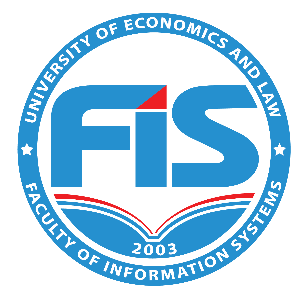
**UNIVERSITY OF ECONOMICS AND LAW**

**FACULTY OF INFORMATION SYSTEMS**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

****

**FINAL PROJECT REPORT**

**Data Warehouse and Integration**

**TOPIC:**

**Lecturer: Nguyen Van Ho**

**Group:  *Máo Wỷ***

**Ho Chi Minh City, December 25th, 2022­­**

# Members of Group Máo Wỷ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *No.* | **NAME** | **STUDENT ID** | **Contribution** | **Signatures** |
| *1* | Nguyễn Hoàng Lợi | K204060304 | 100% |  |
| *2* | Hồ Mai Minh Nhật | K204161995 | 100% |  |
| *3* | Bùi Thị Tri Tâm | K204161999 | 100% |  |
| *4* | Hoàng Nguyễn Quỳnh Giang | K204161987 | 100% |  |
| *5* | Lê Thị Thùy Linh | K204161990 | 100% |  |

# Acknowledgments

My team would like to thank Mr. Nguyen Van Ho - Lecturer of Data Warehouse and Integration and assistant lecturer Le Ba Thien who equipped us with basic knowledge and skills to complete the final project.

However, in the process of researching the topic, due to limited specialized knowledge, we still have many shortcomings when researching, evaluating and presenting the topic. We hope to receive the attention and suggestions of teachers to improve our topic.

Sincerely thanks Gr. Máo Wỷ

# Commitment

We hereby declare that the above project is the research work of our group under the guidance of lecturer Nguyen Van Ho and assistant lecturer Le Ba Thien. The statements stated in the project are also the results of direct, serious, independent research of the author himself and the basis of searching, understanding and studying scientific documents or translations other have been announced. The project will still help ensure objectivity, honesty and science.

Gr. Máo Wỷ

**Contents**

[Members of Group Máo Wỷ 1](#_Toc122900426)

[Acknowledgments 2](#_Toc122900427)

[Commitment 3](#_Toc122900428)

[CHAPTER 1: INTRODUCTION 8](#_Toc122900429)

[1. Business case for the project 8](#_Toc122900430)

[2. Objectives of the project 8](#_Toc122900431)

[2.1. General Objective 8](#_Toc122900432)

[2.2. Specific Objectives 8](#_Toc122900433)

[3. Research Objects: 9](#_Toc122900434)

[4. Scope of the project: 9](#_Toc122900435)

[5. Value and desired outcome of the project 9](#_Toc122900436)

[6. Structure of project: 10](#_Toc122900437)

[CHAPTER 2: THEORETICAL BASIS 11](#_Toc122900438)

[1. Overview of DWH 11](#_Toc122900439)

[1.1. What is DWH? 11](#_Toc122900440)

[1.2. DWH Architecture [1] 11](#_Toc122900441)

[1.3. Advantage of BI in enterprises 11](#_Toc122900442)

[2. Data warehouse and Data mart 12](#_Toc122900443)

[2.1. What are Data warehouse and Data mart? [4] 12](#_Toc122900444)

[2.2. Who needs Data warehouse and Data mart? 14](#_Toc122900445)

[2.3. Advantages and disadvantages of Data warehouse 14](#_Toc122900446)

[3. Snowflake and Star schemas [5] 15](#_Toc122900447)

[4. Integration and ETL Process 17](#_Toc122900448)

[4.1. What is ETL? 17](#_Toc122900449)

[4.2. Why do we need ETL? 17](#_Toc122900450)

[4.3. ETL Process [6] 18](#_Toc122900451)

[CHAPTER 3: REQUIREMENTS ANALYTICS 20](#_Toc122900452)

[1. Business processes of Sales & HR 20](#_Toc122900453)

[1.1. Sales & HR department 20](#_Toc122900454)

[1.2. The purpose of Sales & HR 21](#_Toc122900455)

[1.3. Sales & HR process 21](#_Toc122900456)

[2. Data source and challenges 22](#_Toc122900457)

[2.1. Overview about Adventure World [9] 22](#_Toc122900458)

[2.2. HR [10] 23](#_Toc122900459)

[2.3. Sales 24](#_Toc122900460)

[3. Business Requirements Analysis of Sales & HR 27](#_Toc122900461)

[4. IT requirements Analysis (IT &amp; Infrastructure) (optional) 27](#_Toc122900462)

[CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA 28](#_Toc122900463)

[1. Designing Data Warehouse: 28](#_Toc122900464)

[1.1. Bus Matrix: 28](#_Toc122900465)

[1.2. Transactional and master data for fact and dimension tables: 28](#_Toc122900466)

[1.3. Data Warehouse model (Snowflake or Star): 37](#_Toc122900467)

[2. ETL processes 38](#_Toc122900468)

[2.1. Dimension Table’s ETL Process 43](#_Toc122900469)

[2.2. Fact Table’s ETL Process 45](#_Toc122900470)

[CHAPTER 5: CONCLUSION AND FUTURE WORKS 47](#_Toc122900471)

[1. Results: 47](#_Toc122900472)

[2. Limitations 49](#_Toc122900473)

[3. Future works 50](#_Toc122900474)

[REFERENCES 51](#_Toc122900475)

[APPENDIX 52](#_Toc122900476)

**LIST OF FIGURES**

Figure 1: Data Mart is a part of Data Warehouse 13

Figure 2: Snowflake schema 16

Figure 3:Star schema 17

Figure 4: Module HR 23

Figure 5: Module Sales 24

Figure 6: Bus matrix 28

Figure 7: Relational model 37

Figure 8: Source to staging process 38

Figure 9: Load Dimension for Currency 39

Figure 10: Load Dimension for Customer 39

Figure 11: Load Dimension for Employee 40

Figure 12: Load Dimension for Date 40

Figure 13: Load Dimension for Product 41

Figure 14: Load Dimension for Reseller 41

Figure 15: Load Dimension for Sales Territory 42

Figure 16: Load Fact Internet Sales 42

Figure 17: Load Fact Reseller Sales 43

Figure 18: Load Fact Employee Pay 43

Figure 19: Data in Currency Dimension 44

Figure 20: Data in Customer Dimension 44

Figure 21: Data in Date Dimension 44

Figure 22: Data in Employee Dimension 44

Figure 23: Data in Product Dimension 45

Figure 24: Data in Sales Territory Dimension 45

Figure 25: Data in Reseller Dimension 45

Figure 26: Data in Internet Sales Fact 45

Figure 27: Data in Reseller Sales Fact 46

Figure 28: Data in Employee Pay Fact 46

Figure 29: Tableau BI tool 47

Figure 30: Reseller sales by countries. 48

Figure 31: Sales over time(quaterly). 48

Figure 32: Compare sales amount of Reseller and Internet by product category. 49

Figure 33: Top 10 salesman 49

Figure 34: Fact - Dim connection in Tableau example. 50

**LIST OF TABLES**

[Table 1: Differences between Data Warehouse and Data Mart 13](#_Toc122884519)

[Table 2: HR 23](#_Toc122884520)

[Table 3: Sales 25](#_Toc122884521)

[Table 4: Dim Currency 28](#_Toc122884522)

[Table 5: Dim Customer 29](#_Toc122884523)

[Table 6: Dim Date 30](#_Toc122884524)

[Table 7: Dim Employee 31](#_Toc122884525)

[Table 8: Dim Product 32](#_Toc122884526)

[Table 9: Dim Reseller 33](#_Toc122884527)

[Table 10: Dim Sales Territory 33](#_Toc122884528)

[Table 11: Fact Employee Pay 34](#_Toc122884529)

[Table 12: Fact Internet Sales 34](#_Toc122884530)

[Table 13: Fact Reseller Sales 35](#_Toc122884531)

# CHAPTER 1: INTRODUCTION

## Business case for the project

Organization challenges may be classified by departments; in this case, we will discuss Sales and Human Resources (HR), two of the most significant departments in the business. These two schemas are used in this project to provide simple business requirements.

Sales data analysis is a critical business activity that influences the company's survival and growth. The sales department must identify the challenges they will face while contacting clients. Who are the most loyal consumers across all product categories? Who should the sales team prioritize in terms of long-term relationships? What are the customer's issues? What items are buyers purchasing and at what price point?

Regarding Human Resources Most businesses are concerned about employee retention. They need a method to track both employee reviews and potential candidate profiles. They also want to increase their capacity to forecast how many employees will be needed and plan for changes in pay and benefits on the job.

## Objectives of the project

### 2.1. General Objective

The purpose of the project is to successfully build a data warehouse which makes data available to the business users for decision making purposes, enabling the dealership to use analytics and statistics to help the company sell more products.

### 2.2. Specific Objectives

Identification of particular types of business processes that are affected by the introduction of a data warehouse and also how these business processes are affected.

Identification and categorization of the factors that cause the need for business process change during the pre-deployment stage of a data warehouse project.

Set up and store a centralized set of data linked by two modules (Sales and HR) from the AdventureWork database (Including the data needed to build a DW according to business operations)

Build and deploy a DataWarehouse with the goal of solving the proposed business operations. Specifically, the goal in this project is that with the created DW, it is possible to draw some insights about the sales and HR management department of Adventure Works Cycles company, from which to make a few suggestions to solve.

Learn more about DW, partly understand how to do a DW project from the beginning to the end

## Research Objects:

Our research objects are based on the AdventureWork2019 database. As we mentioned above, we will mainly focus on the Sales and Human Resources schema of this database in order to build a data warehouse for the final subject assignment.

## Scope of the project:

The database, knowledge and document used in this project served our final assignment project, not for any commercial purposes.

We only implement a data warehouse for Sales and HR departments solving simple business requirements due to our shortage of time and experience.

The scope of this project should be managed carefully, especially when we want to scale up or make some specific changes. Therefore, it would be a great honor to receive comments from the professor for our project in the future.

## Value and desired outcome of the project

* Save Time

Since business users can quickly access critical data from a number of sources in a single platform-they can rapidly make informed decisions on key initiatives. They won’t waste precious time retrieving data from multiple sources.

Users can query the data themselves with little or no support from IT-saving more time and money. That means business users won’t have to wait until IT gets around to generating the reports, and hardworking IT analysts can focus on keeping the business running.

