Neptun let me explain first what a RDBMS means. A relational database consists of a collection of tables, each of which is assigned a unique name. The relational model is the primary data model for commercial data processing applications. It is being used in a number of applications outside the domain of traditional data processing. (Relational Database Management System, ISRD Group, p:35).

An abstract theory exists for relational databases. The relational model represents the database as a collection of relations. Each relation correspondents to a table of values

or to some extent, a “file” of records. (Relational Database Management System, ISRD Group, p:35).

The RDBMS software that we use in our company called “Collab”. Collab is powerful web-based ERP (Enterprise resource planning) solution designed based on the needs of the company.

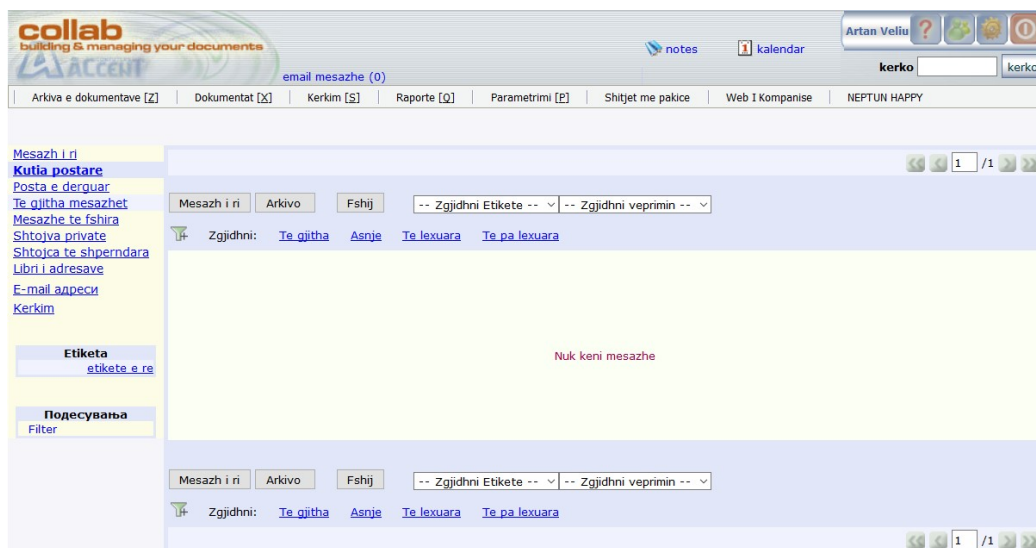


Figure 3 Collab

Below I will list the components that Collab as a web-based software can do for us as clients based on the comments of the company that has developed it:

- Record of SKUs with code, name, barcode, normative/recipes. The possibility for defining unlimited number of additional attributes on whole application level that can be used on many places is a unique function. For example, a company that works with auto-parts has the possibility to add unlimited number of catalogue numbers, names, codes, barcodes etc.
- Grouping the articles into groups, subgroups etc.
- Record of customers with code, name, address, contact information and possibility for customers differentiation and making different price lists, additional attributes, organization sectors etc.
- Creating a regular price list, unlimited number of different price lists.
- Creating different types of documents with appropriate characteristics according to the needs of the company.
- Logging system for users with different access level to the system options.

Here are some of the managerial materials that Collab offers:

- Procurement from domestic businesses
- Import calculation from foreign partners with possibility of forming original cost through assessment of costs such as customs payments, transportation, forwarding etc.
- Sales to foreign buyers via Export Invoices on foreign language and foreign currency.
- Production through defining the products and recipes for their production from raw materials and other articles via Production Order (assemblage or dismantlement).
- Stock correction via stocktaking and report for passed expired dates.
- Wholesale via Dispatch notes/Invoices, Consignments.
- Retail sales via cash invoices, bills, Certificates of receipt, fiscal receipts.
- Return of the sold articles from buyers via returning documents or cancelation of documents.
- Creating Orders to suppliers/from buyers, creating Offers/Preform invoices with generating automatic Dispatch note/Invoice after accepting the Offer.
- Transfer the stocks between warehouses, from warehouse to the shop or backward.

Commercial works and management:

- Record of payments from buyers and payments to suppliers with creating bank statement and payment charges.
- Record of non-paid items /invoices from clients/suppliers.
- Creation of Invoices from Dispatch note or creating Invoice from more different Dispatch notes (group document).
- Dealer price lists for dealers or loyal buyers.
- Assignment of percentage discounts or nominal and generating automatic document (approval for discount).
- Record of cash payments or payments with debit cards.

Finally, the reports that can be generated based on Accent:

- Real time stock list and stock list of concrete date from the past.

- Material card for article with whole chronology of changes.
- Detailed report for the sale via bills through different filters (for example: buyer, shift, time etc.).
- Sales specification through all types of documents and all methods by filtering and sorting.
- Report for erased documents and possibility to see who has done the action
- Report for retail sales.
- Report for trading services.

All the points that I have listed in this chapter are based on what the outsource company reviews about their software. Collab uses a database developed using Microsoft SQL but unfortunately, we cannot access on that database because it is permitted from the company.

### 3.2 Reporting tools

There are different reporting tools that we use in the company in order to modify the data and present them in tables, pivot tables, charts and other visualisations so the users can find the data that they need in order to do their job correctly. There are different kind of reports that can be generates: the users cannot change static reports while interactive repots allow us to modify them and use different types of hierarchy to visualise the data as we like. Interactive reports allow us to drill down in different levels of the data that we have with just one click of button.

The purpose of reporting tools is to modify the data and translate them on a meaningful information. Reporting is very important on achieving the strategic goals of the company because based on that the decisions are going to be made.

With the raising of the company, the reporting capabilities of the company must rise also. The tools that we use to generate and modify reports must be centralised in order to have a better efficiency with the data. The ability to make reports in many different ways has grown in the past years and the technology as evolved also. Embedded reporting tools offer us different ways of exporting and analysing the data.



The two reporting tools that we use today besides Collab, which is the RDBMS that we use, and it generates reports from the MySQL database are Microsoft Access and Excel. The reason for using other reporting tools besides Collab are because Collab as a reporting tool does not give us the ability to generate different kinds of reports and the ones that we can see are not as helpful as we would like them to be. For this, reasons we generate excel files form Collab and use those data after modifying them to make different reports.

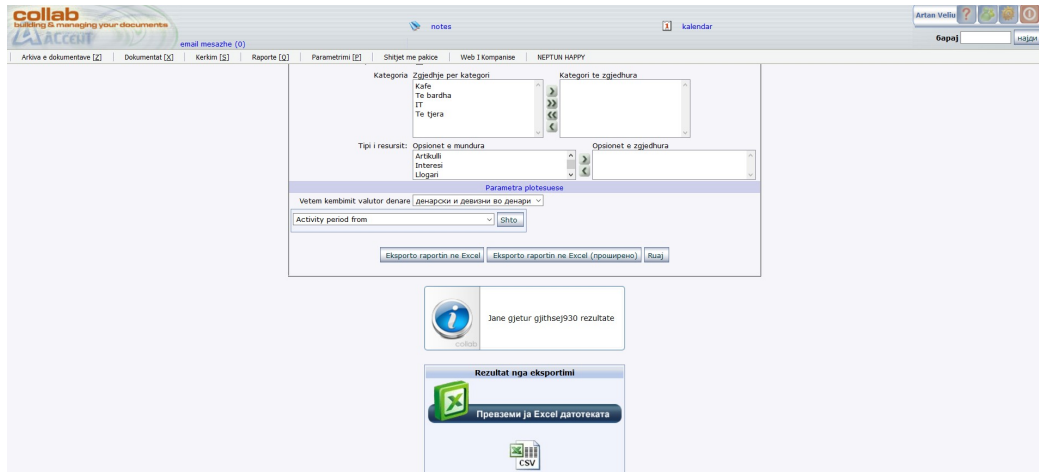


Figure 4 Report generated in Excel from Collab

In the Figure 4 that you can see, the sales are generated for one specific day from Collab in Excel. Collab also gives us the opportunity to see the sales there without extracting them in excel but they are not divided in months, days or years it just shows the period that you have selected example the sales from 01.10.20 until 10.10.20 will be showed as a total we cannot see those sales separated in days or weeks. In Figure 5, you can see the sales based on the stores for a specific day of the month October 2020.

Nr.	Njesia Afariste	Sasia dalese	Vlera e mallit vleresuar me kosto
1.	01 Shumica	78.00	
2.	101 Shitorja 1 - Albi	10.00	
3.	103 Shitorja 3 - Peje	9.00	
4.	106 Shitorja 6 - Prizren 2	9.00	
5.	107 Shitorja 7 - Fushe Kosove	32.00	
6.	108 Shitorja 8 - Ferizaj	5.00	
7.	109 Shitorja 9 - Gjilan	7.00	
8.	114 Shitorja 14 - Royal Mall	1.00	
9.	115 Shitorja 15 - Mitrovice	1.00	
10.	117 Shitorja 17 - Vushtrri	4.00	
		156.00	

Figure 5 Sales per store in Collab

There are also other forms to check different things in Collab like stock per store or stock of products in total or you can check detailed sales for each product or each brand in each store but that is also not grouped in months or years:

Nr.	Shifra	Artikull	Shifra e komitentit	Komitent	Data	Lloji I dokumentit	Nr I dok	Autor	Grupi i artikullit	Тип бренд	Sasia dalese	Vlera e mallit vleresuar me kosto
45.	84053230	FILTER ROWENTA ZR006001	1007	Neptun Kosova SHPK	27.10.2020 09:10	Kupon tatimor	101-502/265211	Leonora Krasniqi	618430	ROWENTA	1.00	19.00

Figure 6 Sales per product in Collab

### 3.3 Microsoft Access & Excel

Microsoft Access and Excel are two of the most powerful tools from Microsoft that are used for generating different reports and doing data analyses on a daily basis. Excel is an old product from Microsoft whereas Access is a newer and more complex product.

In our case, we use these two Microsoft products a lot in order to generate and modify different reports based on the data that we extract from Collab. The data extracted from Collab are in Excel and after transforming them with the important things that we need in order for the data to be ready to for Microsoft access.

#### 3.3.1 Sales

In the chapters, we are going to see the structure of the reports that are actually using in our company. One of the most important reports that we use sure is Sales. Based on this report we see the sales based on Stores, Brands, Types of Sales, and many other possibilities, which are listed in the table below. The report is firstly extracted from Collab and these are the attributes that it gives us:

Header in extracted from Collab	Translated in English
Шифра	Code
Артикл	Article
КОМИНТЕНТ	Customer
Град	City
Датум	Date

Време	Time
тип на продажба	type of sale
продажба/набавка	sale / purchase
Деловна единица	Business unit
Автор	Author
Количина излез	Quantity output
Продажна вредност со ДДВ	Sales value with VAT
Набавна цена	Purchase price
Набавна вредност	Dealer cost
Рабат во денари	Discount
Одобрил рабат	Approved rebate
Главна Група	Main Group
Група 1	Group 1
Група 2	Group 2
Група 3	Group 3
Група 4	Group 4
Група 5	Group 5
Категорија	Category
Бренд	Brand
Тип на документ	Document type
Број на документ	Document number
Ден	Day
Седмица	Week
Месец	Month
Тромесечје	Quarter
Полугодие	Half a year
Година	Year
Furnitori	Supplier
Промоција ID	Promotion ID
Број на партија	Party number
Каматна стапка	Interest rate
Број на рати	Number of installments

The documents that we use for sales and must be selected in Collab in order to get the sales are these:

DocumentType	TypeOfSales
Fature shitje malli nga dyqani	Retail
Fature shitje malli	WS
Fature shitje malli Kesh	WS
Diferencat e stokut pas inventarit	Retail
Fature- me jashte shtetit	Export
Kupon tatimor	Retail
Credit Note-(Dok.jone- brenda vendit)	WS
Fature shitje malli nga servisi	SVC
Fature shitje malli - sindikate	Retail
Credit Note-(bonuset e shumices )	Retail
Fature shitje malli me kredi – NLB	Retail

Fature shitje malli me IUTE CREDIT	Retail
Fature shitje malli me kredi – New Co Ferrnikeli	Retail
Fature shitje malli nga dyqani per person juridik	Retail
Fature shitje malli online	Online
Kupon tatimor - online	Online

This is how the attributes look after transformation and ready to extract in the Microsoft access:

Attributes	Entity
Pcode	84054540
Product	NXEMSE FUEGO PTC-3000R
TAX	1.18
GroupCode	422000
SEGMENT	HVAC
FAMILY	HEATING
CLASS	FAN HEATERS
BRICK	0
CAT	0
SUBCAT	0
Psegmentation	0
PM	Alex D.
MainBrand	FUEGO
Brand	FUEGO
Client	Neptun Kosova SHPK
ClientType_L	Other
ClientCity_L	Mitrovicë
SUB_L	Shitorja 15 - Mitrovice
SUB	H.Mitrovicë
DocNo	115-502/xxxx
DocType	Kupon tatimor
TypeOfSales	Retail
DocAuthor	Filan Fisteku
QTY	1
GS	39.00
NS	33.05
COGS	13.73
Discount	0.00
Promo	Normal
Time	9:03:01 AM
Date	26.10.2020
Day	26
WD	1
W	44
MON	OCT
M	10
Q	Q4
H	H2

Year	2020
Y	ACT
Area	0
Disc	0
Supplier	0
PromoID	0

After the data are transformed on the format that we use we import those data with an excel file in Microsoft access.

Field Name	Data Type		
Pcode	Number		
Product	Short Text		
TAX	Number		
GroupCode	Number		
SEGMENT	Short Text		
FAMILY	Short Text		
CLASS	Short Text		
BRICK	Short Text		
CAT	Short Text		
SUBCAT	Short Text		
PSegmentation	Short Text		
PM	Short Text		
MainBrand	Short Text		
Brand	Short Text		
Client	Short Text	Time	Date/Time
ClientType_L	Short Text	Date	Date/Time
ClientCity_L	Short Text	Day	Number
SUB_L	Short Text	WD	Number
SUB	Short Text	W	Number
DocNo	Short Text	MON	Short Text
DocType	Short Text	M	Number
TypeOfSales	Short Text	Q	Short Text
DocAuthor	Short Text	H	Short Text
QTY	Number	Year	Number
GS	Number	Y	Short Text
NS	Number	Area	Short Text
COGS	Number	Disc	Short Text
Discount	Number	Supplier	Short Text
Promo	Short Text	PromoID	Number

Figure 7 Data in Microsoft Access

The data that we generate daily form Collab are transformed and then imported to Access every day in this form:

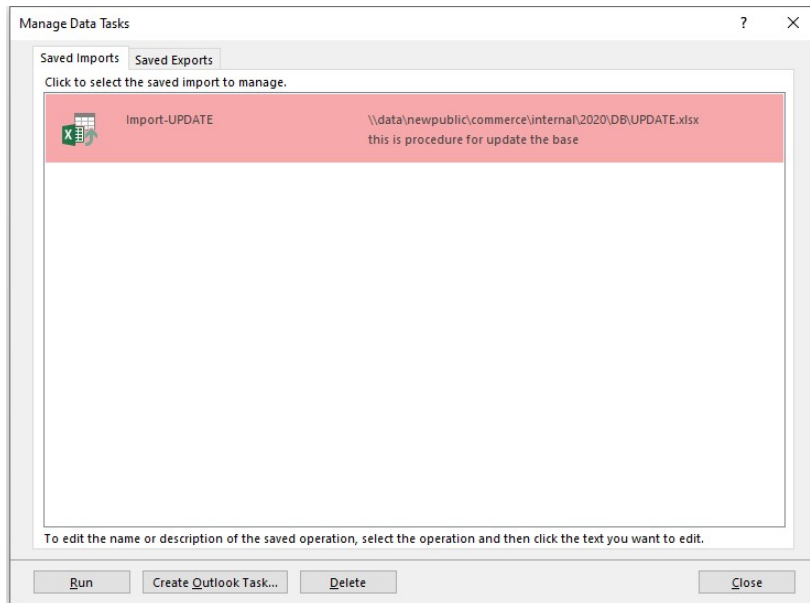


Figure 8 Import data to Access

The data can be imported daily or you can group two or more days and import them at once. This is done based on the needs. As this report is for daily sales, we do it daily.