* Increase Data Quality and Consistency

A data warehouse implementation includes the conversion of data from numerous source systems into a common format. Since data from various departments is standardized, each department will produce results in line with all the other departments. With data virtualization capabilities, you can have more confidence in the accuracy of your data. And accurate data is the basis for strong business decisions.

* Provide Historical Intelligence

A data warehouse stores large amounts of historical data so you can analyze different time periods and trends to make future predictions. Such data typically cannot be stored in a transactional database or used to generate reports from a transactional system.

In short, the result we want to achieve is to build a Data warehouse to solve the bus matrix about Sales and HR modules.

## Structure of project:

The project consists of 5 chapters:

Chapter 1: Introduction

Present the requirement of the research problem, the research purpose, the object and scope of the research, the contribution of the study, and the structure of the report in detail.

Chapter 2: Theoretical basis

Present the theoretical basis and theory used in the project..\

Chapter 3: Requirement Analytics

Present the sales and HR departments' businesses and how they satisfy the demands of the company. Apart from providing an overview of data sources and data challenges.

Chapter 4: Building data warehouse and Intergrating data

Describe in detail how we built the data warehouse in accordance with the requirements of the project.

Chapter 5: Conclusion and Future works

Present the results, limitations of the project and future solutions.

# 

# CHAPTER 2: THEORETICAL BASIS

## Overview of DWH

### 1.1. What is DWH?

For the purposes of supporting data analysis, data mining, artificial intelligence (AI), and machine learning, a data warehouse is a system that collects data from several sources into a single, central, consistent data storage. In ways that a traditional database cannot, a data warehouse system enables an organization to do complex analytics on huge volumes (petabytes and petabytes) of historical data. [[1]](https://www.google.com/url?q=https://www.ibm.com/cloud/learn/data-warehouse&sa=D&source=docs&ust=1671983033423206&usg=AOvVaw1vgLmM2KzJPt42VYQK6v70)

Large amounts of data from many sources are centralized and consolidated in a data warehouse. Organizations may gain useful business insights from their data using its analytical skills to enhance decision-making. It creates a historical record over time that data scientists and business analysts may use to their advantage. A data warehouse may be thought of as the "single source of truth" for an organization because of these features. [[2]](https://www.google.com/url?q=https://www.oracle.com/database/what-is-a-data-warehouse&sa=D&source=docs&ust=1671983033424587&usg=AOvVaw1iLiGpItC1gcsBkgWlqeTb)

### 1.2. DWH Architecture [[1]](https://www.google.com/url?q=https://www.ibm.com/cloud/learn/data-warehouse&sa=D&source=docs&ust=1671983033423206&usg=AOvVaw1vgLmM2KzJPt42VYQK6v70)

Data warehouses have a three-tier architecture:

* Bottom tier: This area contains a server, usually a relational database system, which collects, cleanses, and transforms data from multiple data sources through a process known as Extract, Transform, and Load (ETL) or a process known as Extract, Load, and Transform (ELT).
* Middle tier: The middle tier consists of an OLAP (i.e. online analytical processing) server which enables fast query speeds. Three types of OLAP models can be used in this tier, which are known as ROLAP, MOLAP and HOLAP. The type of OLAP model used is dependent on the type of database system that exists.
* Top tier: The top tier is a visual representation of their search results, often in a chart format, which enables end users to conduct ad-hoc data analysis on their business data.

### 1.3. Advantage of BI in enterprises

Business intelligence (BI) processes help organize data so that it can be conveniently accessed and evaluated. Business decision makers gather deeper insights to make more informed business decisions.

BI is an essential component that connects big data to a company's business strategy. It enables companies to track, gather, analyze, and deliver relevant insights into their sales, operations, employees, customer service, and a variety of other elements. It combines and organizes data, analyses it, and presents it in a cohesive manner — in the form of dashboards or reports — fostering a better knowledge of the business and assisting in decision-making.

Business intelligence systems may uncover bottlenecks in manufacturing and supply chains, gaps in customer service, untapped income sources, waste, and general costs, among other things. “The ability to use this data to innovate and compete will set winners apart from losers,” says Bob Rogers, chief data scientist for IT Transformation, Data Center Group at Intel [[3]](https://www.google.com/url?q=https://mytechdecisions.com/compliance/benefits-business-intelligence-enterprise/&sa=D&source=docs&ust=1671983033420301&usg=AOvVaw3T1ZDCGzD2mtNM2nWltPP4).

## Data warehouse and Data mart

### 2.1. What are Data warehouse and Data mart? [[4]](https://www.google.com/url?q=https://www.talend.com/resources/what-is-data-mart/&sa=D&source=docs&ust=1671983033421681&usg=AOvVaw2I-IUPVtHBXR27hKnTpfxp)

A data warehouse is a kind of data management system that enables and supports business intelligence (BI) operations, particularly analytics. Data warehouses exist exclusively to do queries and analysis on enormous volumes of historical data. A data warehouse's data is often generated from a variety of sources, including application log files and transaction programs.

A data mart is a subject-oriented database that is frequently a partitioned section of a larger data warehouse. A data mart's subset of data is often associated with a certain business unit, such as sales, finance, or marketing. Data marts speed up company operations by providing access to essential information in a data warehouse or operational data store in days rather than months or years. A data mart is a cost-effective solution to acquire meaningful insights rapidly since it only contains data relevant to a certain business sector.

Data marts and data warehouses are both highly organized repositories for storing and managing data until it is required. However, they differ in the scope of data stored: data warehouses are designed to serve as the core data repository for the whole organization, whereas data marts are designed to meet the needs of a single division or business function. Because a data warehouse holds information for the whole firm, it is important to strictly manage who has access to it. Furthermore, retrieving the data you want in a data warehouse is an extremely challenging operation for the organization. As a result, the fundamental objective of a data mart is to distinguish partitions - a smaller collection of data from a larger set in order to allow better data access for the end users.

A data mart can be created from an existing data warehouse - the top-down approach or from other sources, such as internal operating systems or external data. It is a relational database that contains transactional data (time value, numerical order, reference to one or more objects) in columns and rows to make it simple to organize and retrieve. This is similar to a data warehouse.

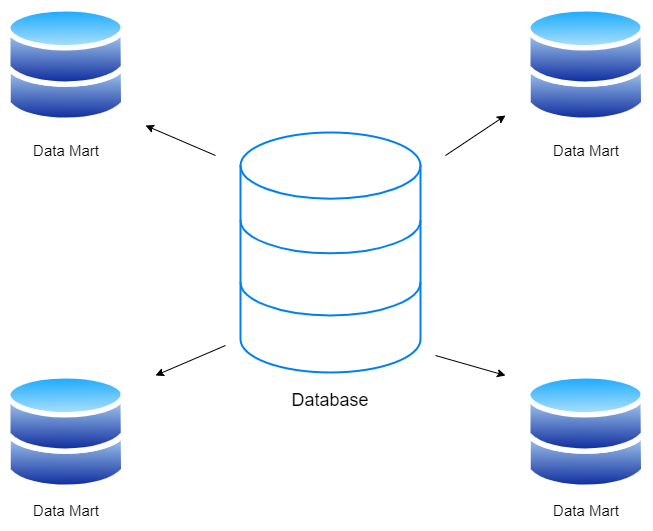


Figure 1: Data Mart is a part of Data Warehouse

On the other hand, different business units may develop their own data marts in accordance with their own data needs. Multiple data marts can be combined to produce a single data warehouse if business needs so demand. The bottom-up approach to development is this.

Table 1: Differences between Data Warehouse and Data Mart

|  |  |  |
| --- | --- | --- |
|  | Data Mart | Data Warehouse |
| Size | less than 100 GB | 100 GB to 1TB+ |
| Subject | Single Subject | Multiple Subject |
| Usage | It helps to take tactical decisions for the business | It helps to take a strategic decision |
| Scope | Line-of-business | Enterprise-wide |
| Data Source | Few sources | Many sources |
| Data Integration | One subject | All business data |
| Time to build | Minutes, Weeks, Months | Many months to years |

### 

### 2.2. Who needs Data warehouse and Data mart?

Because the major purpose of a data warehouse is to assist a company in making decisions about creating various analytics and reports based on stored data, business owners and decision makers will be required.

A data mart provides users with particular data on one of the organization's departments or an area of the company, is dependent on a department, and is exclusively utilized for decision making inside the department. As a result, personnel from each department will need to use

### 2.3. Advantages and disadvantages of Data warehouse

* Advantages
* Save fund and times: A data warehouse allows you to discover any information in minutes, saving time for employees when accessing specific data.
* Improves quality of data while maintaining consistency: Data warehousing adds value to data by transforming it to a standard format. As a result, data from several sources can be combined into a single pattern. This data standardization may help an organization achieve well-developed business results, allowing the company to function with consistency and precision.
* Provides enhanced business insights: Through effective business procedures, data may be integrated and retrieved from many sources. And this feature permits business intelligence (BI) to significantly boost.
* Improves data security: Using a data warehouse, the data sources may be maintained connected and properly safeguarded, reducing the risk of data infringement.
* Disadvantages
* Data homogenization: Data warehousing involves numerous comparable types of information gathered from several sources, which can lead to data homogeneity and inflexibility. This depicts data loss followed by limited access during aggregation.
* Hidden issues in data sources: Data acquired from several internal sources might go undiscovered for years, resulting in hidden flaws in data warehousing systems. For example, when adding new data, some fields may allow null values, resulting in incomplete information even while the facts are accessible. These concealed concerns take time to resolve.

## Snowflake and Star schemas [[5]](https://www.google.com/url?q=https://www.javatpoint.com/data-warehouse-star-schema-vs-snowflake-schema&sa=D&source=docs&ust=1671983033426203&usg=AOvVaw1w9vh3mBS-DGFK4Jii6d_F)

Snowflake schema is a multidimensional database with logical tables in which the entity-relationship diagram is designed to look like a snowflake. It is frequently represented as a centralized fact table that is related to many and distinct dimensions. The dimension tables are separated into many dimension tables that are often normalized to eliminate redundancy.

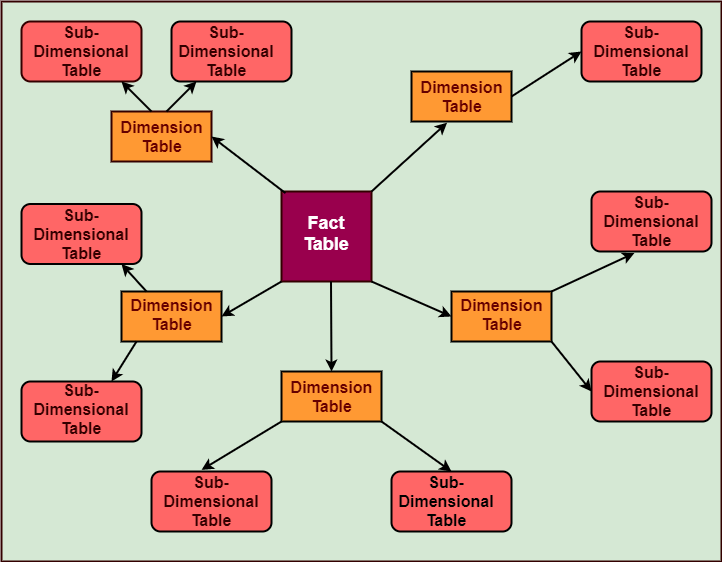


Figure 2: Snowflake schema

A star schema is a relational schema in which the design represents a multidimensional data model. The explicit data warehouse schema is the star schema. The entity-relationship diagram of this schema is called as a star schema because it resembles a star with points extending from a center table. The schema's core is a large fact table, while the points of the star are dimension tables.

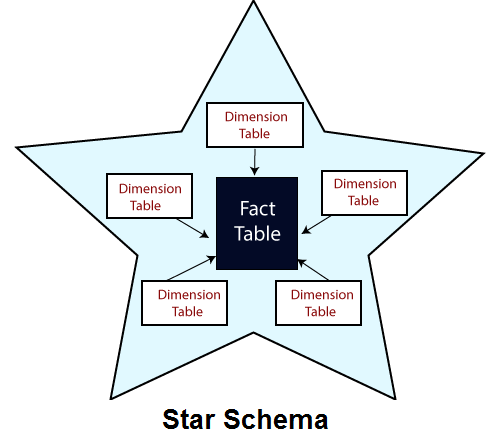


Figure 3:Star schema

## Integration and ETL Process

### 4.1. What is ETL?

ETL (extract, transform, and load) is a data integration process that integrates data from various sources into a single, consistent data store that is then put into a data warehouse or other target system.

The foundation for data analytics and machine learning workstreams is ETL. ETL cleanses and organizes data using a set of business rules to fulfill particular business intelligence needs, such as monthly reporting, but it may also handle more complex analytics to improve back-end operations or end user experiences. An organization will frequently employ ETL to:

* Data extraction from legacy systems
* Cleanse the data to increase its quality and consistency.
* Load data into the specified database.

### 4.2. Why do we need ETL?

The importance of ETL in an organization is in direct proportion to how much the organization relies on data warehousing. For example:

* Marketing firms that desire to extract data from numerous client relationship management (CRM) systems must first convert all of the data into a single format before loading it into business intelligence programs that can uncover trends in customer behavior.
* Hospitals that seek to extract patient data from legacy systems, reformat the data into a format that the new system understands, and put the reformatted patient data into a single system that enables healthcare professionals to make better health choices.
* Federal agencies that want to extract information from hundreds of local databases, categorize the data, and load it into a national database so authorities can obtain a seeing perspective of common problems in towns throughout the country.

→ The ETL procedure saves time and improves data quality. An ETL solution could be useful if someone has to transport, categorize, or standardize data.

### 4.3. ETL Process [[6]](https://www.google.com/url?q=https://www.ibm.com/cloud/learn/et&sa=D&source=docs&ust=1671983033427516&usg=AOvVaw30ysu0sDcDqgydoMbbP8fZ)

The ETL process is comprised of 3 steps that enable data integration from source to destination

Step 1: Extraction

Most firms handle data from several sources and employ a range of data analysis technologies to provide business insight. To carry out such a complicated data strategy, data must be free to transfer across systems and apps.

Data must first be extracted from its source, such as a data warehouse or data lake, before it can be transported to a new location. Structured and unstructured data are imported and merged into a single repository in this initial phase of the ETL process. Large amounts of data may be retrieved from a variety of data sources, including:

* Existing databases and legacy systems
* Cloud, hybrid, and on-premises environments
* Sales and marketing applications
* Mobile devices and apps
* CRM systems
* Data storage platforms
* Data warehouses
* Analytics tools

Step 2: Transformation

The raw data is processed. The data is converted and consolidated here for its intended analytical use case. This phase may include the following tasks:

* Filtering, cleansing, de-duplicating, validating, and authenticating the data.
* Performing calculations, translations, or summarizations based on the raw data. This can include changing row and column headers for consistency, converting currencies or other units of measurement, editing text strings, and more.
* Conducting audits to ensure data quality and compliance
* Removing, encrypting, or protecting data governed by industry or governmental regulators
* Formatting the data into tables or joined tables to match the schema of the target data warehouse.

Step 3: Load

The converted data is transported from the staging area into the target data warehouse in this final stage. This often involves an initial load of all data, followed by recurrent loads of incremental data updates and, less frequently, complete refreshes to wipe and replace data in the warehouse. Most ETL processes are automated, well-defined, continuous, and batch-driven in most businesses. ETL is often performed during off-hours when traffic on source systems and the data warehouse is at its lowest.

# 

# 

# CHAPTER 3: REQUIREMENTS ANALYTICS

## Business processes of Sales & HR

### 1.1. Sales & HR department

* Sales department [[7]](https://wiki.tino.org/sales-department-la-gi/)

Sales Department is a department responsible for selling the products of the business. The Sales Department includes business operations and sales processes that help distribute products efficiently, support strategies, and accomplish business-oriented goals.

And in the Sales department, each person (or group of people) will be in charge of specialized tasks, specifically:

* Distributing products to large and small distribution agents and consumers, reaching many different markets.
* Monitor market needs and tastes, manage inventory, process processing and distribute products to the market.
* Ensure outreach and sales activities are in accordance with the business plan.
* Manage the changing needs of each region, control and deal with problems arising from competition.
* Product supply management is always guaranteed to the market.
* HR department [[8]](https://amis.misa.vn/30711/hr-department-la-gi/)

HR department or Human Resources Department is understood as the department responsible for administration, supervision and direct communication with human resources in each company or enterprise.

The HR department is a complete system in charge of HR related issues in the enterprise. And in the HR department, each person (or group of people) will be in charge of specialized tasks, specifically:

* Recruitment (Recruitment)
* Developing and implementing salary and benefits policy (Compensations and Benefits)
* Training and Human Resource Development (Learning and Development)
* Labor Relations Management
* Administration - Human Resources (HR Admin)

As the person who works directly with the company's personnel, the HR Department is expected to be the department that understands the thoughts and feelings of the employees the most, and is considered the key to building corporate culture and uniting members together.

### 1.2. The purpose of Sales & HR

The main and biggest purpose of Sales & HR is to make the company more able to grow. While Sales tries to make the total revenue of the company increase, HR tries to recruit candidates, expand the company size as well as motivate and encourage employees. If you combine Sales and HR together, the working efficiency of the whole company will be enhanced. Sales leaders can help HR understand sales strategy, customers and products, and the type of selling resources needed. HR, in turn, can help sales apply a higher level of discipline to talent management so that the sales organization is capable of performing at its best.

### 1.3. Sales & HR process

* HR process

Let’s think about this in terms of the employee lifecycle. End-to-end processes are all the different steps that introduce an employee, manage their performance, help them advance, or help them exit a company. When you have a process for each stage, you are effectively guiding an employee on their professional journey.

Some of the most common include:

* Recruiting
* Onboarding & Off-boarding
* Workforce & Succession Planning
* Employee Engagement & Relations
* Compensation, Benefits, Rewards
* Absence Management
* Performance Management

In short, all of these touch employees on their first, last, and most important days at work. That’s why having them in place can often play a key role in organizational success.

* Sales process:

A sales process is a set of repeatable steps that a sales person takes to take a prospective buyer from the early stage of awareness to a closed sale.

Typically, a sales process consists of 5-7 steps:

* Prospect: is the process of sourcing new, early-stage leads to begin working through the sales process. It's a vital part of the sales process and part of most reps' daily or weekly workflow.
* Connect and qualify leads: The connect step of the sales process involves reps initiating contact with those early-stage leads to gather information. The second part of this step is qualifying new leads — deciding whether or not they're a good-fit lead for your business and whether or not they'll likely move forward in the buyer's journey.
* Research the company: Research helps your reps put themselves in the customer's shoes to offer a more tailored and personalized experience, thus improving the likelihood of closing a deal.
* Give an effective pitch: The presentation step is typically when your salesperson runs a formal product or service demonstration for your prospect.
* Handle objections: Listening to your prospect's objections and questions can help your reps better tailor your product to fit their needs. Through their research and presentation preparation, reps should identify and anticipate possible objections, whether about cost, onboarding, or other parts of the proposed contract.
* Close the deal: This step of the sales process refers to any late-stage activities that happen as a deal approaches closing. It varies widely from company to company and may include delivering a quote or proposal, negotiation, or achieving the buy-in of decision-makers.
* Nurture and continue to sell: The final step of the sales process also involves continuing to communicate and reinforce value to customers. This can provide opportunities to upsell and cross-sell, as well as opportunities to get secure referrals from delighted customers.

## Data source and challenges

### 2.1. Overview about Adventure World [[9]](https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/analysis-services-tutorial-scenario?view=asallproducts-allversions)

This project is based on Adventure Works Cycles, a fictitious company. Adventure Works Cycles is a large, multinational manufacturing company that produces and distributes metal and composite bicycles to commercial markets in North America, Europe, and Asia. The headquarters for Adventure Works Cycles is Bothell, Washington, where the company employs 500 workers. Additionally, Adventure Works Cycles employs several regional sales teams throughout its market base.

In recent years, Adventure Works Cycles bought a small manufacturing plant, Importadores Neptuno, which is located in Mexico. Importadores Neptuno manufactures several critical subcomponents for the Adventure Works Cycles product line. These subcomponents are shipped to the Bothell location for final product assembly. In 2005, Importadores Neptuno became the sole manufacturer and distributor of the touring bicycle product group.

Following a successful fiscal year, Adventure Works Cycles now wants to broaden its market share by targeting advertising to its best customers, extending product availability through an external Web site, and reducing the cost of sales by reducing production costs.