Microsoft Access file, which is located in a public folder in our serves, is connected with Excel files where we generate all the reports that we need. The connection is done in this form:

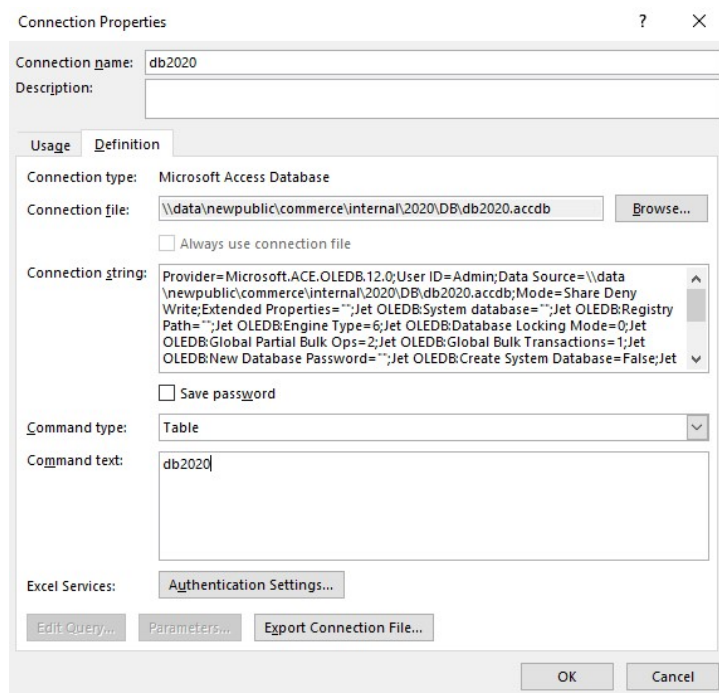
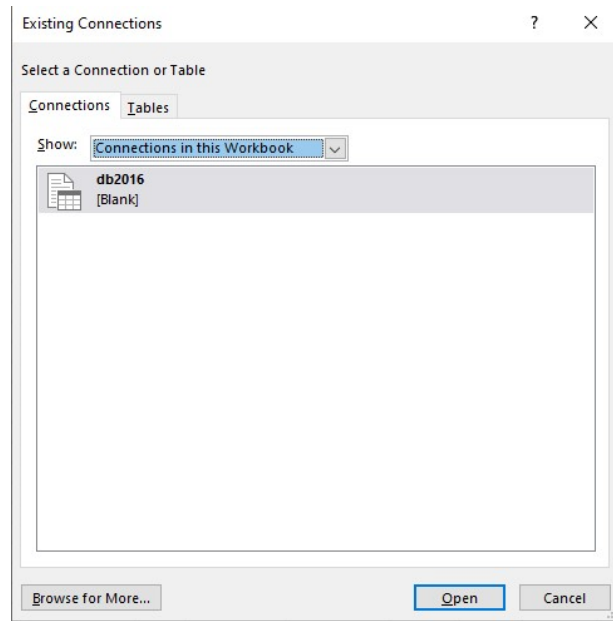


Figure 9 Connection of Excel with Access Base

After the connection is finished, we generate pivot tables that give us the information that we need about sales values, budgeted values per store per category and many other information. Based in the attributes that I showed in Figure 7 you can generate different reports depending from the needs of the employees.



*Figure 10 Connections for a pivot table*

All the employees or managers that need information about sales have their files connected with this Access base and every day they refresh the file once and all the data will be transferred at their files.

### **3.3.2 Other reports**

In our company currently, we are using Microsoft Access as an external database for modifying and generating the reports that we need. As you may have seen in the beginning of this chapter, Collab, which is our database, does not give us many opportunities on generating the files and data on the formats that are suitable for us as employees and managers.

Using Access has given us some advantages but for a company like Neptun we cannot use Access for a long time and there are some reasons for that. First, space that you are allowed to store data in access is 2GB, we are closely approaching to that amount

just with five years history of sales data, and we cannot use it anymore. Second is the data security. For example, you will require more extensive database security than Access can offer. Finally, Access cannot be used on MAC PC's it is designed to be used just in business PC.

The need for different kind of reports in the company high so we have also other reports generated from Collab and transformed as we do with the sales report. For the moment there are many other reports done in the same way using Microsoft Access and Excel. Departments like finances or marketing they only work by using Excel based on the data that they generate in Collab but there are many files which can be deleted or have inefficient data that are not usable at any time.

Other reports generated and transformed from the commercial department are done Daily, Weekly, Monthly or Yearly. All these reports have a specific meaning and show different kind of information that are helpful for the company. The reasons for transformation and load of the data from Collab to Access or Excel is same as in sales report. The data are extracted because you cannot get all the kind of information that you need from Collab. Collab as a software/database does not give you the opportunity to see the reports on the way that are suitable for managers and other decision makers.

Below you can find the data extracted from Collab for the Stock Age report that we do on weekly bases:

Шифра	Артикул	Вкупен лагер	Просечна цена	Вкупна просечна сума	Денови										
					0 deri ne 60		60 deri ne 90		90 deri ne 120		120 deri ne 180		mbi 180		
					кол.	Просечна сума	кол.	Просечна сума	кол.	Просечна сума	кол.	Просечна сума	кол.	Просечна сума	
84053128	ADAPTER ACME WIRELESS PER MOBILE CH-301	6.00	17.80	106.78	6.00	106.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 11 Stock Age in Collab

As you can see in the Figure 11, the data generated for the age of the stock, which we do every week, has just the basic things that a report must have. In this form, you cannot see the weeks of the year 2020 or the product category of the products if the user likes to group the products based on categories. In order to see more data, we usually do the transformation of these data and import them in Access based on these attributes and data types:



Field Name	Data Type
ProductCode	Number
Product	Short Text
Short Description	Short Text
TAX	Number
GroupCode	Number
SEGMENT	Short Text
FAMILY	Short Text
CLASS	Short Text
BRICK	Short Text
CAT	Short Text
SUBCAT	Short Text
PM	Short Text
MainBrand	Short Text
Brand	Short Text
WH	Short Text
SUB_L	Short Text
SUB	Short Text
QTY	Number
LCP	Currency
COG	Currency
0-60Q	Number
0-60A	Currency
60-90Q	Number
60-90A	Currency
90-120Q	Number
90-120A	Currency
120-180Q	Number
120-180A	Currency
>180Q	Number

Field Name	Data Type
Date	Date/Time
W	Number
MON	Short Text
M	Number
Q	Short Text
H	Short Text
Year	Number
Supplier	Short Text

Figure 12 Stock Age in Access

The connection between Access and Excel is same as in Sales report just with different path as the file now is Stock Age.

Next, I will describe a monthly report with is about imported and exported products from our suppliers. Suppliers can be local or non-local it depends. Collab does not give us the opportunity to generate a meaningful report about imported and exported product. The documents that we select in order to get the data are these:

Tipa te dokumenteve	Tipa te zgjedhura te dokumenteve
Fature shitje malli	Fature ne valute- hyrje ne depo doganore
Fature shitje malli Kesh	Fature ne valute- hyrje ne depo kryesore
Fature shitje malli me IUTE CREDIT	Fature- me jashte shtetit
Fature shitje malli me kredi - New Co Fernikeli	Hyrje malli ne magazine -furnitori eshte nga kosova
Fature shitje malli me kredi - NLB	
Fature shitje malli nga dyqani	

Figure 13 Import & Export Documents

The report generated:

Nr. Rend.	ID	Shifra	Artikull	Shifra e komitentit	Komitent	Data	Lloji I dokumentit	Nr I dok	Autor	Grupi i artikullit	Тип бренда	Sasia hyrese	Sasia dalese	Vlera furnizuese me TVSH
1	10154	84050017	FUEGO DERRASE	1009	Depo Dogand	04.11.2020	Fature ne valute- hyrje	04-310/30538	Rrezart&	618320	FUEGO	90	0	

Figure 14 Import-Export data from Collab

Same as with the other reports here are the attributes and data types that are going to be imported in Microsoft Access:

Field Name	Data Type
PCode	Number
Product	Short Text
Client	Short Text
Date	Date/Time
DocType	Short Text
DocNo	Short Text
Brand	Short Text
MainBrand	Short Text
GCode	Number
SEGMENT	Short Text
FAMILY	Short Text
CLASS	Short Text
BRICK	Short Text
CAT	Short Text
SUB	Short Text
PM	Short Text
PSegmentation	Short Text
QTY	Number
AmountL	Number
Curr	Short Text
Rate	Number
Amount	Number
Price	Number
Export	Number
MON	Short Text
M	Number
Q	Short Text
Year	Number
Type	Short Text

Figure 15 Import & Export in Access

Finally, we are going to see the stock report generated from Collab and transformed to be used in Excel. About stock, there are some reports that you can check directly in Collab or export the data in Excel but still you are going to miss a lot of necessary information for making decision. The data that we see in Collab are usually a faster way for checking something but you really cannot make decisions based on those data. In the Figure 15, you can see a report for stock generated from Collab:

Magazina	Shifra	Artikull	Cmimi mesatar	Sasia e momentit	Totali
04 Depo Kryesore 2	84037075	LG SYZE 3D	0.33	5.00	1.66

Figure 16 Stock Report from Collab

This is a simple report about stock at a specific date generated in Collab. The report is based on stores/warehouses and products. As you can see there is also included the cost of the product and the quantity that is available now in a specific store. There are other attributes that can be added from Collab for this report like prices because we have different kind of prices for a product and also you can list the stores to be in a row not in column. In Figure 16, we can see the attributes that this report has after transformation.

Field Name	Data Type
MAG	Short Text
DATE	Date/Time
PCODE	Number
PRODUCT	Short Text
LCP	Number
QTY	Number
TTL	Number
WH	Short Text
MON	Number
Y	Number
Brand	Short Text
MainBrand	Short Text
GCode	Number
SEGMENT	Short Text
FAMILY	Short Text
CLASS	Short Text
BRICK	Short Text
CAT	Short Text
SUB	Short Text
PM	Short Text
TAX	Number
RP	Number
WSP	Number
Supplier	Short Text

Figure 17 Stock Report in Access

There are many other reports that we use in our everyday work but the ones listed in this chapter are the most important ones. Nowadays we are also using power pivot and queries to generate different kind of PSI (Purchase Sales Inventory) reports that are very important for getting a clear view of sales and inventory movement between stores. This kind of reports are a huge help for the product managers on their everyday tasks but unfortunately these reports are all done manually via using Microsoft Excel tools.

## 4.0 Data Warehouse Modelling

### 4.1 System Design

In this chapter, the focus is going to be translating the requirements on a set of specifications through deriving logical and physical data models or data marts for the data warehouse. These requirements are going to be used on the future on implementing the data marts and the data warehouse tables. Many occasions can affect the decision-making in the company starting from daily life events to technical reasons.

As I maintained earlier the two main people who evolved the data warehouse in the past years were Ralph Kimball and Bill Inmon and as we said there are differences on the way these two approaches to the design of data warehouse. For many reasons on this thesis, I will use Ralph Kimball's method of developing the data warehouse. The data from each department are going to be gathered and then separated in different data marts. Then we will design the data warehouse and business intelligence processes from start until the end and the case study is going to be about Neptun.

As we are going to use Ralph Kimball's development methodologies, we are going to use also star schema for the data warehouse design. The dimensional model of a business process is made up of two components: measurements and their context. Known as facts and dimensions, these components are organized into a database design that facilitates a wide variety of analytic usage. Implemented in a relational database, the dimensional model is called a star schema. Implemented in a multidimensional database, it is known as a cube. If any part of your data warehouse includes a star schema or a cube, it leverages dimensional design (Star Schema – Complete Reference, Christopher Adamson, p:3). A dimensional design for a relational database is called a star schema. Related dimensions are grouped as columns in dimension tables, and the facts are stored as columns in a fact table. The star schema gets its name from its appearance: when drawn with the fact table in the centre, it looks like a star or asterisk (Star Schema – Complete Reference, Christopher Adamson, p:10).

At the core of a star schema is the fact table. In addition to presenting the facts, the fact table

Includes surrogate keys that refer to each of the associated dimension tables. Each row in the fact table stores facts at a specific level of detail. This level of detail is known as the fact table's grain (Star Schema – Complete Reference, Christopher Adamson, p:12).

In a star schema, a dimension table contains columns representing dimensions. As discussed, these columns will provide context for facts. The dimension tables serve to provide the rich context needed for the study of facts. In queries and reports, the dimensions will be used to specify how facts will be rolled up—their level of aggregation. Dimension values may be used to filter reports. (Star Schema – Complete Reference, Christopher Adamson, p:10).

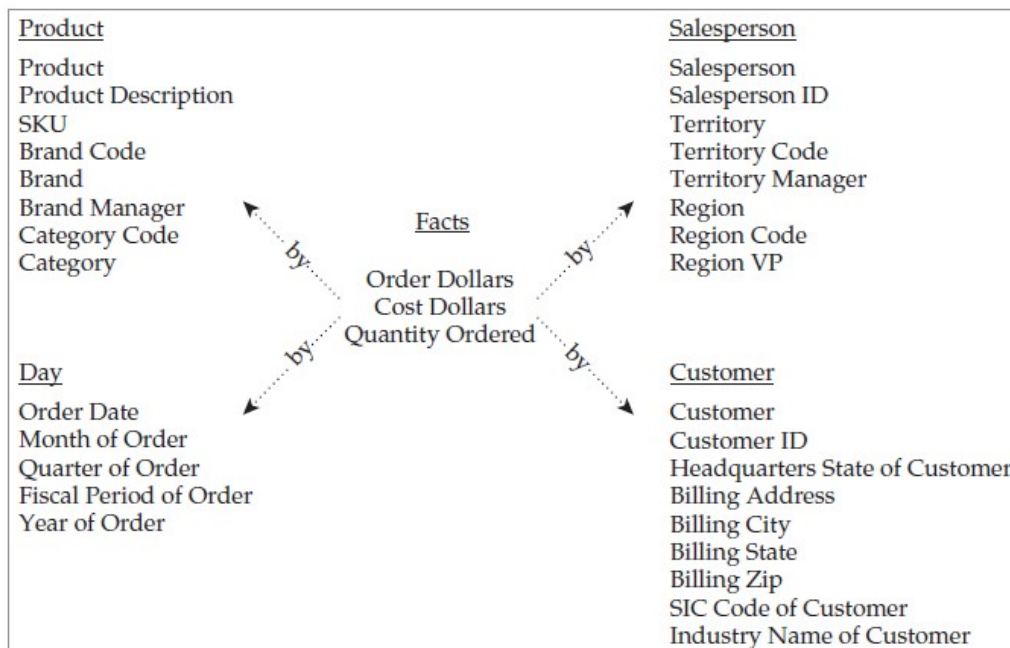


Figure 18 Example of a star schema (Star Schema – Complete Reference, Christopher Adamson, p:10)

## 4.2 System Development & Verification

When we are talking about the system development and verification of the data carried out, we are actually thinking about the actual implementation of the fact and

dimension tables and the ETL (Extract, Transform and Load) process of the data warehouse. After the implementations is going to be done, we will start with the data verification, are the data valid compared to the data that we had before, are those data valise compared to the source data that we have and finally which are the result that we are getting from those data.

The best way to make the data validation is by preparing reports and filtering the data. Those reports will be generated from our data warehouse on a reporting tool like Power BI or other reporting tools depend which one is going to be more suitable for our case.

### 4.3 Design and Implementation

In this chapter, we are going to be concentrated in three main topics that are going to be: system analyses, design and finally the implementation of the data warehouse for a retail company, which in this case is Neptun.