### 2.2. HR [[10]](https://www.google.com/url?q=https://dataedo.com/download/AdventureWorks.pdf&sa=D&source=docs&ust=1671982895670171&usg=AOvVaw3yqFsJAoqVn5i-ucY8_5e6)

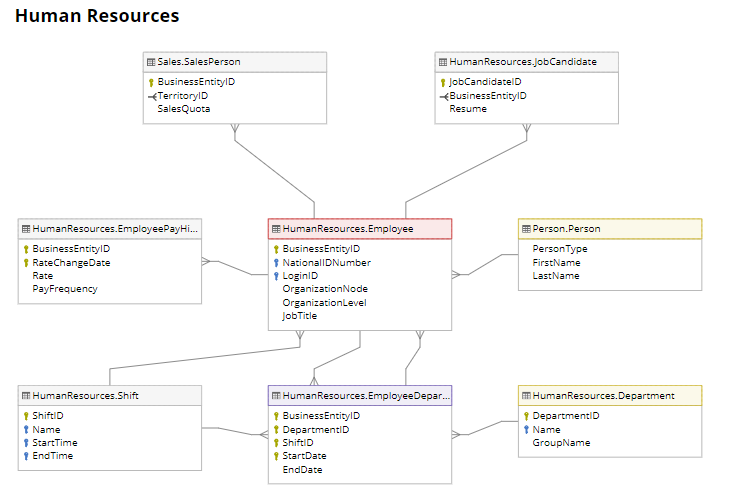
****

Figure 4: Module HR

Table 2: HR

|  |  |  |
| --- | --- | --- |
| STT | Tables | Descriptive |
| 1 | HumanResources.Department | Lookup table containing the departments within the Adventure Works Cycles company. |
| 2 | HumanResources.Employee | Employee information such as salary, department, and title. |
| 3 | HumanResources.EmployeeDepartmentHistory | Employee department transfers. |
| 4 | HumanResources.EmployeePayHistory | Employee pay history. |
| 5 | HumanResources.JobCandidate | Résumés submitted to Human Resources by job applicants. |
| 6 | HumanResources.Shift | Work shift lookup table |
| 7 | Person.Person | Human beings involved with AdventureWorks: employees, customer contacts, and vendor contacts |

### 

### 2.3. Sales

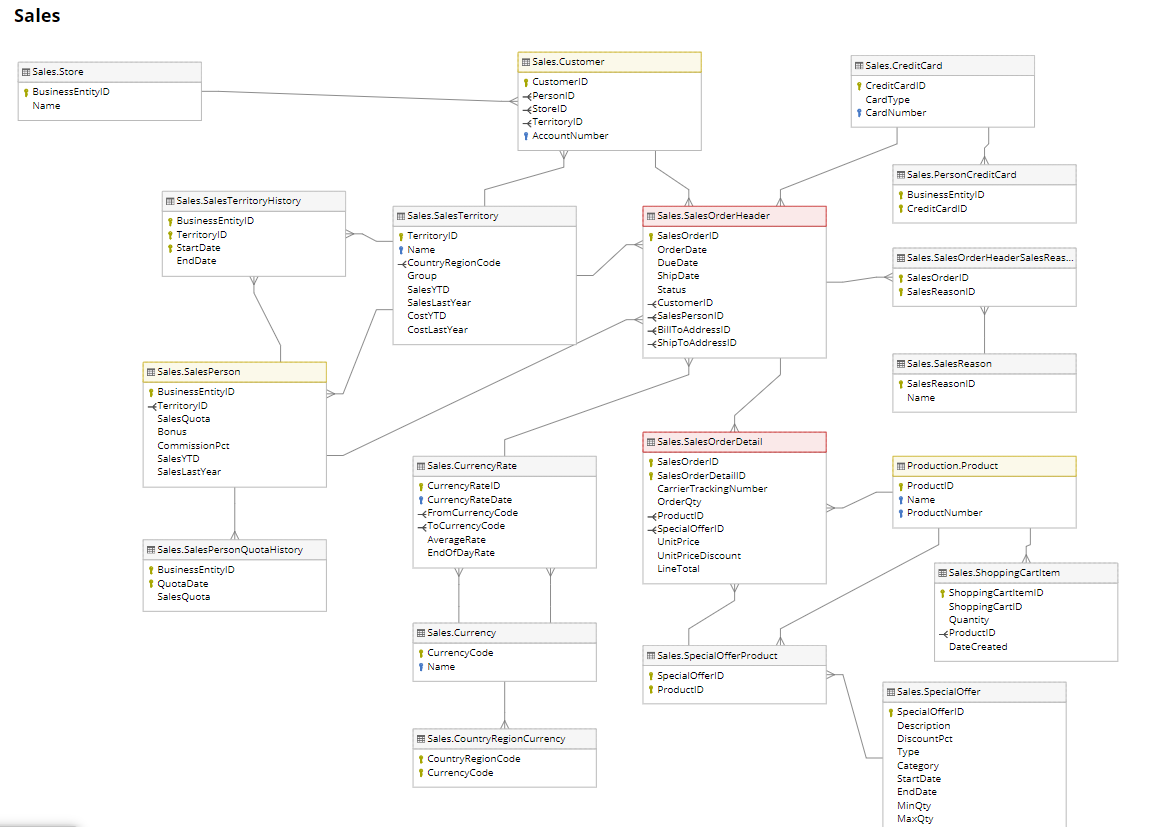


Figure 5: Module Sales

Sales module holds information about shopping cart, sales orders, special offers and sales people

Table 3: Sales

|  |  |  |
| --- | --- | --- |
| **STT** | **Table** | **Descriptive** |
| 1 | Sales.CountryRegionCurrency | Cross-reference table mapping ISO currency codes to a country or region. |
| 2 | Sales.CreditCard | Customer credit card information. |
| 3 | Sales.Currency | Lookup table containing standard ISO currencies. |
| 4 | Sales.CurrencyRate | Currency exchange rates. |
| 5 | Sales.Customer | Current customer information. Also see the Person and Store tables. |
| 6 | Sales.PersonCreditCard | Cross-reference table mapping people to their credit card information in the CreditCard table. |
| 7 | Sales.SalesOrderDetail | Individual products associated with a specific sales order. See SalesOrderHeader. |
| 8 | Sales.SalesOrderHeader | General sales order information. |
| 9 | Sales.SalesOrderHeaderSalesReason | Cross-reference table mapping sales orders to sales reason codes. |
| 10 | Sales.SalesPerson | Sales representative current information. |
| 11 | Sales.SalesPersonQuotaHistory | Sales performance tracking. |
| 12 | Sales.SalesReason | Lookup table of customer purchase reasons. |
| 13 | Sales.SalesTaxRate | Tax rate lookup table. |
| 14 | Sales.SalesTerritory | Sales territory lookup table |
| 15 | Sales.SalesTerritoryHistory | Sales representative transfers to other sales territories. |
| 16 | Sales.ShoppingCartItem | Contains online customer orders until the order is submitted or canceled. |
| 17 | Sales.SpecialOffer | Sale discounts lookup table. |
| 18 | Sales.SpecialOfferProduct | Cross-reference table mapping products to special offer discounts. |
| 19 | Sales.Store | Customers (resellers) of Adventure Works products. |

**2.4. Challenges**

To support the data analysis needs of the modules sales & HR, the company currently takes transactional data from the AdventureWorks2019 database, and non-transactional information such as sales quotas from spreadsheets, and consolidates this information into the relational data warehouse. However, the relational data warehouse presents the following challenges:

* Reports are static. Users have no way to interactively explore the data in the reports to obtain more detailed information, such as they could do with a Microsoft Office Excel pivot table. Although the existing set of predefined reports is sufficient for many users, more advanced users need direct query access to the database for interactive queries and specialized reports. However, because of the complexity of the AdventureWorks2019 database, too much time is needed for such users to learn how to create effective queries. However, after the construction of the DW is complete, visualization tools can be used to solve this problem.
* Query performance is widely variable. For example, some queries return results very quickly, in only a few seconds, while other queries take several minutes to return.
* Aggregate tables are difficult to manage. In an attempt to improve query response times, the data warehouse team at Adventure Works built several aggregate tables in the AdventureWorksDW2019 database. For example, they built a table that summarizes sales by month. However, while these aggregate tables greatly improve query performance, the infrastructure that they built to maintain the tables over time is fragile and prone to errors.
* It is difficult to combine related sets of information. Specialized queries that combine two sets of related information, such as sales and sales quotas, are difficult for business users to construct. Such queries overwhelmed the database, so the company requires that users request cross-subject-area sets of data from the data warehouse team. As a result, only a handful of predefined reports have been defined that combine data from multiple subject areas. Additionally, users are reluctant to try to modify these reports because of their complexity.

## Business Requirements Analysis of Sales & HR

The analysis requirements focus on the following 2 groups of topics: Sales & HR

* Sales : The main objective of this Sales topic data warehouse is to compare two sales methods Retailer Sales and Internet Sales based on total product sales. For Retailer Sales, by determining important criteria such as total sales by period, by product type, by geographic location, by salesperson, and by sales volume. Similarly, for Internet Sales, by determining important criteria such as total sales by period, by product type, by geographical location, by salesperson and by sales volume. The data warehouse provides essential information such as daily income, weekly revenue, monthly revenue, total revenue, goals, employee information and vision. To further clarify the Business requirement, a list of query questions has been listed as follows:
* Sort sales in ascending or descending order by years?
* Statistics of sales by product?
* Top 10 salespeople based on the revenue they sell by year?
* Top 10 most sold products on Internet Sales?
* HR: The main goal of this HR topic data warehouse is to make statistics and calculate the salary of employees by title, time and sales.