System analyses is the process of studying the current system, analysing the needs and problems, finding alternate candidate solutions, evaluating each of them separately, and finally choosing the best solution. (System Analyses and Design, Arunesh Gayal, p:8)

During the analyses that we have done to the system which we use every day in work to perform different tasks we studied the activates of the company and we concluded that the most important and urgent thing to work on are the reports that usually are used form the Commercial Department. We collected all the data and the files that can be generated also we did do interviews in order to get the requirements from the stake holders and the users and by doing this it becomes easier to designee the new system.

### 4.4 Data

One of the most important things about companies is data and the way that they are showed to the decision makers or other users. Companies are around the world in all industries have integrated data as an essential part of their business processes. Now

more than ever this data collected from different systems and sources allow companies to acquire fruitful information and knowledge of the company, other competitors, and customers.

In today's environment that the products offered from different companies are countless, customers have a variety of choices and the competition between the companies is becoming tenser than ever even in our small market that we have in Kosovo. To avoid losing the customers the most important thing is understanding their desires and this is done by gathering the data. Data is the true voice of the customers and with this data companies can draw a picture of their customers and, more importantly, their interest and willingness to buy. It is also found that leading companies are turning data into insights and taking action to offer customers the products that match their needs most. Using this kind of data and gathered from the customers desires and connecting them with the sales and inventory data that we have from our system will help us on organizing the data on our DW and BI and have better information about the sales, purchase and inventory on our company.

#### 4.5 Neptun Kosovo Data

The data that we can extract from our software in Neptun are enough to get all the information that we need for archiving the results that we like to see. However, in order to achieve those results, we must transform those data and link them together. All the reports that we see today are separated in different files and comparing data from different excel files does not give us the detailed information that we are looking for. For that reason, developing a data warehouse and linking the report with each other after will help the management and the other user a lot on a better decision making for the future. As the commercial department in our company is the most important department when comes to decision making on how we are going to develop our company in the coming months or years I will be focused mainly on this department for now and next we can add other departments in the same data warehouse based on their requirements. Retail companies in Kosovo as in other regional countries are growing fast and using the data to ensure we are on the right track is very important for the company, Using the multiple data sources and the

available technology that we have today can empower quick decision making and this will increase the efficiency. This are some of the important points that we can achieve based on the data that will be used:

- Accurate trend forecasting
- Design Effective Marketing Strategies
- Improve Supply Chain Efficiency

## 4.6 Requirements

There are two kinds of requirements: functional and non-functional requirements.

Functional requirements define the system behaviour. They essentially define what the system dose or must do. Functional requirements usually define if & then behaviours and include calculations, data input, and business processes. Functional requirements are features that allow the system to function as it was intended. Put another way, if the functional requirements are not met, the system will not work. Functional requirements are product features and focus on user requirements

In the other hand, non-functional requirements specify how the system should do it. Non-functional requirements do not affect the basic functionality of the system (hence the name, non-functional requirements). Even if the non-functional requirements are not met, the system will still perform its basic purpose. Non-functional requirements define system behaviour and other characteristics that will affect the user experience while working on it. Based on the non-functional requirements we are going to know how easy the system to use is.

We formulated a questioner and did interweaves with all managers and other users who work in our department in order to gather the requirements from them. As this is an academic project, the requirements will be listed below. These requirements are done based on the questioners and interviews together. Then those requirements are validate based on the data sources that we have in order to see if we can achieve the results that we like to achieve.

Functional Requirements:



- The users can access the data in the company using Microsoft Excel or BI tools.
- The users should be able to analyse the sales based on stores, products and those should be separated based on months and years.
- The user should be able to see the sales based on the type of sales they can be Retail, Wholesale, Online, Export or Service.
- The user should be able to see the margin in percentage and profit in Euros based on the sales per store, month, year and day.
- The user should be able to see the sales based on Brands and Main Brands.
- The user should be able to see the sales based on the product categories.
- The user should be able to see the sales and stock based on price range.
- The user should be able to see the sales, stock based on products, categories, stores.
- The user should be able to see the sales based on customers and customer types.
- The users should be able to see the imported products based on categories and products per supplier and compare them with years before.
- The user should be able to see the charts where will be displayed sales, stock or import/exported values based on products, categories, stores, suppliers etc.
- The user should be able to see the sales based on the geographical locations. This is going to be done based on the client cities or store cities.

#### Non-Functional Requirements:

- The system should be based on windows authentication in order to save the data integrity. Each user has his own PC and account based on that they can only see the data after they are logged in in PC.
- The applications the user uses will be software as it is Excel and web-based app, which will be the BI app.
- The data will be accessible via internet and public folders.
- The system should perform well in order to not have latency and also should not crash the reports in the middle of the work because this will cause data lose.
- The system should be updated daily every 24h.

User requirements:

- Ability to drill down in reports to get more information.
- Ability to extract reports in Excel or PDF.
- Ability to get data from data warehouse and do other modifications in order to get the results that we like.
- Ability to get in the system and get data all the time while the user is inside work facilities.

System Requirements:

For this academic project, we will use Microsoft SQL, which is a product from Microsoft Company. SQL is one of the market leaders on the database technology. SQL is relational database management system that supports a number of other applications, including BI, transaction processing and analytics. In SQL, we will be storing all the data that we need for our data warehouse. There you can store data from different sources or other databases using another Microsoft tool called Integration Services (SSIS) with Visual Studio. We will also use Microsoft Excel and a power BI tool to do the reports. The BI tool can be Power BI, which is a Microsoft Product or another free tool that we can use for the purpose of this project. The data can be extracted using queries based on the user requirements or using the cube from Analyses Services (SSAS).

#### 4.7 Fact and Dimensions Tables

Dimension tables are the foundation of powerful analytics. In a star schema, a dimension table contains columns representing dimensions (Star Schema – Christopher Adamson, p:10). Dimensional tables serve to provide information to the fact tables. In queries or reports, dimensional table is usually used to specify how the facts are rolled up. Dimensional values are usually used to filter in the reports. They will be used to provide context for each measurement, usually in the form of textual labels that precede facts on each row of a report. They may also be used to drive master-detail relationships, subtotalling, cross-tabulation, or sorts (Star Schema – Christopher Adamson, p:10).

At the core of a star schema is the fact table. In addition to presenting the facts, the fact table includes surrogate keys that refer to each of the associated dimension tables (Star Schema – Christopher Adamson, p:12). Fact tables contains the measurements that are associated with a specific business process.

In this project, we are going to design three data marts that are going to be separated like this: sales, stock and Import & Export. According to Ralph Kimball, data marts represent a unit or departmental process within an organization. A data mart is a simple form of data warehouse focused on a single subject or line of business.

#### 4.8 Data Marts

In this section, we are going to design the three data marts that we will have in our data warehouse which are going to be: Sales, Import & Export and Stock.

Let us start with the first data mart, which will be sales data mart. The business events for this mart you can see in the image below:

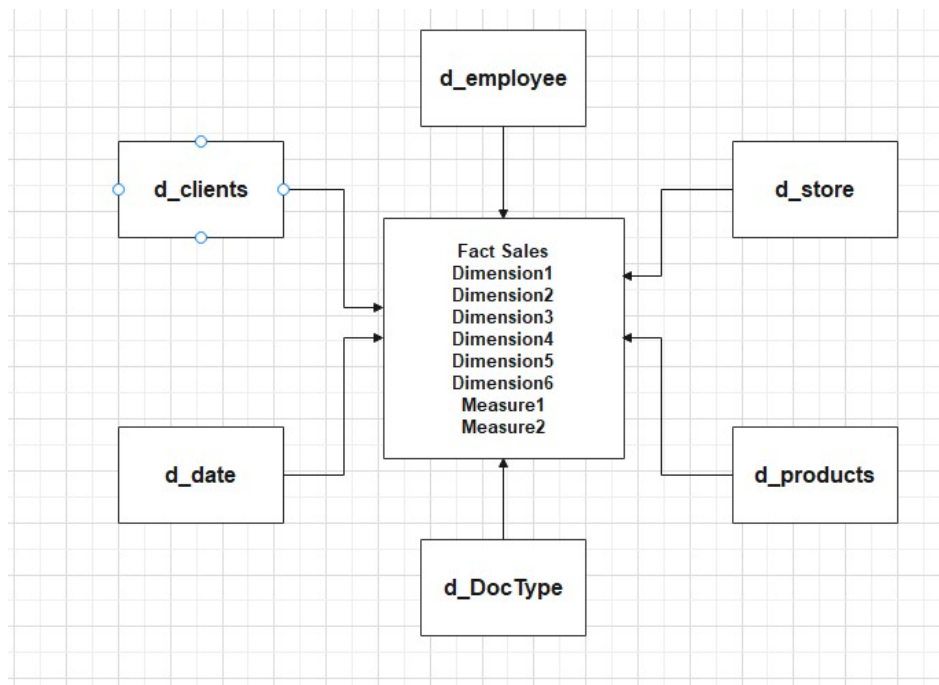


Figure 19 Star schema for Sales Data Mart

As you can see in the image the dimension tables and the fact table for sales data mart are listed. The sales data mart is populated by six-dimension tables: clients, employees, store, products, doctype and date. Some of the data in this tables are going to be from the source and the others are going to be calculated in the database based on the information that we need. All dimensional tables will have a primary key which will be generated automatically by auto-increment command while designing the tables in SQL. There will also be a natural key for some of the tables. This primary keys in dimensional tables will be presented as foreign keys in fact table. So, the link between fact and dimensional tables is done via referential integrity which in this case is foreign key.

Dimensional tables usually contain the attributes that explain the business processes that are places in the fact tables. Those attributes explain the grid that the tables will have and the information that we will gate using them. Those attributes can change. Example a product name can change based on the supplier income. So, we have to decide if we are going to track those changes or not. Based on what we need we will decide if we are going to keep track on the changes, SCD (Slowly Changing Dimension) are the dimension that have changed so we have to decide which type we will use. There are six types of SCD but four of them are the most important:

1. Type 0 – Fixed Dimension
2. Type 1 – No History
3. Type 2 – Row Versioning
4. Type 3 – Previous Value column

In the figure below you can see the full sales data mart with all the attributes in dimension and fact table:

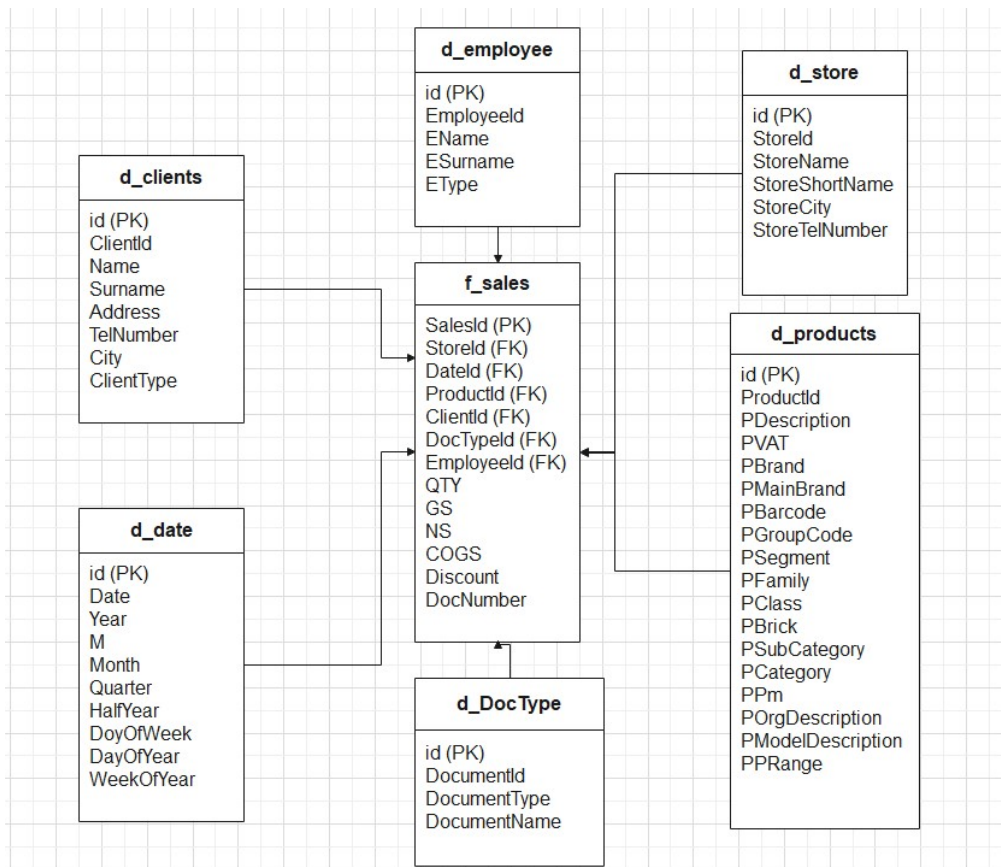


Figure 20 Logical Model for Sales Fact Table

The second data mart in which we will work now is the stock data mart. The star schema dimensional model for this process is below:

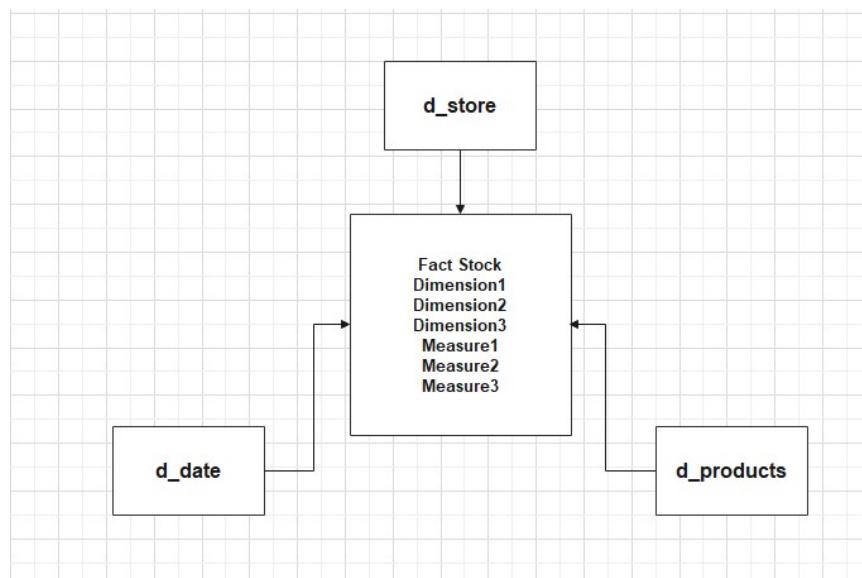


Figure 21 Star schema for Stock Data Mart

In the next image, we are going to see the designed stock data mart with all the attributes in dimension and fact table:

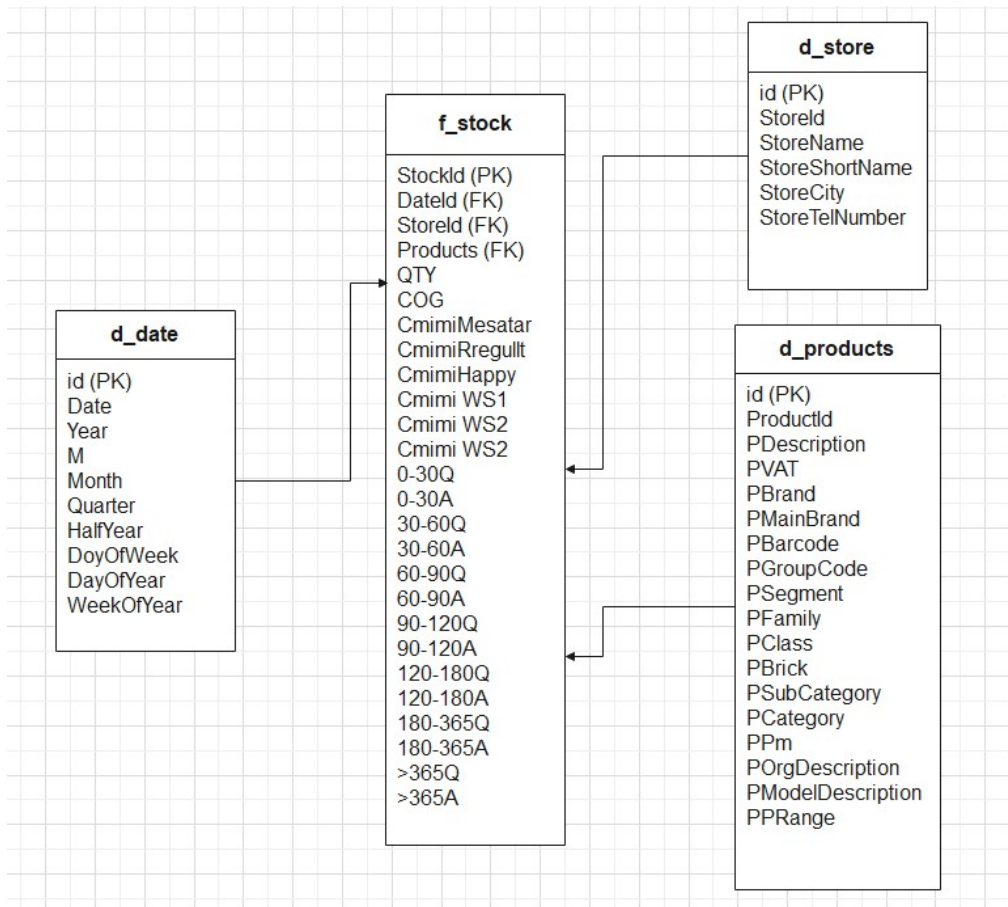
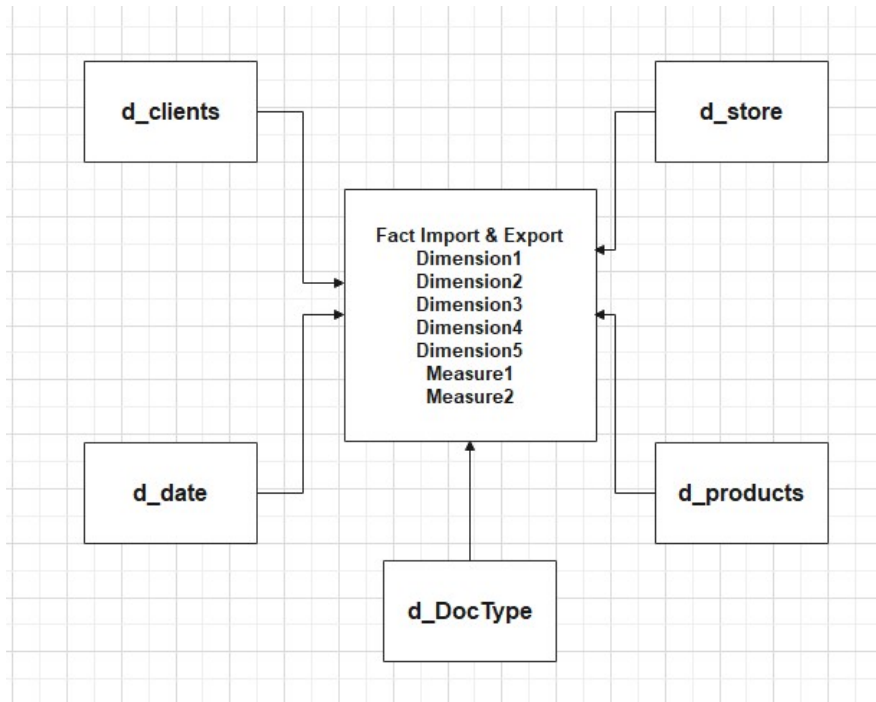


Figure 22 Logical Model for Stock Fact Table

Third and the last data mart that we are going to design for this data warehouse is import & export data mart. The star schema dimensional model for this process is below:



*Figure 23 Star schema for Import & Export Data Mart*

In the next image, we are going to see the designed Import & Export data mart with all the attributes in dimension and fact table:

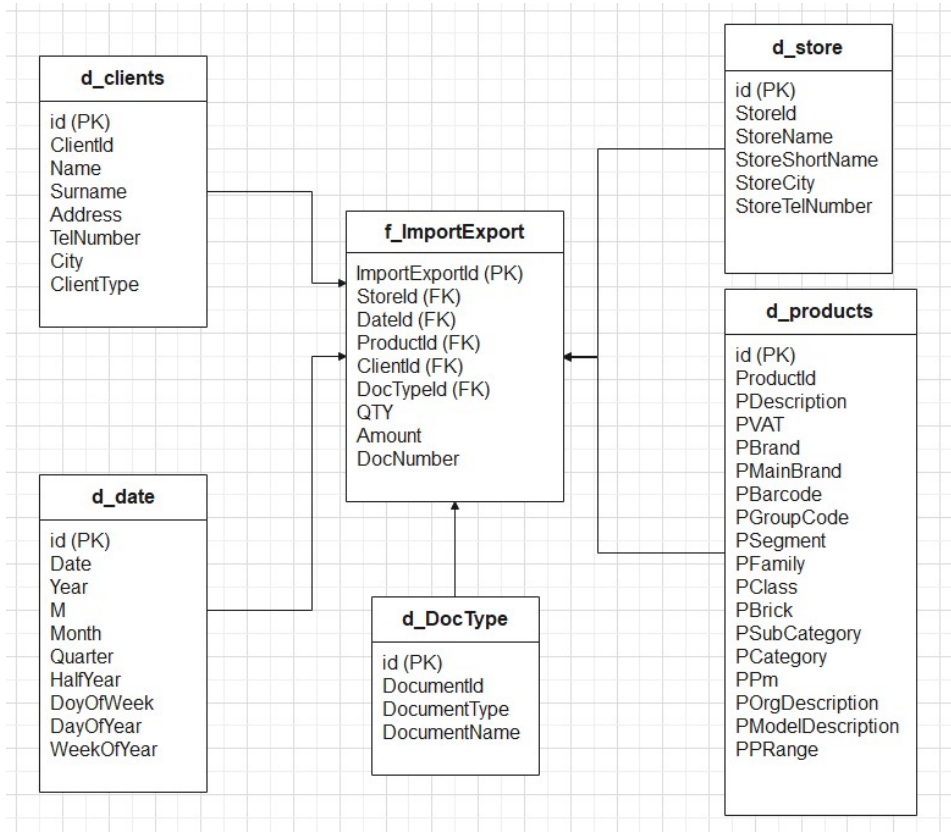


Figure 24 Logical Model for Import & Export Fact Table

#### 4.9 Physical Database

After having designed the logical database both for dimensional and fact tables now we are going to put these tables in SQL Server and use them in reality. The database is going to be designed in Microsoft SQL Management Studio (Express Edition) which is an open-source edition so we can use it for our project.

As we pointed out in our logical database design, we are going to work with three data marts which will separate three different processes but they will connect with each other to achieve the reporting requests that we got from the users. All data marts as described in the logical design will include dimension and fact tables. Each dimension table will have a primary key which will be auto-incremented and a natural key which will be extracted from the data sources. All the primary keys will link the dimension and fact tables.

In the image below, we will see the database and the tables created in SQL Server



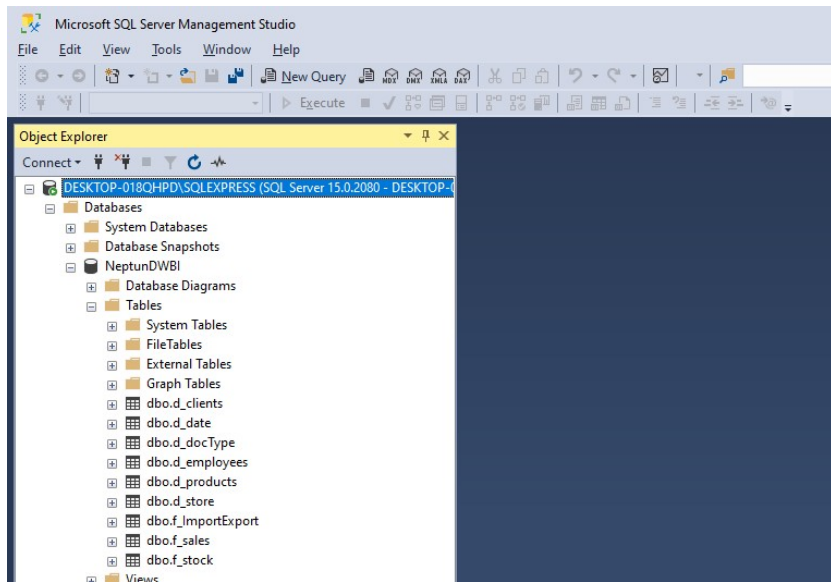


Figure 25 Database in SQL Server Management Studio

The physical design of the data marts are as follows:

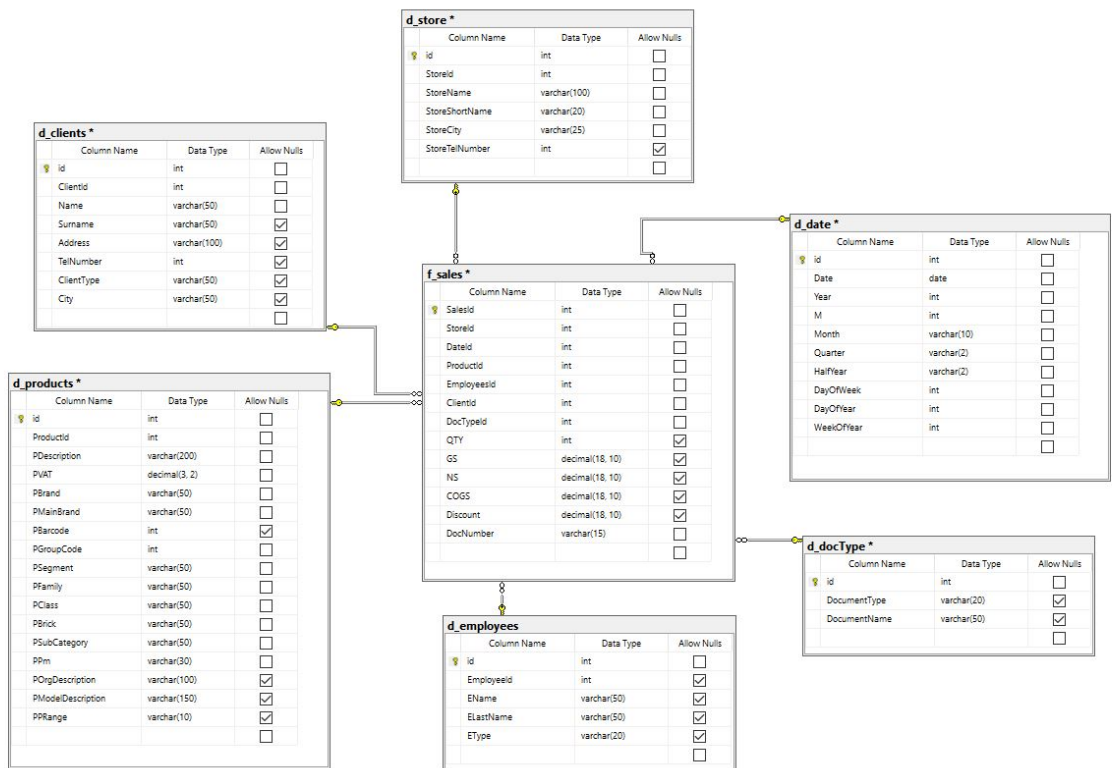


Figure 26 Physical Design for Sales Data Mart

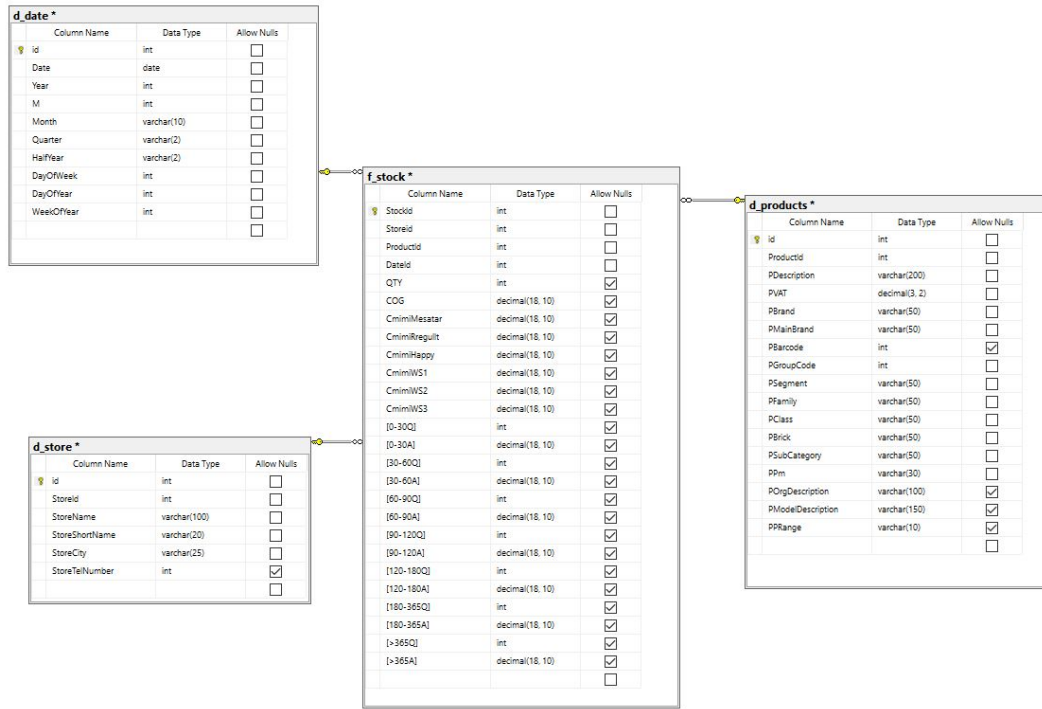


Figure 27 Physical Design for Stock Data Mart

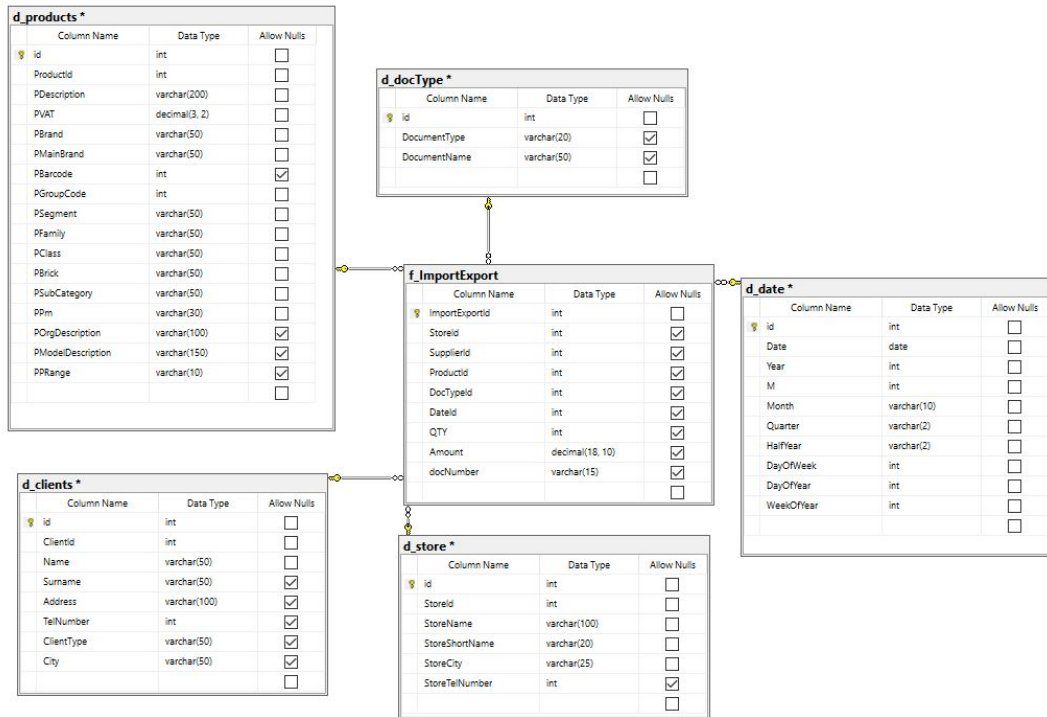


Figure 28 Physical Design for Import & Export Data Mart

## 4.10 ETL Process

ETL process is one of the most important processes in a project like Data Warehouse. After we successfully completed the design of the table and create them in the SQL Server now is the time to populate those tables. Dimensional and Fact tables must be populated with the right data in order for the to be populated without any problems and what's more important form those tables we will generate reports that must have legit data.