## IT requirements Analysis (IT &amp; Infrastructure) (optional)

# 

# CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

## Designing Data Warehouse:

### Bus Matrix:

A bus matrix is a consolidated tabular view that depicts the relationships between dimension and fact tables. A bus matrix enables you to identify easily which facts share the same dimension(s). By using this information, you can assess which tables can be combined into one report.[[11]](https://docs.genesys.com/Glossary:Bus_Matrix)

After defining business process, fact and common dimension tables, here are our bus matrix:

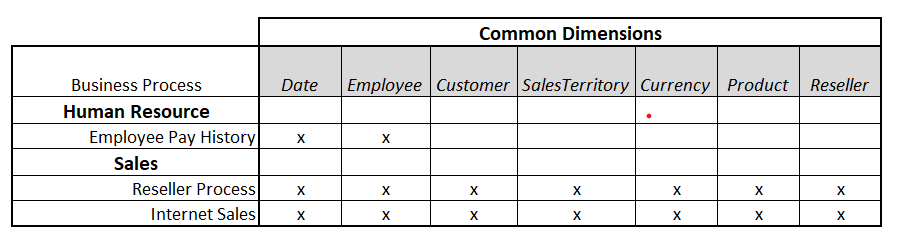


Figure 6: Bus matrix

### 1.2. Transactional and master data for fact and dimension tables:

Dim Currency:

Table 4: Dim Currency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **MasterData** | **Transactional Data** |
| CurrencyKey (PK) | Primary key for Currency | int |  |  |  |
| CurrencyAlternateKey |  | nvarchar |  | x |  |
| CurrencyName | Currency name | nvarchar | x | x |  |

Dim Customer:

Table 5: Dim Customer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Field Name | Descriptive | Type | Null | Master Data | Transaction Data |
| CustomerKey (PK) | Surrogate key from CustomerID | int |  |  |  |
| GeographyKey | Geograply Key | int | x | x |  |
| CustomerAlternateKey | Customer ID | int | x | x |  |
| Title | A courtesy title. For example, Mr. or Ms | nvarchar | x | x |  |
| FirstName | First name of customer | nvarchar | x | x |  |
| MiddleName | Middle name or middle initial of customer | nvarchar | x | x |  |
| LastName | Last name of customer | nvarchar | x | x |  |
| NameStyle | 0 = The data in FirstName and LastName are stored in western style (first name, last name) order. 1 = Eastern style (last name, first name) order | bit | x | x |  |
| BirthDate | Date of birth | date | x | x |  |
| MaritalStatus | M = Married, S = Single | nchar | x | x |  |
| Suffix | Surname suffix. For example, Sr. or Jr | nvarchar | x | x |  |
| Gender | M = Male, F = Female | nvarchar | x | x |  |
| EmailAddress | E-mail address for the customer | nvarchar | x | x |  |
| YearlyIncome | USD | money | x | x |  |
| TotalChildren | Total children of Customer | tinyint | x | x |  |
| NumberChildrenAtHome | Number of children of customer | tinyint | x | x |  |
| NumberCarsOwned | Number of cars of customer | nvarchar | x | x |  |
| AddressLine1 | First street address line | nvarchar | x | x |  |
| Phone | Telephone number identification number | nvarchar | x | x |  |
| DateFirstPurchase | Date first purchase date | date | x | x |  |
| CommuteDistance |  | nvarchar | x | x |  |

Dim Date:

Table 6: Dim Date

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| DateKey(PK) | Surrogate key for date | int |  |  |  |
| FullDateAlternateKey | Date business key | date |  |  | x |
| DayNumberOfWeek | Day number of week | tinyint | x | x |  |
| EnglishDayNameOfWeek | English day name of week | nvarchar | x | x |  |
| DayNumberOfMonth | Day number of month | tinyint | x | x |  |
| DayNumberOfYear | Day number of year | smallint | x | x |  |
| WeekNumberOfYear | Week number of year | tinyint | x | x |  |
| EnglishMonthName | English month name | nvarchar | x | x |  |
| MonthNumberOfYear | Month number of year | tinyint | x | x |  |
| CalendarQuarter | Calendar quarter | tinyint | x | x |  |
| CalendarYear | Calendar year | smallint | x | x |  |
| CalendarSemester | Calendar semester | tinyint | x | x |  |
| FiscalQuarter | Fiscal quarter | tinyint | x | x |  |
| FiscalYear | Fiscal year | smallint | x | x |  |
| FiscalSemester | Fiscal semester | tinyint | x | x |  |

Dim Employee:

Table 7: Dim Employee

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| EmployeeKey (PK) | Surrogate key for EmployeeID | int |  |  |  |
| EmployeeNationalIDAlternateKey | Employee National ID | nvarchar | x | x |  |
| FirstName | First name of employee | nvarchar | x | x |  |
| LastName | Last name of employee | nvarchar | x | x |  |
| Title | Work title such as Buyer or Sales Representative | nvarchar | x | x |  |
| HireDate | Employee hired on this date | date | x | x |  |
| BirthDate | Date of birth | date | x | x |  |
| EmailAddress | E-mail address for the person | nvarchar | x | x |  |
| Phone | Telephone number identification number | nvarchar | x | x |  |
| MaritalStatus | M=Married, S=Single | nchar | x | x |  |
| Department | Department in which the employee worked including currently | nvarchar | x | x |  |
| StartDate | Date the employee started work in the department | date | x | x |  |
| EndDate | Date the employee left the department. NULL = Current department | date | x | x |  |
| Status |  | nvarchar | x | x |  |

Dim Product:

Table 8: Dim Product

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| ProductKey (PK) | Surrogate key for Product records | int |  |  |  |
| ProductAlternateKey | ProductID | nvarchar | x | x |  |
| EnglishProductName | Product name | nvarchar |  | x |  |
| StandardCost | Standard cost of the product | money | x | x |  |
| Color | Product color | nvarchar |  | x |  |
| Size | Product size | nvarchar | x | x |  |
| SizeRange | Size range for product | nvarchar | x | x |  |
| EnglishDescription | Description | nvarchar | x | x |  |
| FrenchDescription | Description | nvarchar | x | x |  |
| ArabicDescription | Description | nvarchar | x | x |  |
| ProductSubcategoryCode | Subcategory code | nvarchar | x | x |  |
| ProductcategoryCode | Category code | nvarchar | x | x |  |
| Status |  | nvarchar | x | x |  |

Dim Reseller:

Table 9: Dim Reseller

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| ResellerKey (PK) | Surrogate key for ResellerID | int |  |  |  |
| ResellerAlternateKey | Reseller ID | nvarchar | x | x |  |
| ResellerName | Name of reseller | nvarchar | x | x |  |
| NumberEmployees | Number of employee in reseller | int | x | x |  |
| YearOpened |  | int | x | x |  |

Dim Sales Territory:

Table 10: Dim Sales Territory

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| SalesTerritoryKey (PK) | Surrogate key for Sales Territory | int |  | x |  |
| SalesTerritoryAlternat-eKey | Sales Territory Key | int | x | x |  |
| SalesTerritoryRegion | Territory region | nvarchar | x | x |  |
| SalesTerritoryCountry | Territory country | nvarchar | x | x |  |
| SalesTerritoryGroup | Geographic area to which the sales territory belong | nvarchar | x | x |  |
| SalesTerritoryImage | Territory image | varbinary | x | x |  |

Fact Employee Pay:

Table 11: Fact Employee Pay

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| EmployeeKey (FK) | Foreign Key from DimEmployee | int |  |  |  |
| DateKey(FK) | Foreign Key from DimDate | int |  |  |  |
| Pay Frequency | 1 = Salary received monthly, 2 = Salary received biweekly | nvarchar | x | x |  |
| Rate | Salary hourly rate | nvarchar | x | x |  |

Fact Internet Sales:

Table 12: Fact Internet Sales

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| ProductKey(FK) |  | int |  |  |  |
| OrderDateKey(FK) |  | int |  |  |  |
| DueDatKey(FK) |  | int |  |  |  |
| ShipDateKey(FK) |  | int |  |  |  |
| CustomerKey(FK) |  | int |  |  |  |
| CurrencyKey(FK) |  | int |  |  |  |
| SalesTerritoryKey(FK) |  | int |  |  |  |
| SalesOrderNumber(PK) |  | nvarchar |  |  | x |
| SalesOrderLineNum(PK)ber |  | tinyint |  |  | x |
| OrderQuantity |  | smallint | x | x |  |
| UnitPrice | Selling price of a single product | money | x | x |  |
| DiscountAmount | Discount amount on a order | float | x | x |  |
| SalesAmount | Sales total for a order | money | x | x |  |
| TaxAmt | Tax amount | money | x | x |  |
| OrderDate | Dates the sales order was created | datetime | x |  | x |
| DueDate | Date the order is due to the customer | datetime | x |  | x |
| ShipDate | Date the order was shipped to the customer | datetime | x |  | x |

Fact Reseller Sales:

Table 13: Fact Reseller Sales

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Descriptive** | **Type** | **Null** | **Master Data** | **Transaction Data** |
| ProductKey(FK) |  | int |  |  |  |
| OrderDateKey(FK) |  | int |  |  |  |
| DueDatKey(FK) |  | int |  |  |  |
| ShipDateKey(FK) |  | int |  |  |  |
| ResellerKey(FK) |  | int |  |  |  |
| EmployeeKey(FK) |  |  |  |  |  |
| CurrencyKey(FK) |  | int |  |  |  |
| SalesTerritoryKey(FK) |  | int |  |  |  |
| SalesOrderNumber(PK) |  | nvarchar |  |  | x |
| SalesOrderLineNumber(PK) |  | tinyint |  |  | x |
| OrderQuantity | Order quantity of a order | smallint |  | x |  |
| UnitPrice | Unit price on a product | money |  | x |  |
| DiscountAmount | Discount amount on the order | float |  | x |  |
| SalesAmount | Sales total for the order | money |  | x |  |
| OrderDate | Order date | datetime |  |  | x |
| DueDate | Due date | datetime |  |  | x |
| ShipDate | Ship date | datetime |  |  | x |

### 1.3. Data Warehouse model (Snowflake or Star):

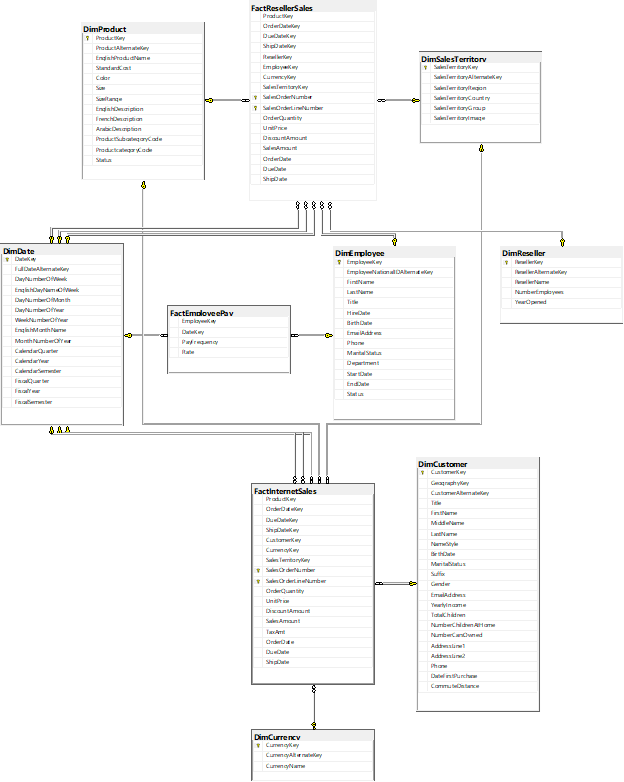


Figure 7: Relational model

When it comes to the relational model, our project decided to implement a galaxy schema. A galaxy schema (or Fact Constellation Schema) can be thought as star schema with fully interlinking and normalization which can prevent data redundancy and inconsistency.[[6]](https://ieeexplore.ieee.org/abstract/document/7848419)

## ETL processes

In our process in order to load data to the warehouse we use SQL Server Integration Service(SSIS) to support us. Following above ETL process in section 2, our works are illustrated below:

* Load data from AdventureWorks source to data staging.