Extract, transform and load process is very complex and its development is a big challenge for the data warehouse developers. The way that this process is designed is influenced by the tools used in the process and the architecture of the data warehouse. The data that will be loaded are the data that are required in order to generate the reports needed. Those requirements were gathered through interviews with the employees or managers that are going to use the reports in their everyday work.

The ETL process must work with different source systems, based on heterogeneous technologies, to create a single integrated view of information. It must do this with data that is often of dubious quality. It must function within tightly prescribed processing windows and have minimal impact on operational systems. It must be fully automated, able to process exceptions, and able to recover from errors (Star Schema – Christopher Adamson, p:404).

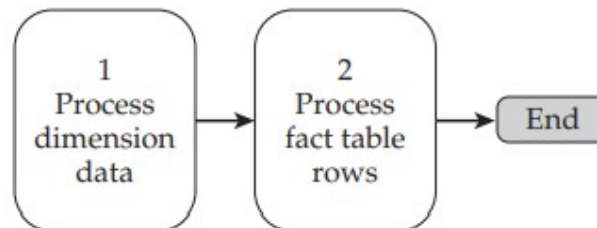
The process of loading the dimension tables is made complex in the need of handling the slowly changing dimension. The other cases are not so complex all you have to do is match the data type and load the tables with them. In the other hand the process of loading fact tables has other complex issues like handling the foreign keys. As supplied by a source system, transactions will be accompanied by natural keys. The load process must replace these with surrogate keys (Star Schema – Christopher Adamson, p:403).

The process of loading the tables is usually under time pressure so it must be done in the most efficient and fastest way. Optimizing the process of loading the data is one of the most important things to do in the process of loading the data but this process is very difficult. The bast way of having better results in this process is to identify the data changes before they are loaded in the tables. There are many data cleansing tasks

that can be automated in the load process but for some data, it is better to do the cleansing before loading them.

The ETL process is a coordinated effort between numerous tools and utilities that are configured and assembled to load the star schema as required by its design documentation (Star Schema – Christopher Adamson, p:404).

The process of loading a star schema data warehouse involves two major activities: processing the data for dimensional tables and fact tables. Loading the dimensional tables is made complex by the data changes that must be identified or if you have the data updated in your source tables which is in a file outside the data warehouse you must delete all the data from the table and load them again with the changes made this is a manual way of handling the changes. In the image below you can see how the process of loading of star schema data warehouse goes:



*Figure 29 loading a star*

The process of loading a dimensional table normally goes in the way that you can see in the image below:

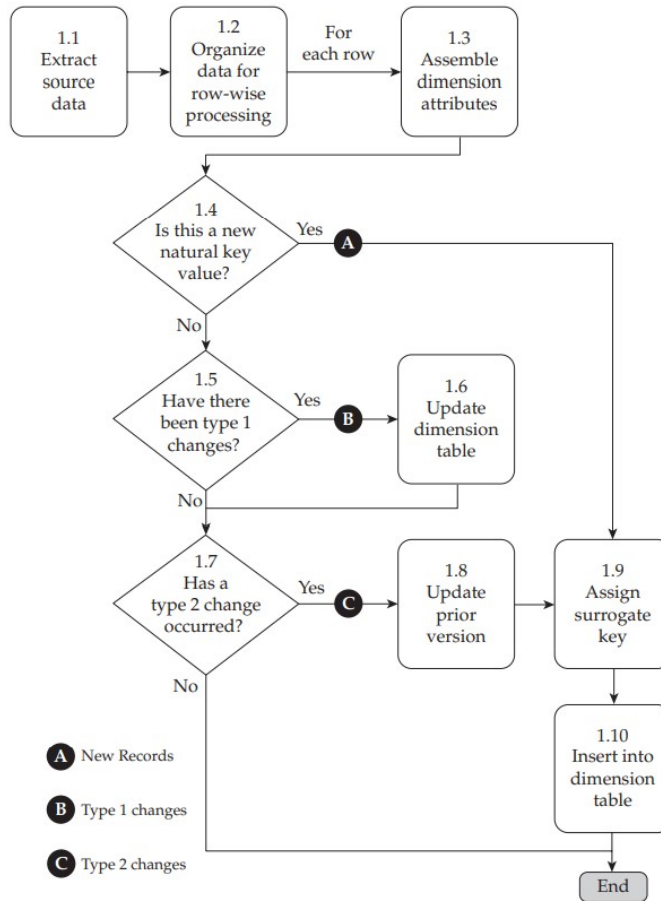


Figure 30 Loading a dimensional table

The process of loading dimensional tables described in steps:

- Extract source data from source(s)
- Assemble the dimensional attributes
- Identify new and changed dimension records
- Manage surrogate keys
- Process new records
- Process type 1 changes
- Process type 2 changes

The most difficult work on loading the data is usually done in the dimensional tables. The process of loading the fact tables is simpler as described in the book Star Schema from Christopher Adamson these are the steps taken in the moment that the data are loaded in the fact tables:

- Extract data from the source systems

- Compute the facts
- Aggregate facts to match the grain of the fact table
- Obtain surrogate keys for each of the dimensions
- Load fact table records into the warehouse

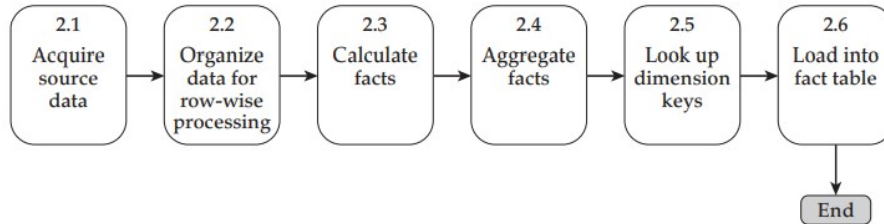


Figure 31 Loading a fact table

#### 4.11 ETL and SSIS (SQL Server integration Services)

In this chapter, we are going to describe step by step the process of loading the tables using SSIS or SQL Server Integration Services. After we finished the design of the tables and created them in the SQL Server now it is time to load the data, using the data sources, which in our project will be Excel Files.

##### 4.11.1 Dimensional Tables

In the image below, we are going to see all the dimensional tables that are going to be loaded using the SSIS:

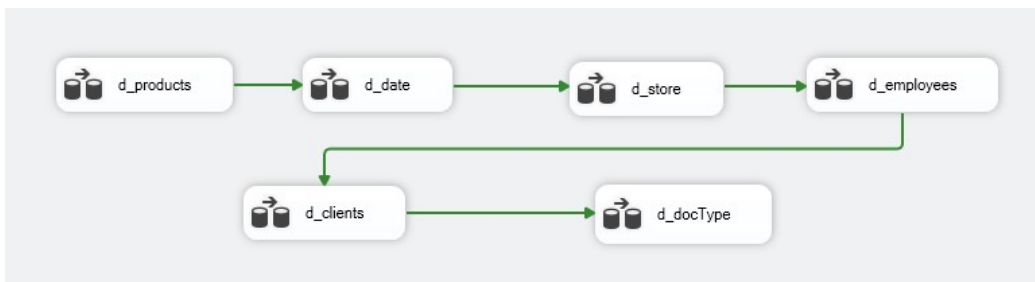


Figure 32 Dimensional Tables

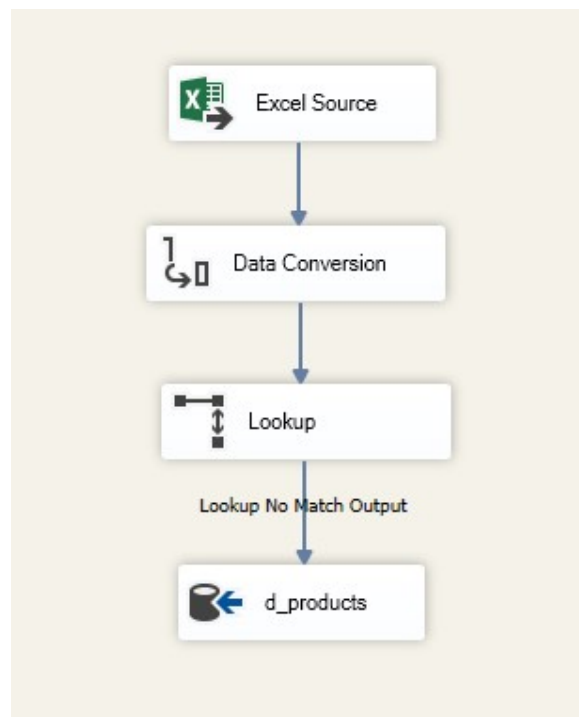
There are six dimensional tables in our data warehouse:

1. D\_products

2. D\_date
3. D\_store
4. D\_employees
5. D\_clients
6. D\_doctype

The sources of this table are Excel files, which in this case are data that we could get from my company “Neptun” and manipulate them in order to use those data as examples for our case study.

The first dimensional table that is going to be loaded is d\_products:



*Figure 33 Products table*

As you can see we have the data source which is an excel file d\_products.xls which is located in the servers. There are data types that will be modified using the Data Conversion because the SQL Server data types and SSIS data types do not match in all cases. Then we have used lookup in order to check for the new data. After the full table is loaded, we do not have to load the table every time for that we use lookup in order to load just the new record that are added in our dimensional tables. In the image below you can see the data loaded in the data warehouse:

SQLQuery1.sql - DE...18QHPD\artan (64) \* - X

```
select * from d_products;
```

0 %

Results Messages

	id	ProductId	PDescription	PVAT	PBrand	PMainBrand	PBarcode	PGroupCode	PSegment	PFamily	PClass	PBrck	PSubCate
1	169314	2925104	Blender Fuego FG-107	1.18	FUEGO	FUEGO	0	364110	H/A	SDA	FOOD PREPARATION	BLENDERS	0
2	169315	2981033	Pegla Beko BKK-2133	1.18	BEKO	ARCELUK	0	365200	H/A	SDA	LAUNDRY	IRONS	0
3	169316	2981034	MICRO SD	1.18	APACER	OTHER	0	615231	OTHER	ACCESSORIES	PHOTO ACC	MEMORY CARDS	<8GB
4	169317	2981035	PEN DRIVE	1.18	APACER	OTHER	0	512631	IT	COMPUTER PARTS	COMPUTER INPUT DEVICES	STORAGE	<8GB
5	169318	2981036	MOUSE LASER	1.08	GENIUS	OTHER	0	512510	IT	COMPUTER PARTS	COMPUTER INPUT DEVICES	PC MOUSES	0
6	169319	2981037	LG MONITOR 23	1.18	LG	LG	0	642100	OTHER	OTHER	OTHER	OTHER	0
7	169320	2981038	LG MONITOR 23	1.18	LG	LG	0	642100	OTHER	OTHER	OTHER	OTHER	0
8	169321	2981039	LG MONITOR 23	1.18	LG	LG	2030072011954	513313	IT	COMPUTER PARTS	COMPUTER PERIPHERALS	PC MONITORS	24"
9	169322	2981040	LG MONITOR 23	1.18	LG	LG	0	642100	OTHER	OTHER	OTHER	OTHER	0
10	169323	2981041	BEKO REF RBI 6300	1.18	BEKO	ARCELUK	0	331120	H/A	REF	REF	BUILT-IN REF	0
11	169324	2981042	BEKO REF RBI 6300	1.18	BEKO	ARCELUK	0	331120	H/A	REF	REF	BUILT-IN REF	0
12	169325	2981043	BE	1.18	OTHER	OTHER	0	642100	OTHER	OTHER	OTHER	OTHER	0
13	169326	2981044	BEKO REF RBI 6300	1.18	BEKO	ARCELUK	0	331120	H/A	REF	REF	BUILT-IN REF	0
14	169327	2981045	SAMSUNG LCD 32B350	1.18	SAMSUNG	SAMSUNG	0	162311	A/V	TV	LCD	30"-39"	HD
15	169328	2981046	Keyboard black De Lux	1.08	DELUX	OTHER	0	512300	IT	COMPUTER PARTS	COMPUTER INPUT DEVICES	PC KEYBOARDS	0
16	169329	2981047	LG LCD 37LH3000	1.18	LG	LG	8808992479369	162323	A/V	TV	LCD	30"-39"	FHD
17	169330	2981048	SAMSUNG LCD LE40B530	1.18	SAMSUNG	SAMSUNG	0	162411	A/V	TV	LCD	40"-49"	HD
18	169331	6	LG A/C ES-H 186 KLAO	1.18	LG	LG	2030072009302	413400	HVAC	A/C	SPLIT SYSTEM	18000 BTU	0

Figure 34 Products table results

The second table will be d\_date:

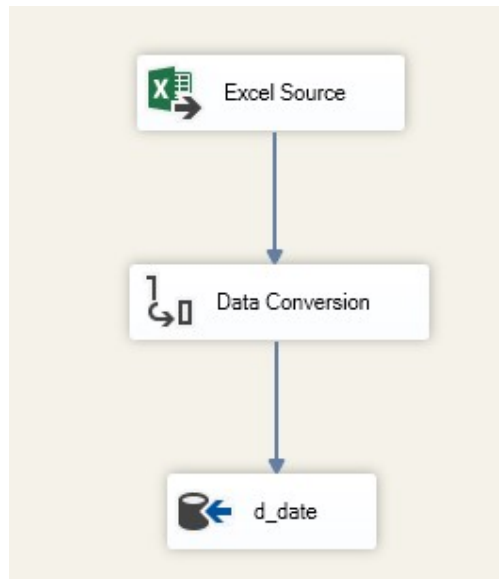


Figure 35 Date table

Here are the results after the table is loaded:



SQLQuery1.sql - DE...18QHPD\artan (64)\* - X

```
select * from d_date;
```

100 %

Results Messages

	id	Date	Year	M	Month	Quarter	HalfYear	DayOfWeek	DayOfYear	WeekOfYear
1	80529	2020-01-01	2020	1	January	Q1	H1	3	1	1
2	80530	2020-01-02	2020	1	January	Q1	H1	4	2	1
3	80531	2020-01-03	2020	1	January	Q1	H1	5	3	1
4	80532	2020-01-04	2020	1	January	Q1	H1	6	4	1
5	80533	2020-01-05	2020	1	January	Q1	H1	7	5	1
6	80534	2020-01-06	2020	1	January	Q1	H1	1	6	2
7	80535	2020-01-07	2020	1	January	Q1	H1	2	7	2
8	80536	2020-01-08	2020	1	January	Q1	H1	3	8	2
9	80537	2020-01-09	2020	1	January	Q1	H1	4	9	2
10	80538	2020-01-10	2020	1	January	Q1	H1	5	10	2
11	80539	2020-01-11	2020	1	January	Q1	H1	6	11	2
12	80540	2020-01-12	2020	1	January	Q1	H1	7	12	2
13	80541	2020-01-13	2020	1	January	Q1	H1	1	13	3
14	80542	2020-01-14	2020	1	January	Q1	H1	2	14	3
15	80543	2020-01-15	2020	1	January	Q1	H1	3	15	3
16	80544	2020-01-16	2020	1	January	Q1	H1	4	16	3
17	80545	2020-01-17	2020	1	January	Q1	H1	5	17	3
18	80546	2020-01-18	2020	1	January	Q1	H1	6	18	3
19	80547	2020-01-19	2020	1	January	Q1	H1	7	19	3

Figure 36 Date table results

Next table is d\_store:

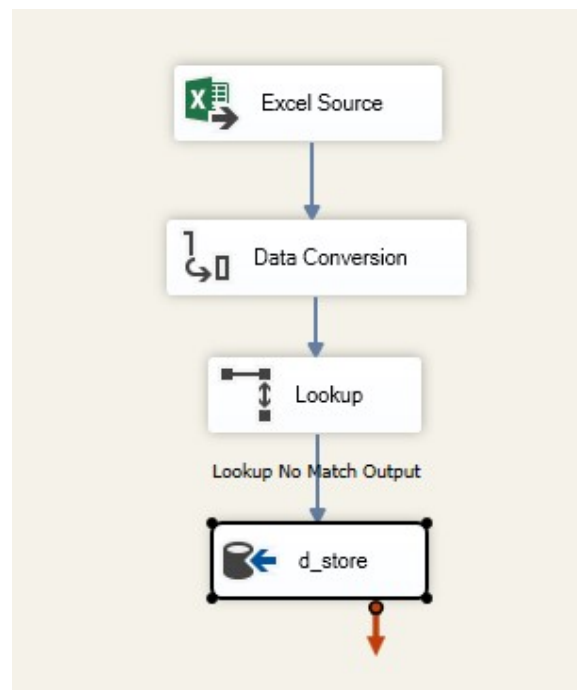


Figure 37 Store table

Here are the results after the table is loaded:

SQLQuery1.sql - DE...18QHPD\artan (64))\*

```
Select * from d_store;
```

100 %

Results Messages

	id	StoreId	StoreName	StoreShortName	StoreCity	StoreTelNumber
1	93	101	Shitorja 1 - Albi	B.Albi	Pristine	44111222
2	94	102	Shitorja 2 - Prizren	Prizren	Prizren	44111223
3	95	107	Shitorja 7 - Fushe Kosove	A.Fushë Kosovë	Pristine	44111224
4	96	105	Shitorja 5 - Podujeve	N.Podujevë	Podujeve	44111225
5	97	106	Shitorja 6 - Prizren 2	D.Prizren	Prizren	44111226
6	98	103	Shitorja 3 - Peje	F.Pejë	Peje	44111227
7	99	108	Shitorja 8 - Ferizaj	E.Ferizaj	Ferizaj	44111228
8	100	109	Shitorja 9 - Gjilan	G.Gjilan	Gjilan	44111229
9	101	115	Shitorja 15 - Mitrovice	H.Mitrovicë	Mitrovice	44111230
10	102	110	Shitorja 10 - Istog	P.Istog	Istog	44111231

Figure 38 Store table results

Next table is d\_employees:

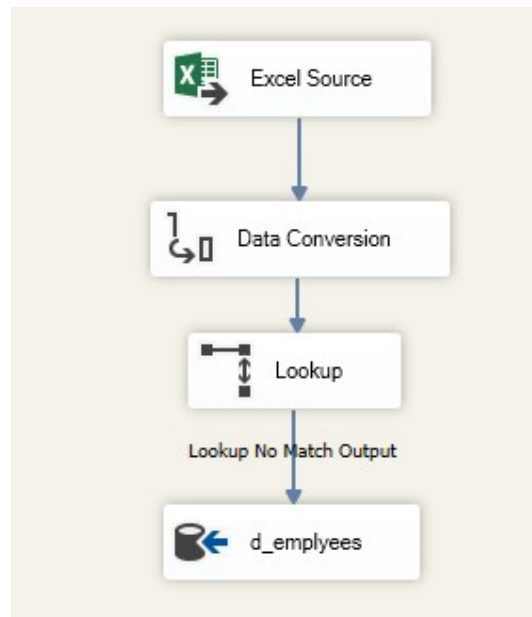


Figure 39 Employees table

Here are the results:

SQLQuery1.sql - DE...18QHPD\artan (64)\*

```
Select * from d_employees;
```

100 %

Results Messages

	id	EmployeeId	EName	ELastName	EType
1	16	101	EmplooyA	EmplooyA	Sales
2	17	102	EmplooyB	EmplooyB	Manager
3	18	103	EmplooyC	EmplooyC	Sales
4	19	701	EmplooyA	EmplooyA	Sales
5	20	702	EmplooyB	EmplooyB	Manager
6	21	703	EmplooyC	EmplooyC	Sales
7	22	501	EmplooyA	EmplooyA	Sales
8	23	502	EmplooyB	EmplooyB	Manager
9	24	503	EmplooyC	EmplooyC	Sales
10	25	601	EmplooyA	EmplooyA	Sales
11	26	602	EmplooyB	EmplooyB	Manager

Figure 40 Employees table results

Now let us continue with d\_clients:

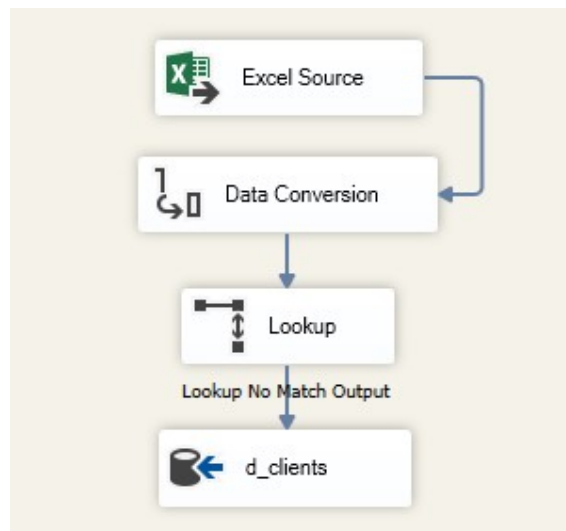


Figure 41 Clients table

Here are the results for d\_clients tables:

SQLQuery1.sql - DE...18QHPD\artan

```
select * from d_client
```

100 %

Results Messages

	id	ClientId	Name
1	109854	95054	Client1
2	109855	74202	Client2
3	109856	91891	Client3
4	109857	18049	Client4
5	109858	12270	Client5

Figure 42 Clients tables results

Moreover, the last dimensional table is d\_docType:

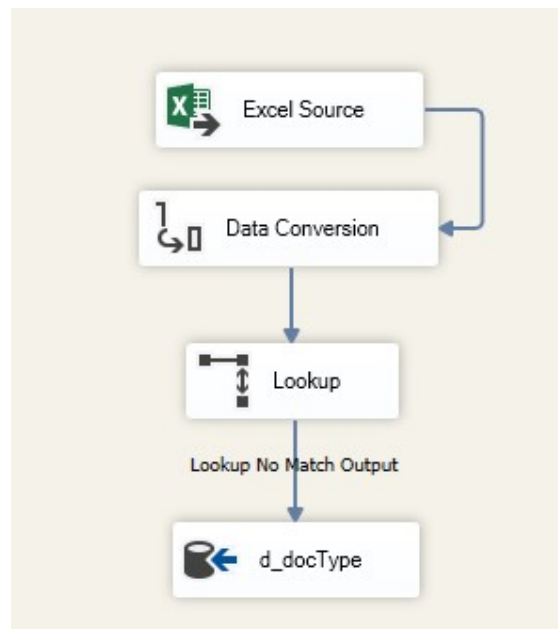


Figure 43 DocType table

In addition, this is the result after we loaded the last dimensional table:

	id	DocumentType	DocumentName	DocumentId
1	1631	Retail	Fature shitje malli nga dyqani	1100
2	1632	WS	Fature shitje malli	2200
3	1633	WS	Fature shitje malli Kesh	2210
4	1634	Export	Fature- me jashte shtetit	3300
5	1635	Retail	Kupon tatimor	1110
6	1636	Retail	Credit Note-(Dok.jone- brenda vendit)	1120
7	1637	SVC	Fature shitje malli nga servisi	4400
8	1638	Retail	Fature shitje malli - sindikata	1130

Figure 44 DocType table results

#### 4.11.2 Fact tables

After successful loading of the dimension tables, it is now time to load the fact table which is the last step in the process of data warehouse loading. The data that are stored in the dimensional tables are all high-quality data and that will make the process of loading the fact tables easier. This data all connected with business intelligence will give the company the best results when it comes to data gathering from the processes that will happen.

We have three fact tables that are going to be loaded:

1. F\_sales
2. F\_stock
3. F\_ImpExp

Fact table's sources for our project are also Excel files, which are located in our servers. Let us start with the load of first fact table f\_sales:

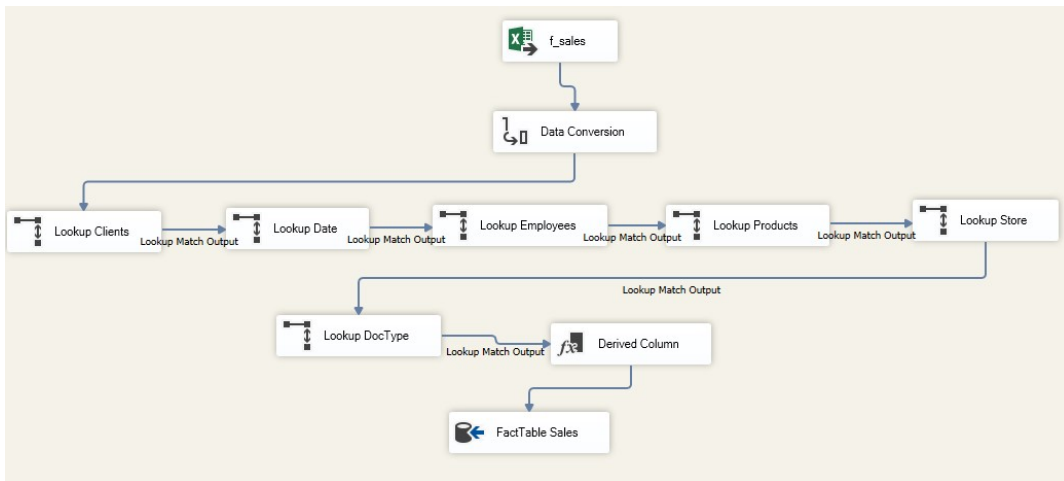


Figure 45 Sales fact table

Here are the results after the data are loaded:

SQLQuery1.sql - DES...n (64) Executing...\*

```
select * from f_sales;
```

	SalesId	StoreId	DateId	ProductId	EmployeesId	ClientId	QTY	GS	NS	COGS	Discount	DocNumber	DocTypeId
1	776599	110	82022	182252	55	119078	1	260.0000000015	220.3389830521	184.9998910000	0.0000000000	04-401/34292	1632
2	776600	110	82022	182252	55	119078	1	270.0000000037	228.8135593251	184.9998910000	0.0000000000	04-401/34291	1632
3	776601	110	82022	182311	55	178379	1	319.9999520000	271.1864000000	214.9802200000	0.0000000000	04-401/34290	1632
4	776602	110	82022	181963	55	178379	1	249.9999920000	211.8644000000	147.7499920000	0.0000000000	04-401/34290	1632
5	776603	110	82022	182252	55	161850	20	5599.9991600000	4745.7620000000	3699.9978000000	0.0000000000	04-401/34289	1632
6	776604	110	82022	182517	55	161850	2	469.9999000000	398.3050000000	340.9387760000	0.0000000000	04-401/34289	1632
7	776605	110	82022	183102	55	161850	2	460.0001080000	389.8306000000	315.2000000000	0.0000000000	04-401/34289	1632
8	776606	110	82022	175997	55	161850	10	2099.9999800000	1779.6610000000	1297.7239900000	0.0000000000	04-401/34289	1632
9	776607	110	82022	175827	55	161850	10	1800.0000002050	1525.4237288347	1293.2341700000	0.0000000000	04-401/34289	1632
10	776608	110	82022	170045	55	161850	10	1499.9999995900	1271.1864406432	1029.9592000000	0.0000000000	04-401/34289	1632
11	776609	110	82022	182978	55	231558	2	489.9999560000	415.2542000000	350.0000000000	0.0000000000	04-401/34288	1632

Figure 46 Sales fact table results

Now let us see the second fact table f\_stock:

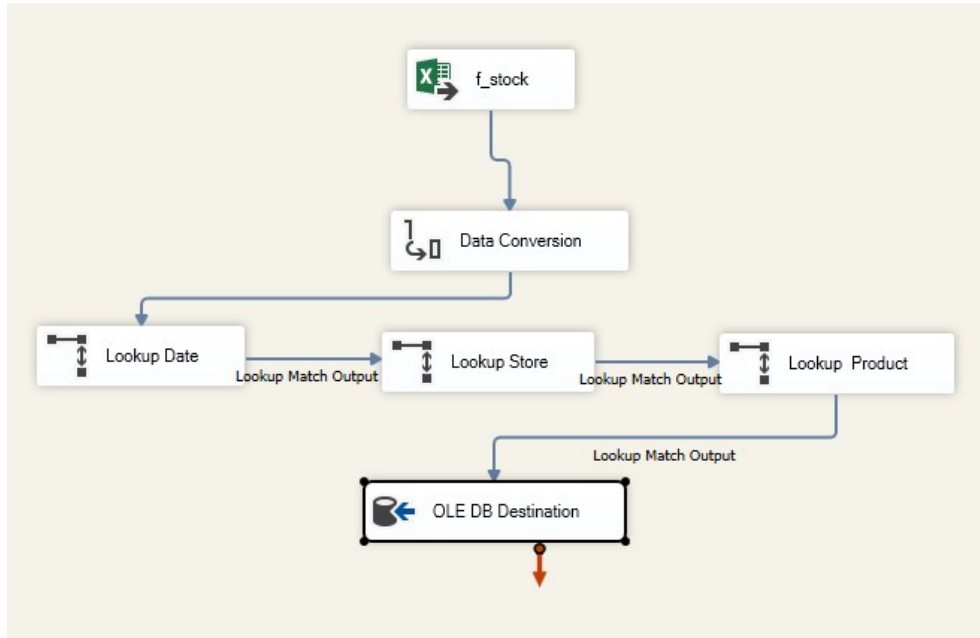


Figure 47 Stock fact table

Here are the results after the data are loaded:

SQLQuery1.sql - DE...18QHPD\artan (64)\* - X

```
select * from f_stock;
```

100 %

	StockId	StoreId	ProductId	DateId	QTY	COG	CmimiMesatar	CmimiRegult	CmimiHappy	CmimiWS1	CmimiWS2	CmimiWS3	0-30Q
1	40405	93	181518	80807	3	86.7100000000	28.9033333333	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
2	40406	93	181519	80807	326	12376.8700000000	37.9658588957	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
3	40407	93	181520	80807	7	183.7800000000	26.2542857142	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
4	40408	93	181521	80807	111	11441.5700000000	103.0772072072	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
5	40409	93	181522	80807	1	40.8500000000	40.8500000000	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
6	40410	93	181523	80807	171	999.3300000000	5.8440350877	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
7	40411	93	181524	80807	34	1308.1900000000	38.4761764705	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
8	40412	93	181525	80807	82	5889.9300000000	71.8284146341	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
9	40413	93	181526	80807	88	11096.1800000000	126.0929545454	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
10	40414	93	179212	80807	70	7419.2500000000	105.9892857142	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
11	40415	93	179211	80807	847	100862.8700000000	119.0824911452	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
12	40416	93	179267	80807	1	18.4200000000	18.4200000000	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0
13	40417	93	179264	80807	1	7.4000000000	7.4000000000	1.0000000000	2.0000000000	3.0000000000	4.0000000000	5.0000000000	0

Figure 48 Stock fact table results

Moreover, the last fact table is f\_ImpExp:



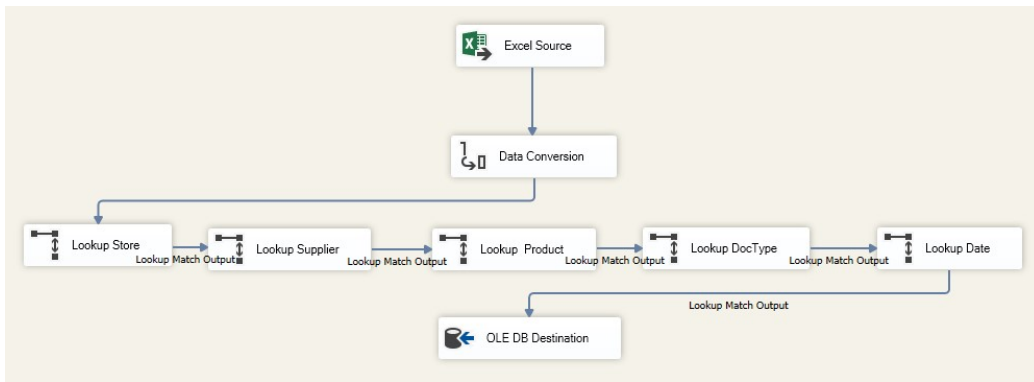


Figure 49 ImpExp fact table

The results after we successfully loaded the last table in the data warehouse:

SQLQuery1.sql - DE...18QHPD\artan (64)\*

```
select * from f_ImportExport;
```

	ImportExportId	StoreId	SupplierId	ProductId	DocTypeId	DateId	QTY	Amount	docNumber
1	10705	110	209075	181661	1647	81972	2	38.9830508474	04-202/1949
2	10706	110	179597	180457	1647	81972	1	10.3389830508	04-202/1948
3	10707	110	179597	180456	1647	81972	1	8.4745762720	04-202/1948
4	10708	110	209075	181661	1647	81995	1	19.4915254237	04-202/1964
5	10709	110	179597	183107	1647	81995	2	23.7288135594	04-202/1963
6	10710	110	179597	180459	1647	81995	14	172.0338983048	04-202/1963
7	10711	110	179597	180457	1647	81995	1	10.3389830508	04-202/1963
8	10712	110	179597	180456	1647	81995	3	25.4237288160	04-202/1963
9	10713	110	179597	180443	1647	81995	2	39.8305084746	04-202/1963
10	10714	110	179597	180442	1647	81995	7	139.4067796611	04-202/1963
11	10715	110	179597	180733	1647	81995	7	130.5084745762	04-202/1963
12	10716	110	179597	180731	1647	81995	2	8.9830508542	04-202/1963

Figure 50 ImpExp fact table results

#### 4.12 Creating views and reports

Now after the ETL process is finished the aim is to create some reports that are requested from the management in order to fully complete the project and see the results in real life. These reports will help the management and company on achieving better results and take better decisions for the future of the company.



Dimension tables as we know are connected with fact table via foreign keys so we do not have to make other relationships between them. The three fact tables are going to be linked with each other with attributes from the dimension tables.

One of the most powerful features of SQL is its capability to display views of the data that are structured differently from how the database tables stored the data (SQL for Dummies 8th Edition - Allen G. Taylor, p:141). You can manipulate a view as just as you can manipulate a real table. The difference is that a view does not have an independent relationship (SQL for Dummies 8th Edition - Allen G. Taylor, p:141). The view gets the data from one table or different tables and from them we can see those tables' columns. In our case, we will use SQL Server for creating the views. These views will be linked with Excel files and Power Bi in order to generate reports that we need.

As organizations continue to pursue their goals in an economy that seems more like a battlefield, it's no wonder that they, too, feel the need for reliable information based on real and readily usable data — business intelligence (Microsoft Business Intelligence for Dummies – Ken Withee, p:1). For this project, we are going to use Power Bi as a business intelligence tool. What is Power BI? Power BI is a collection of software's and other applications that work together to link our unrelated data sources and turn those results in data in visually immersive and interactive insights. The source tables can we form different applications. In our case, we will create views and connect those views with Power BI, after that we will share the results with the others.

Now let us create some views using the SQL Server and the data that we have loaded in our data warehouse using the SSIS tools. These views will be connected with Microsoft Excel and Power Bi for analysing the data. The first view that we created is about checking the Sales over the years in Neptun based on the Product Categories:

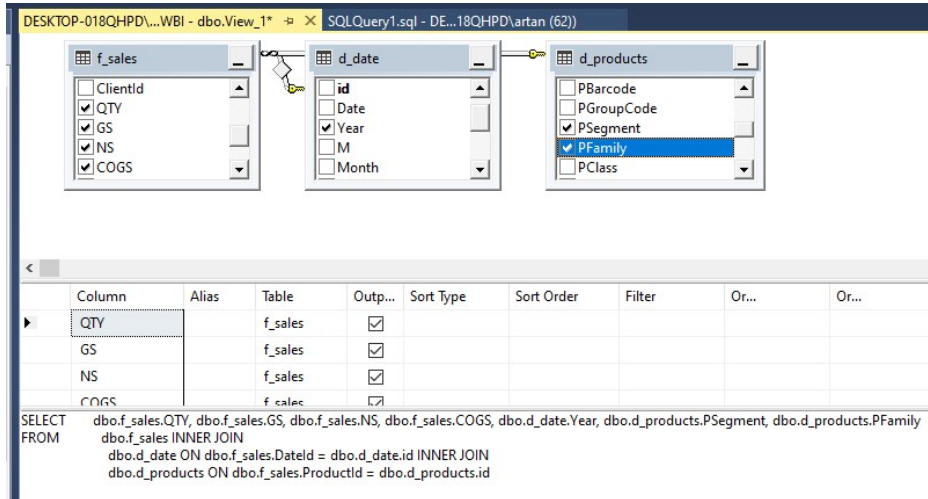


Figure 51 View of sales of products in categories

As you can see in the image, we have created a view by linking three different tables: f\_sales, d\_date and d\_products in order to generate the data that we need and we will name this view “Sales by Category”.

PSegmen	PFamily	Year					
		2019		2020		2021	
		QTY	GS	QTY	GS	QTY	GS
A/V	A/V ACCESSORIES	320	12,578 €	235	8,692 €	387	11,664 €
	HOME A/V	256	56,896 €	246	54,997 €	345	71,096 €
	IN-CAR ELECTRONICS	16	1,121 €	14	810 €	29	1,580 €
	PHOTO	22	6,555 €	14	4,252 €	13	3,286 €
	PORTABLE A/V	1,369	49,467 €	1,454	67,108 €	1,839	103,796 €
	TV	9,939	3,838,967 €	9,403	3,572,970 €	8,922	4,054,330 €
	<b>A/V Total</b>	<b>11,922</b>	<b>3,965,584 €</b>	<b>11,366</b>	<b>3,708,829 €</b>	<b>11,535</b>	<b>4,245,753 €</b>
COMM	MOBILE COMM	5,131	1,433,625 €	6,340	1,586,099 €	7,472	1,798,550 €
	<b>COMM Total</b>	<b>5,131</b>	<b>1,433,625 €</b>	<b>6,340</b>	<b>1,586,099 €</b>	<b>7,472</b>	<b>1,798,550 €</b>
H/A	COOKING	10,939	1,996,171 €	9,724	1,789,552 €	15,136	2,799,736 €
	D/W	4,582	1,260,635 €	4,872	1,266,295 €	7,367	1,947,735 €
	PERSONAL CARE	10,488	294,485 €	10,439	291,123 €	12,986	374,472 €
	REF	8,716	2,186,190 €	7,631	1,923,693 €	10,127	2,747,078 €
	SDA	19,061	921,235 €	19,702	919,742 €	21,087	1,058,985 €
	W/M	7,278	2,002,292 €	7,782	2,054,404 €	9,558	2,663,155 €
	<b>H/A Total</b>	<b>61,064</b>	<b>8,661,009 €</b>	<b>60,150</b>	<b>8,244,809 €</b>	<b>76,261</b>	<b>11,591,161 €</b>
HVAC	A/C	1,210	440,031 €	736	251,394 €	1,608	584,471 €
	HEATING	3,643	132,186 €	4,717	157,102 €	4,701	171,156 €
	<b>HVAC Total</b>	<b>4,853</b>	<b>572,217 €</b>	<b>5,453</b>	<b>408,496 €</b>	<b>6,309</b>	<b>755,627 €</b>
IT	COMPUTER PARTS	3,918	108,123 €	3,053	87,403 €	5,807	167,352 €
	PERSONAL COMPUTING	42	32,093 €	35	28,683 €	50	35,196 €
	PORTABLE COMPUTING	1,714	486,761 €	2,097	624,020 €	2,239	721,383 €
	<b>IT Total</b>	<b>5,674</b>	<b>626,976 €</b>	<b>5,185</b>	<b>740,105 €</b>	<b>8,096</b>	<b>923,931 €</b>
OTHER	ACCESSORIES	11,078	136,266 €	9,164	124,068 €	13,474	184,619 €
	ENTERTAINMENT	395	78,395 €	388	62,443 €	99	18,111 €
	HOME	2,943	40,611 €	3,354	72,863 €	2,419	53,838 €
	OUTDOOR			2	793 €	86	26,456 €
	<b>OTHER Total</b>	<b>14,416</b>	<b>255,272 €</b>	<b>12,908</b>	<b>260,167 €</b>	<b>16,078</b>	<b>283,024 €</b>
	<b>Grand Total</b>	<b>103,060</b>	<b>15,514,683 €</b>	<b>101,402</b>	<b>14,948,506 €</b>	<b>125,751</b>	<b>19,598,046 €</b>

Figure 52 Sales in categories over the years

The results of the view that we just created are showed in the Figure 52 and as we can see from the results, the company has performed much better on year 2021. These result and others are made easier with the data warehouse that we implemented in this project. Now let us move on and create some more views based on the management requests that we got over the interviews that we had with them to get their requests. Next report that is going to be generate will use another view called “Sales by Stores”:

Figure 53 View of Sales by Stores

Year	2021														
GS	Month														
StoreShortName	January	February	March	April	May	June	July	August	September	October	November	December	Grand Total		
A.Fushë Kosovë	206,216 €	176,673 €	223,773 €	358,510 €	231,185 €	255,343 €	329,790 €	299,985 €	323,544 €	352,557 €	351,497 €	387,878 €	3,496,949 €		
B.Albi	140,451 €	116,210 €	123,914 €	104,899 €	147,909 €	153,893 €	209,723 €	185,628 €	146,980 €	199,718 €	205,666 €	215,083 €	1,950,074 €		
C.Royal Mall	67,563 €	65,053 €	78,812 €	78,193 €	64,233 €	76,810 €	107,303 €	89,058 €	97,850 €	104,252 €	106,059 €	139,824 €	1,075,011 €		
D.Prizren	112,394 €	93,065 €	103,844 €	257,735 €	141,672 €	138,005 €	227,096 €	219,501 €	176,588 €	219,565 €	172,195 €	196,160 €	2,057,821 €		
E.Ferizaj	105,995 €	86,236 €	100,766 €	189,028 €	120,562 €	114,932 €	201,444 €	176,513 €	139,342 €	151,205 €	173,982 €	171,585 €	1,731,591 €		
F.Pejë	71,972 €	60,635 €	47,949 €	115,433 €	73,710 €	87,867 €	132,541 €	126,110 €	87,821 €	119,926 €	102,058 €	138,405 €	1,164,428 €		
G.Gjilan	65,049 €	44,932 €	60,471 €	108,414 €	64,494 €	80,811 €	102,308 €	92,962 €	80,711 €	100,314 €	92,035 €	128,416 €	1,020,917 €		
H.Mitrovicë	57,249 €	36,775 €	56,190 €	116,327 €	83,755 €	62,765 €	86,583 €	77,690 €	71,655 €	97,398 €	69,111 €	120,658 €	936,153 €		
J.Suharekë	26,416 €	18,135 €	17,354 €	53,899 €	26,548 €	31,123 €	50,634 €	49,694 €	40,285 €	37,258 €	37,217 €	62,387 €	450,949 €		
K.Vushtrri	31,485 €	22,122 €	22,321 €	53,219 €	22,316 €	36,833 €	53,341 €	43,990 €	41,232 €	50,048 €	39,443 €	66,584 €	482,933 €		
L.Lipjan	18,317 €	10,785 €	12,539 €	37,237 €	26,142 €	22,911 €	39,343 €	44,326 €	27,124 €	29,259 €	30,959 €	52,951 €	351,893 €		
M.Klinë	10,438 €	7,198 €	16,116 €	31,772 €	6,673 €	17,085 €	31,872 €	22,326 €	14,074 €	24,841 €	19,077 €	36,646 €	238,117 €		
O.Drenas			8,052 €	38,447 €	19,072 €	19,455 €	41,107 €	37,641 €	39,509 €	32,276 €	29,795 €	49,382 €	314,735 €		
X.Depo Kry.	239,199 €	214,283 €	351,856 €	398,987 €	186,086 €	451,910 €	653,138 €	307,486 €	334,044 €	327,926 €	285,296 €	576,264 €	4,326,474 €		
Grand Total	1,152,744 €	952,101 €	1,223,958 €	1,942,100 €	1,214,358 €	1,549,743 €	2,266,223 €	1,772,909 €	1,620,758 €	1,846,542 €	1,714,389 €	2,342,221 €	19,598,046 €		

Figure 54 Sales in Stores per month

As we can see from the view that is created, we fulfilled another requested report for this data warehouse. On the report, we can see the sales in each of the stores based on the months and years. In the next view that is called “Sales based on bus unit”, we can

check the sales based on the sales channels, which are Retail, Wholesale and Online. Through this report, the management will understand which of the selling channels is performing better over the years.

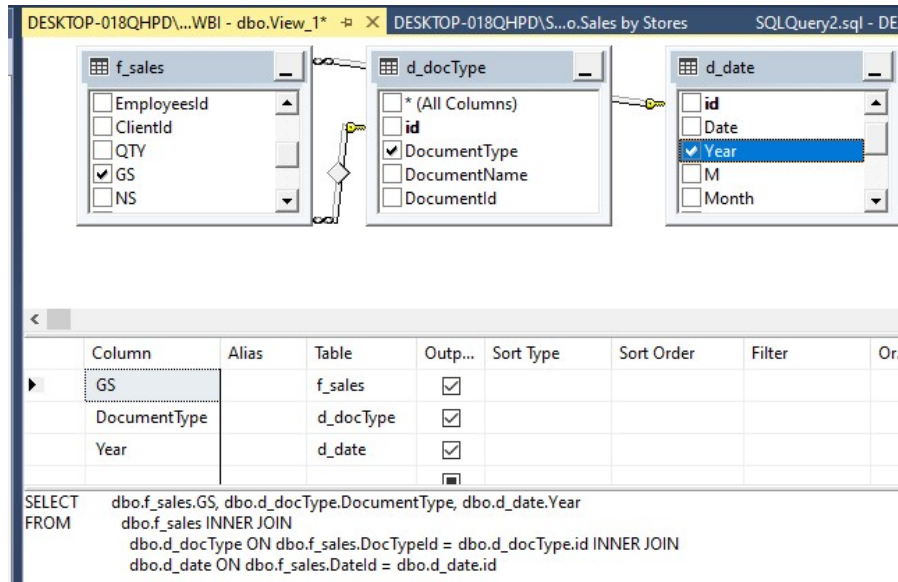


Figure 55 View of Sales by channels

DocumentType	2018		2019		2020		2021	
	Amount	%	Amount	%	Amount	%	Amount	%
Retail	9,623,482 €	78.88%	12,152,334 €	78.33%	11,962,700 €	80.03%	15,249,989 €	77.81%
WS	2,577,040 €	21.12%	3,362,348 €	21.67%	2,527,690 €	16.91%	4,326,474 €	22.08%
Online		0.00%		0.00%	458,116 €	3.06%	21,583 €	0.11%
<b>Grand Total</b>	<b>12,200,522 €</b>	<b>100.00%</b>	<b>15,514,683 €</b>	<b>100.00%</b>	<b>14,948,506 €</b>	<b>100.00%</b>	<b>19,598,046 €</b>	<b>100.00%</b>

Figure 56 Sales by channels

Now let's move one and create a view about the products imported from abroad the country and this report we will check based on the suppliers from which we took the goods. This report is very important for the management because based on the data we can check which our biggest supplier is and check the products that we imported or exported to other clients. The view can we created to see the data in different ways but we will use name of our suppliers and the amounts of goods in values that we imported over the years.