Figure 8: Source to staging process

This staging area is non-relational modeling in which each table doesn’t have a primary key or foreign key between these tables.

* Load data from the staging table to the data warehouse. For the loading dimension table we write SQL procedures to load data for each dimension. With fact table, in order to loading data we use lookup table for business key of with each dimension.
* Load Currency Dimension

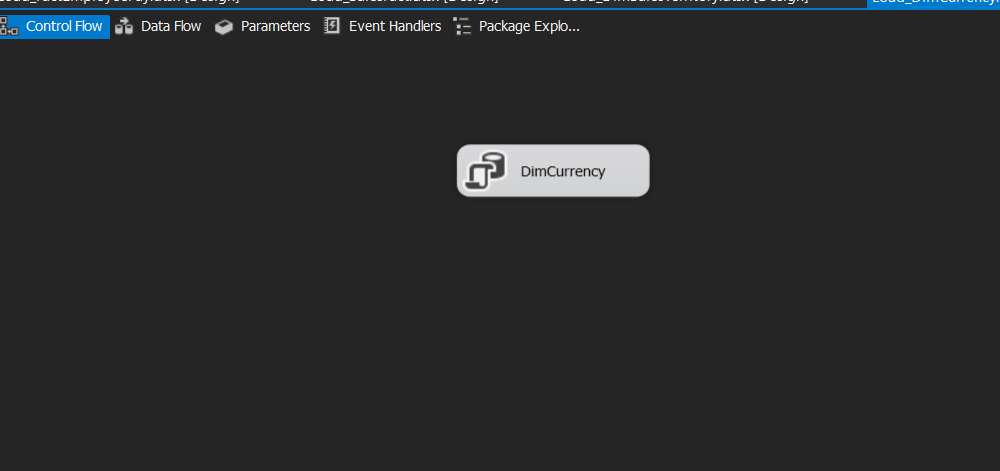


Figure 9: Load Dimension for Currency

* Load Customer Dimension

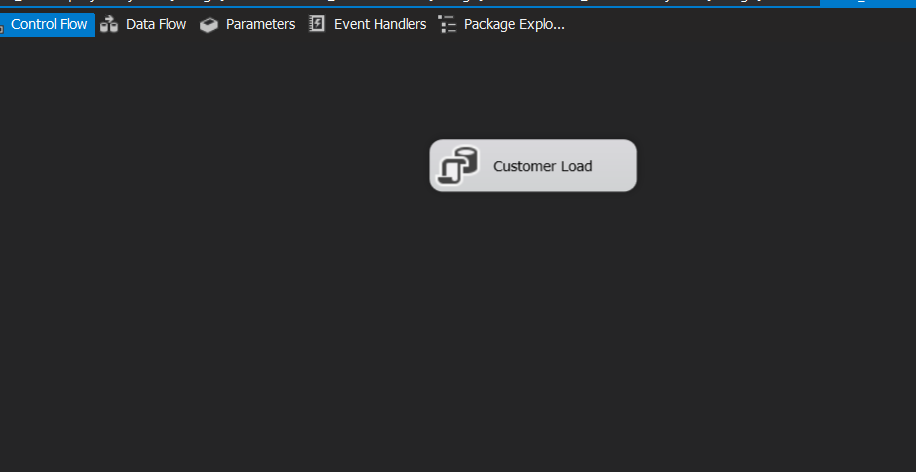


Figure 10: Load Dimension for Customer

* Load Employee Dimension

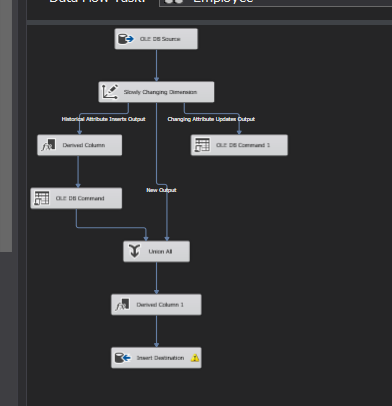


Figure 11: Load Dimension for Employee

* Load Date Dimension

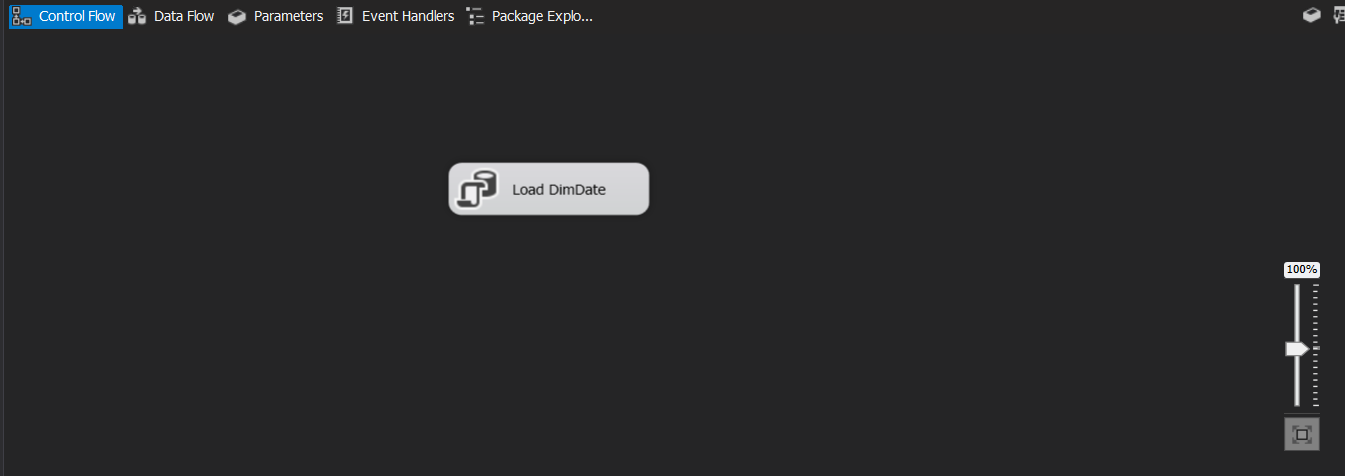


Figure 12: Load Dimension for Date

* Load Product Dimension