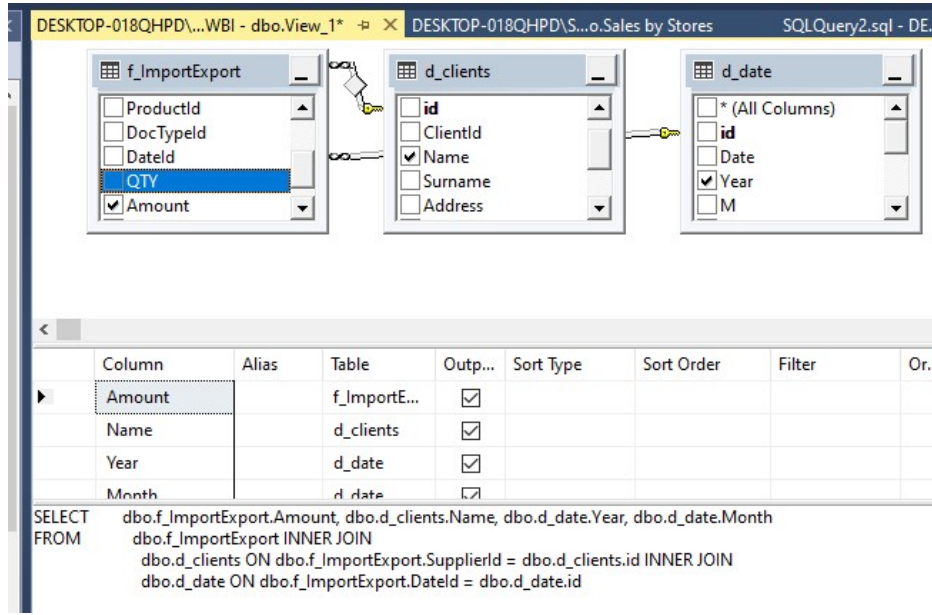


Figure 57 View of Imported products

Amount	Year			Grand Total
Name	2019	2020	2021	
ALBA-TRADE Handels GmbH	1,141,150 €	861,500 €	2,797,550 €	4,800,200 €
Beko Balkans	2,017,004 €	2,982,058 €	8,158,842 €	13,157,904 €
BSH Hisni aparati d.o.o Nazarje	139,040 €	134,511 €	561,147 €	834,698 €
Candy Hoover Group S.r.l.	155,143 €		148,987 €	304,129 €
Client69744	40,136 €	51,271 €	115,349 €	206,756 €
Client99222	189,730 €	286,029 €	761,281 €	1,237,039 €
DTM COMPANY DOO			99,607 €	99,607 €
ELICA S.p.A.			61,879 €	61,879 €
GROUPE SEB	130,340 €	66,989 €	280,840 €	478,169 €
LG Electronics Magyar KFT	236,032 €	894,657 €	3,196,376 €	4,327,064 €
PHILIPS ROMANIA SRL	56,318 €	264,314 €	456,008 €	776,640 €
<b>Grand Total</b>	<b>4,104,893 €</b>	<b>5,541,327 €</b>	<b>16,637,865 €</b>	<b>26,284,085 €</b>

Figure 58 Imports by Supplier

By modifying the view about imported products by supplier, we are going to include also products and categories from the d\_products table in order to check top 10 products that are imported on year 2021. Here are the results:



Name	All
Year	2021
PDescription	QTY
ENELARESE BEKO DSN04310X	1807
FRIGORIFER BEKO BCSA285K3SN	1730
ENELARESE BEKO BDIN24322	1694
BEKO FURRE BIE 24301 B	1650
DREJTUESE PER FLOKE PHILIPS BHS375/00	1604
BEKO PLLAKE HIC 64401 1X	1498
ENELARESE BEKO DIN28435	1428
SAMSUNG SM-A125FZKKEUC A12 4/128GB DS BLACK	1312
BEKO PLLAKE HIC 64401	1253
SAMSUNG SM-A125FZKVEUC A12 4/64GB DS BLACK	1095

Figure 59 Top products imported

The next view is going to be about stock of the products based on the Main Brands. In this view we can see also the age of the stock based on the days. A Main Brand is a super brand that can have more than one brand in their family. This view is called “Stock by Brands”:

DESKTOP-018QHPD\...ck by Main Brands\* - X SQLQuery3.sql - DE...18QHPD\artan (60))\*

f\_stock

- 120-180A
- 180-365Q
- 180-365A
- > 365Q
- > 365A

d\_products

- PMainBrand
- PBarcode
- PGroupCode
- PSegment
- PFamily

	Column	Alias	Table	Outp...	Sort Type	Sort Order	Filter
▶	QTY		f_stock	<input checked="" type="checkbox"/>			
	PMainBrand		d_products	<input checked="" type="checkbox"/>			
	PBrand		d_products	<input checked="" type="checkbox"/>			

```

SELECT  dbo.f_stock.QTY, dbo.d_products.PMainBrand, dbo.d_products.PBrand, dbo.d_products.PSegment,
        dbo.d_products.PFamily, dbo.f_stock.[0-30Q], dbo.f_stock.[30-60Q],
        |dbo.f_stock.[60-90Q], dbo.f_stock.[90-120Q],
        |dbo.f_stock.[120-180Q], dbo.f_stock.[180-365Q], dbo.f_stock.[> 365Q]
FROM    dbo.f_stock INNER JOIN
        dbo.d_products ON dbo.f_stock.ProductId = dbo.d_products.id

```

Figure 60 View of Stock by Main Brands

PMainBrand	QTY	0-30Q	30-60Q	60-90Q	90-120Q	120-180Q	180-365Q	>365Q
ACER	16	3	8	3	0	0	0	2
APPLE	797	553	130	23	59	3	15	14
ARCELIK	20,862	6,915	3,924	4,201	2,056	1,844	1,069	853
BOSCH	3,185	1,001	250	467	790	206	350	121
CANDY	1,601	429	175	0	384	189	360	64
DELL	264	64	111	56	23	0	1	9
ELICA	831	0	0	0	792	0	0	39
FUEGO	50,159	2,255	11,666	5,535	8,477	15,857	5,226	1,137
HP	830	483	160	68	39	37	5	38
HUAWEI	93	0	93	0	0	0	0	0
LENOVO	630	223	94	106	60	51	91	5
LG	3,864	1,501	1,215	718	176	38	194	22
NIKON	1	0	0	0	0	0	0	1
OTHER	15,263	4,040	3,817	1,731	1,264	1,229	1,997	1,171
PHILIPS	14,137	4,769	1,803	2,428	2,039	885	1,462	751
REMINGTON	1,399	0	445	0	622	0	332	0
SAMSUNG	4,001	1,566	539	791	68	311	592	134
SEB	8,896	0	2,148	1,037	2,590	416	1,824	881
SEGMENT	17	0	0	0	0	0	0	17
SONY	1,589	0	530	311	46	286	236	180
WHIRLPOOL	1	0	0	0	0	0	0	1
XIAOMI	829	620	121	63	0	1	5	19
<b>Grand Total</b>	<b>129,265</b>	<b>24,422</b>	<b>27,229</b>	<b>17,538</b>	<b>19,485</b>	<b>21,353</b>	<b>13,759</b>	<b>5,459</b>

Figure 61 Stock by Main Brands

The results of the last view that we are going to create are going to be showed using the Business intelligence tool called Power Bi. The results that will be showed in this report will be updated daily and have a better visualisation that will help the management get a snapshot about the results faster and in a better-looking way. Here is the view created for this report called “PoweBi”;

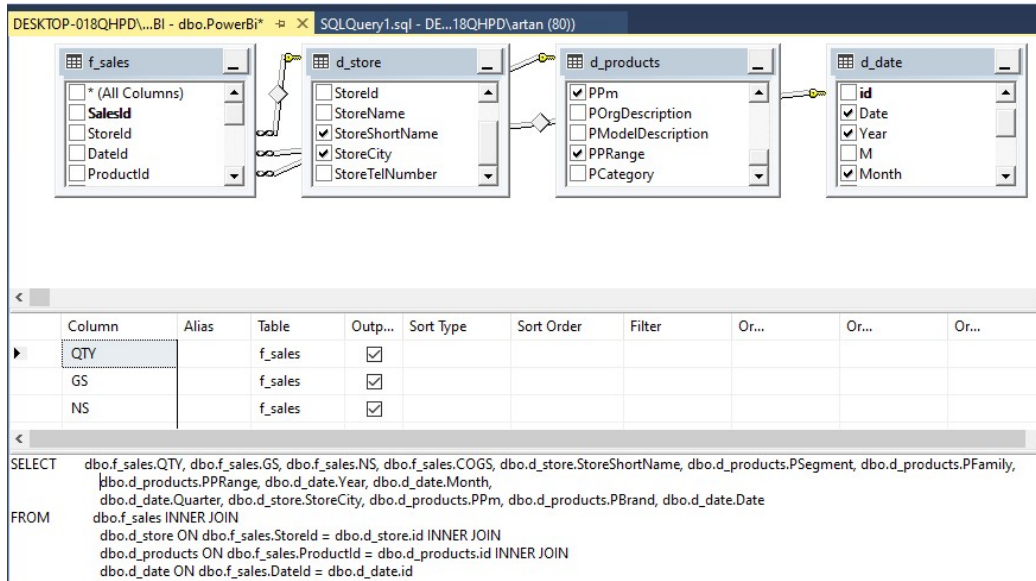


Figure 62 View PowerBi

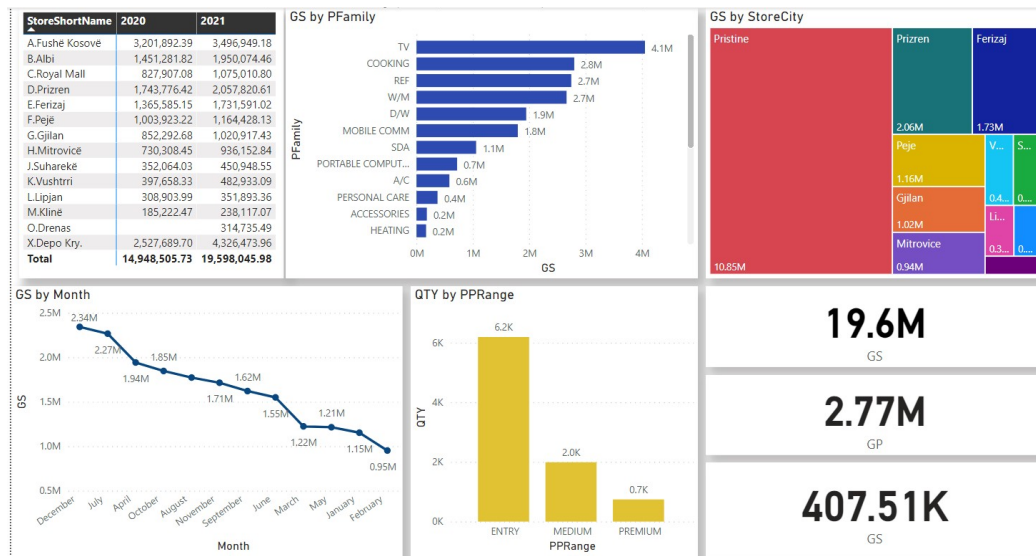


Figure 63 Reports created in PowerBi

As you can see in the figure 63, the we have created some visuals that will show us and the management snapshots of data for the company results. On the first table, we can see the sales per store in two years 2020 and 2021. The other reports are all about year 2021. In the second chart, we have sales in categories from the highest to the lowest category. The third chart shows sales per city and for sure, Pristina has the



highest sales because Neptun in Pristina has three stores. In the fourth chart, we have sales per month and for sure, December is the highest month on sales. Then we showed sales by quantity based on the products price range. Product price range is important to see which the products that we are selling the most are. Finally, as you can see there is 19.6m total sales realization for year 2021, then we can see total Gross Profit and in the last card, there are the sales on the last day of the year 2021. The final card will be updated based on the previous day sales.

## 5.0 CONCLUSION

This thesis introduces the idea to carry out a data warehouse in a retail company such as Neptun. In order to carry out this data warehouse and use the business intelligence tools that a company needs for better reporting data in is necessary to have a strong business commitment by an executive sponsor who believes on the project.

In this thesis, we will see that the key to a data warehouse is a data design. Data design creates the good data warehouse. Here we have been focused on determining the needs of the users in this case the company. We also have located the sources of the data and those data are organized in a data dimensional model that will represent the business needs for the future use. The data that are needed must find a way to get to the right people in ample time. However, the most important thing is that those data should bring benefits to be organization. It is understandable that business intelligence must be part of the companies such as Neptun because it will allow us understand the results of the past but also will help us to find new opportunities to reach a better market share in our country and find new opportunities and be in the time for the new trends on the future.

Like all the other serious companies in retail or e-commerce Neptun has stored working data for some time now, and we have continued to amass huge amount of data at a massive rate the working databases becomes more valuable, it helps grow the rate at which business succeed. In this thesis, we have shown how data warehouse will collect, join, load, organize and summarize this structured and unstructured data and how this data will be used to tell future business decisions.

In this thesis will be included the designing and implementation on a small-scale the of a data warehouse. We will see how star schema works from an organized and effective ETL tool from different database and multiple data sources. This example will show us the advantages of building and implementing a data warehouse and with this; we will achieve an improved experience for users who use this data for better future decisions.

## 6.0 BIBLIOGRAPHY OF THESES

1. Building a Data Warehouse: With Examples in SQL Server Copyright © 2008 by Vincent Rainardi - www.apress.com.
2. Third Edition - Copyright © 2013 by Ralph Kimball and Margy Ross. The Data Warehouse Lifecycle Toolkit, Second Edition - Copyright □ 2008 by Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy and Bob Becker.
3. The Microsoft ® Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft® Business Intelligence Toolset, Second Edition.
4. How to write a research methodology - <https://www.scribbr.com/dissertation/methodology/> (accessed 5 April 2020).
5. How to Write Your Dissertation Methodology - <https://www.wikijob.co.uk/content/postgraduate/phd/dissertation-methodology> (accessed 5 April 2020).
6. Top Five Benefits of a Data Warehouse - <https://www.tibco.com/blog/2011/08/10/how-to-sell-the-business-on-a-data-warehousing-project/> (accessed 7 February 2020).
7. Why the Retail Industry Needs Data Warehousing and Analytics - <https://www.voicendata.com/retail-industry-needs-data-warehousing-analytics/> (accessed 5 May 2020).
8. Star Schema – Complete Reference by Christopher Adamson 2010.
9. System Analyses and Design, Arunesh Gayal.
10. SQL for Dummies 8<sup>th</sup> Edition - Allen G. Taylor
11. Microsoft Business Intelligence for Dummies – Ken Withee
12. Mastering Microsoft Power BI - Brett Powell
13. SQL Server Integration Services Design Patterns - Tim Mitchell, Matt Masson, Andy Leonard, Jessica Moss, Michelle Ufford
14. A Review of Data Warehousing and Business Intelligence in different perspective - Vijay Gupta.
15. Relational Database Management System, ISRD Group,
16. [https://www.accent.mk/?page\\_id=785&lang=en](https://www.accent.mk/?page_id=785&lang=en)
17. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modelling - Kimball R. and Ross M., (1996).

18. Data Warehouse Toolkit: Second Edition, the Complete Guide to Dimensional Modelling - Kimball R. and Ross M.
19. Structured Query Language (SQL) A Practical Introduction - By Akeel I. Din 1994.
20. Inside Microsoft SQL Server 2008 T-SQL Programming - By Itzik Ben-Gan, Dejan Sarka, Greg Low, Roger Wolter – 2010.
21. Business Intelligence Tools for Small Companies A Guide to Free and Low-Cost Solutions - Albert Nogués and, Juan Valladares · 2017.