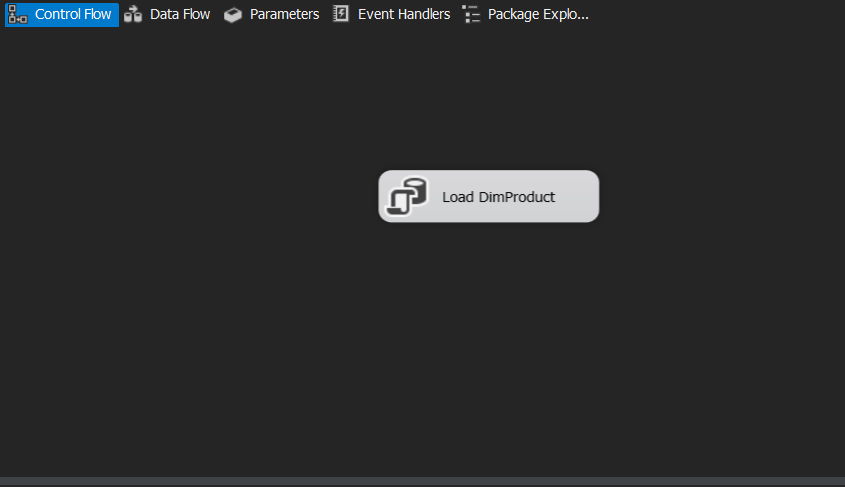


Figure 13: Load Dimension for Product

* Load Reseller Dimension

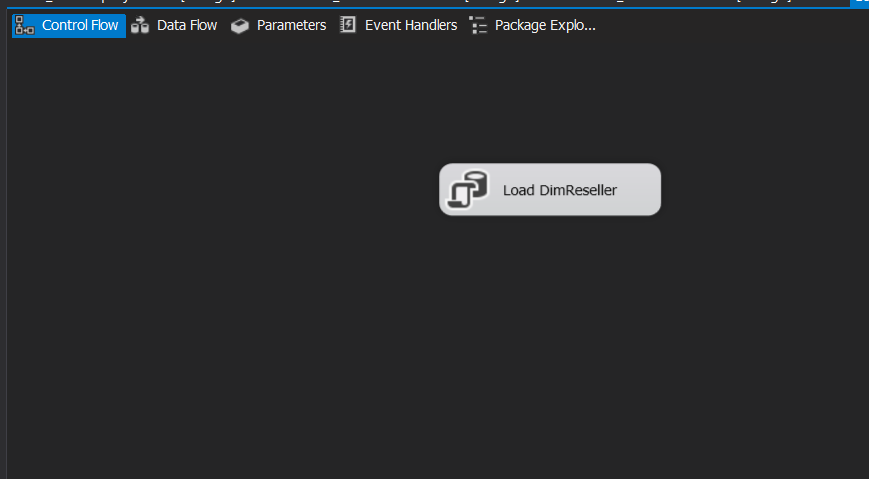


Figure 14: Load Dimension for Reseller

* Load Sales Territory

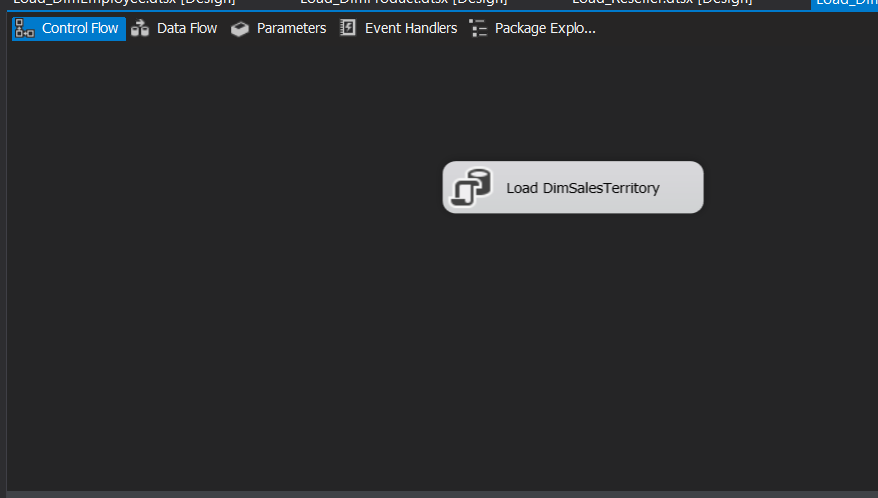


Figure 15: Load Dimension for Sales Territory

* Load Internet Sales Fact

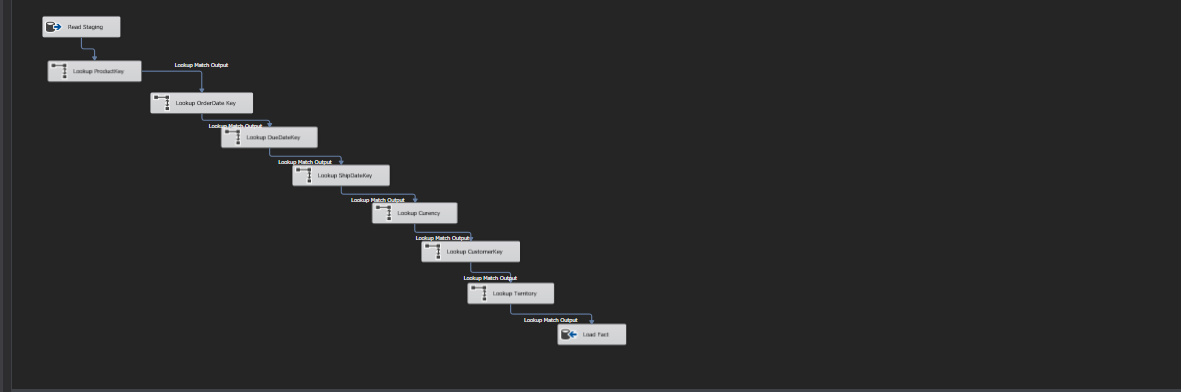


Figure 16: Load Fact Internet Sales

* Load Reseller Sales Fact

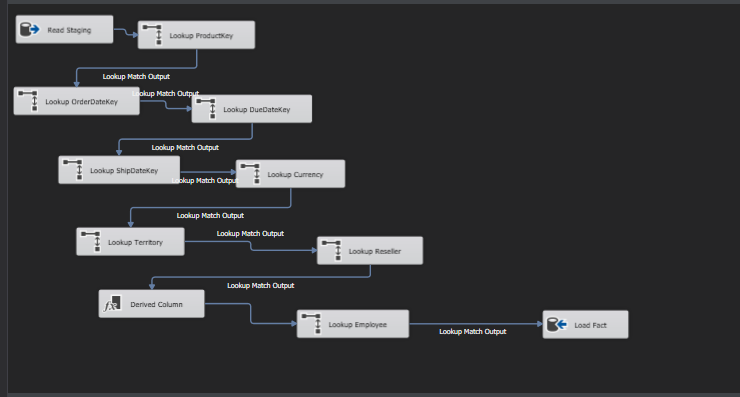


Figure 17: Load Fact Reseller Sales

* Load Employee Pay Fact

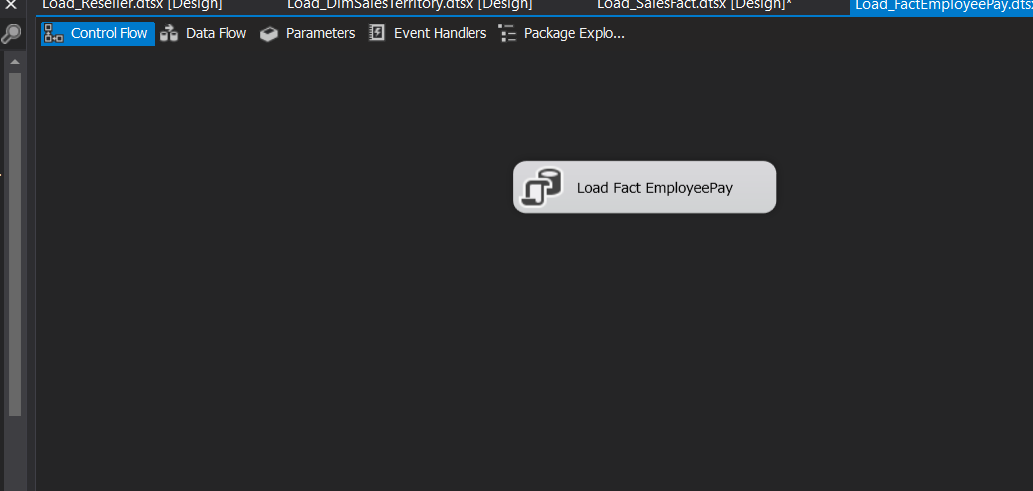


Figure 18: Load Fact Employee Pay

### 2.1. Dimension Table’s ETL Process

* Currency Dimension



Figure 19: Data in Currency Dimension

* Customer Dimension

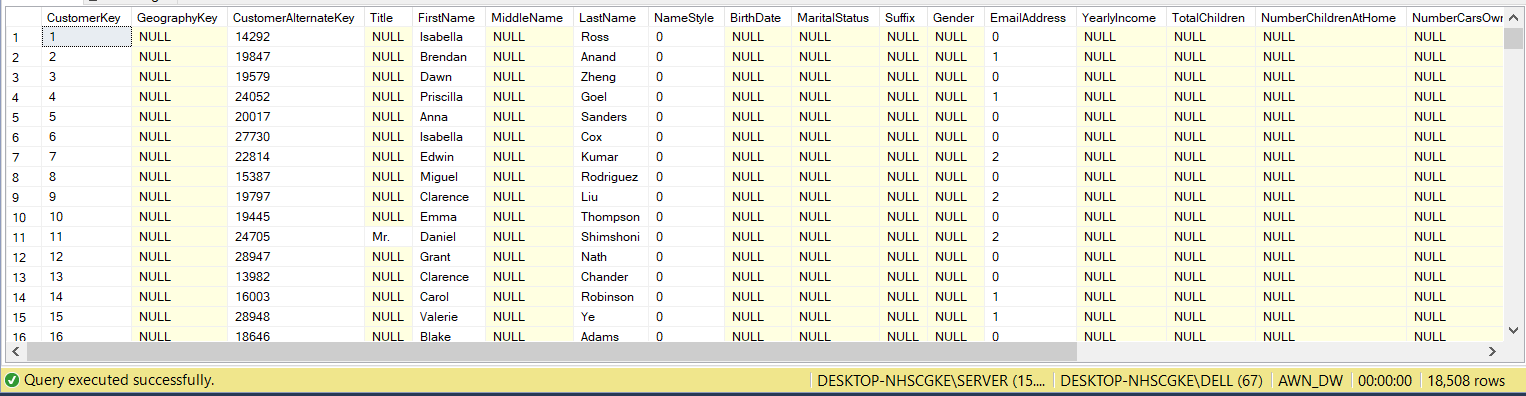


Figure 20: Data in Customer Dimension

* Date Dimension

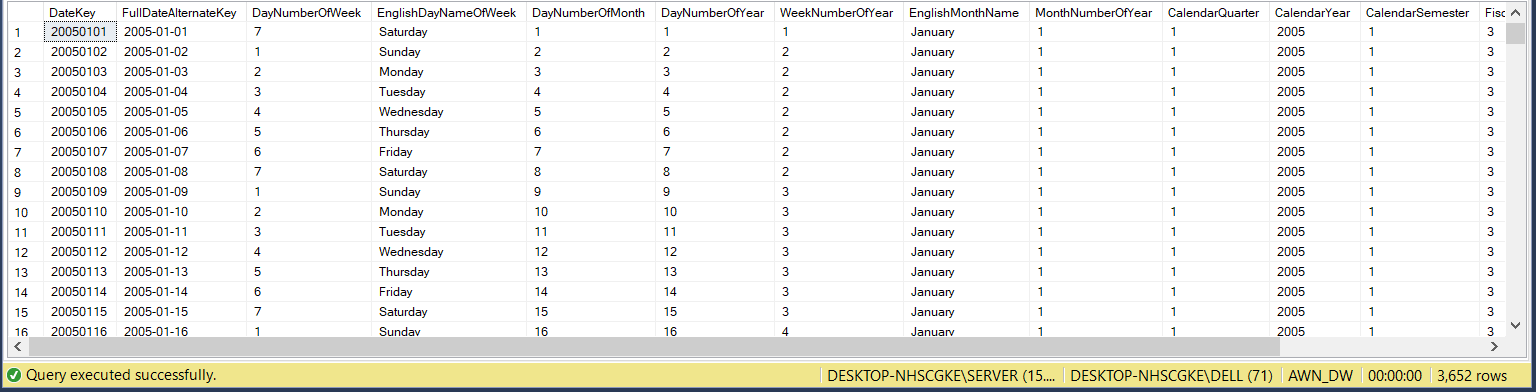


Figure 21: Data in Date Dimension

* Employee Dimension

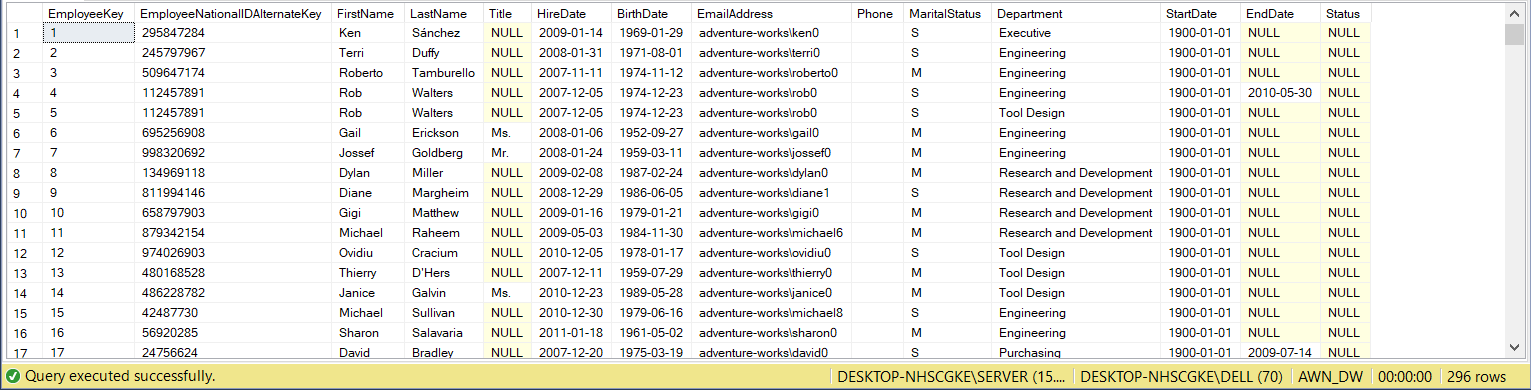


Figure 22: Data in Employee Dimension

* Product Dimension

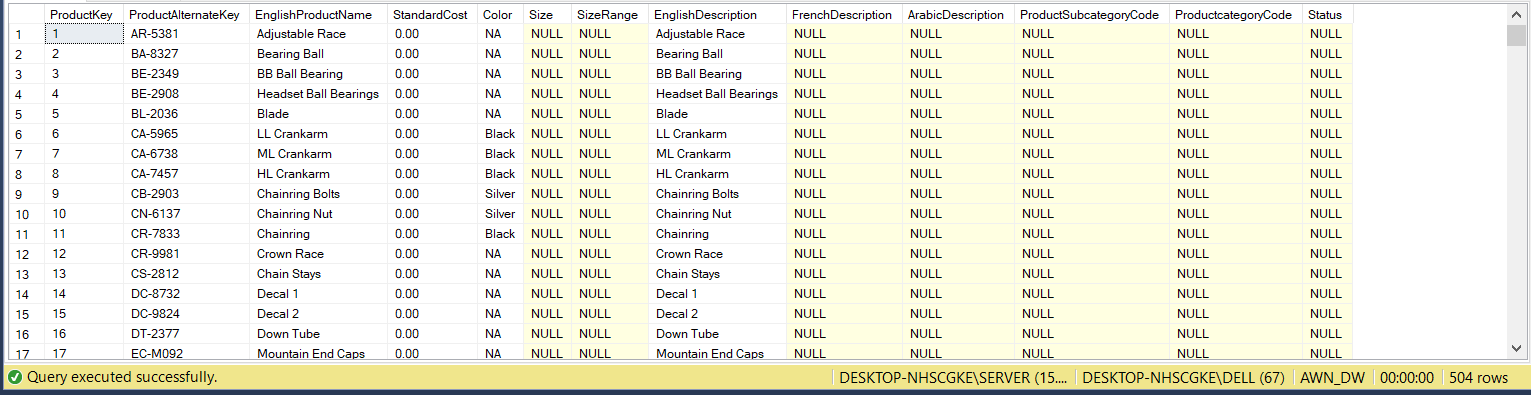


Figure 23: Data in Product Dimension

* Sales Territory Dimension

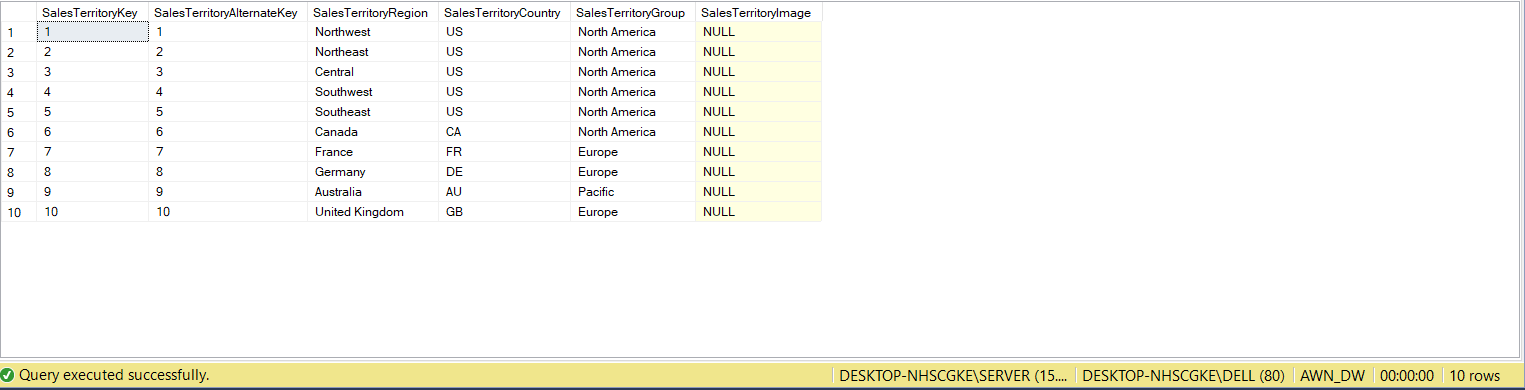


Figure 24: Data in Sales Territory Dimension

* Reseller Dimension

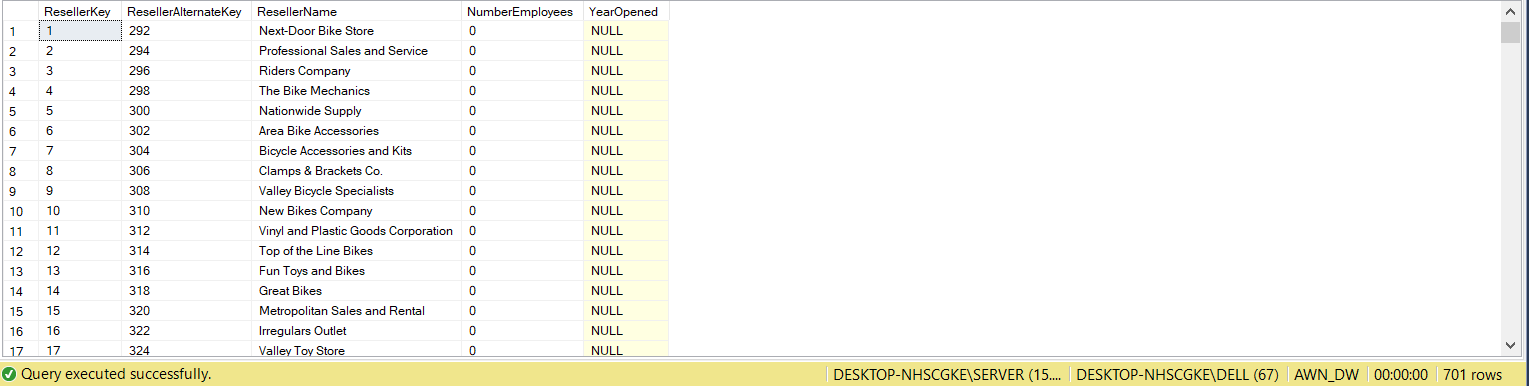


Figure 25: Data in Reseller Dimension

### 2.2. Fact Table’s ETL Process

* Internet Sales Fact

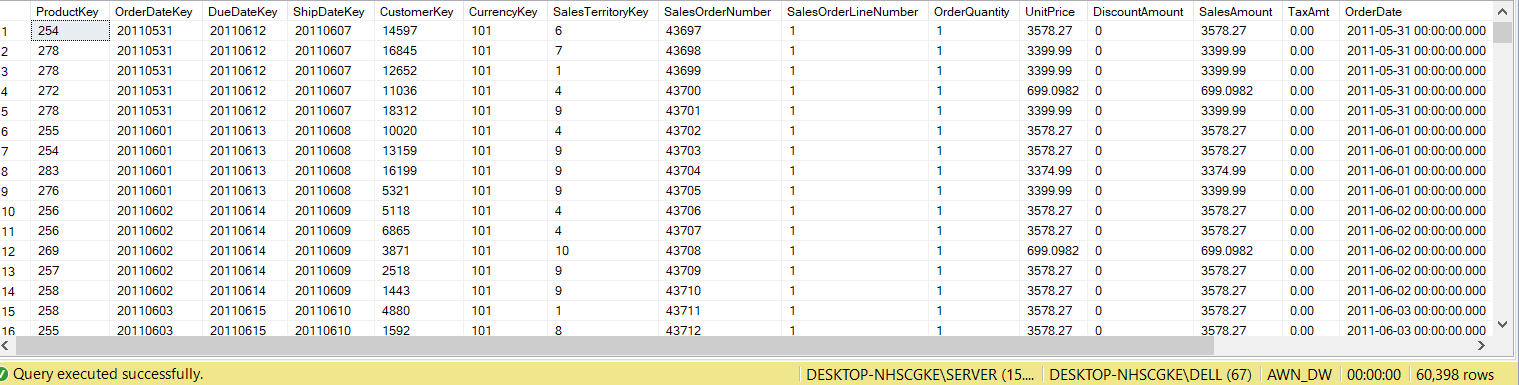


Figure 26: Data in Internet Sales Fact

* Reseller Sales Fact

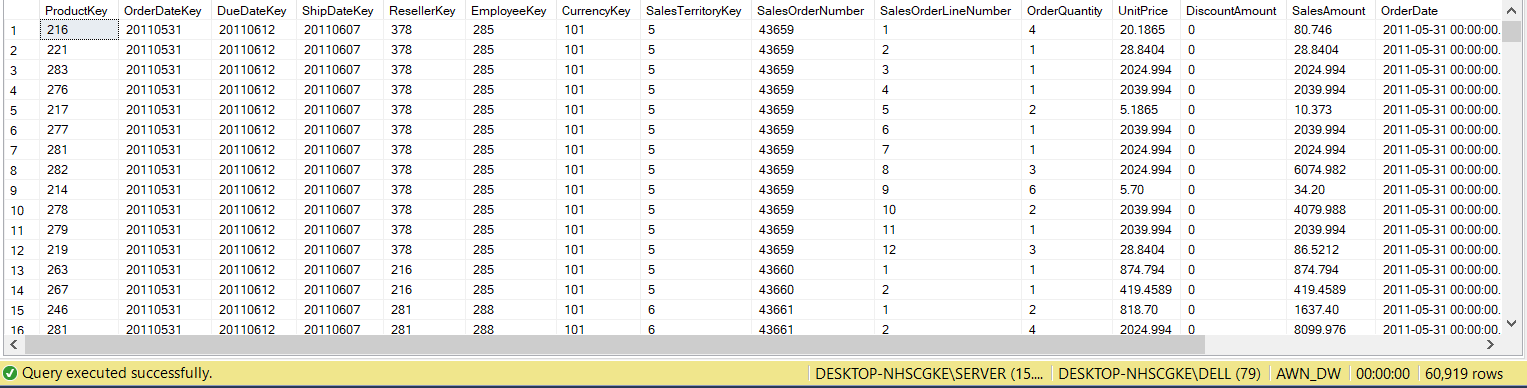


Figure 27: Data in Reseller Sales Fact

* Employee Pay Fact



Figure 28: Data in Employee Pay Fact

# 

# CHAPTER 5: CONCLUSION AND FUTURE WORKS

## Results:

After the ETL process, data from the staging area is all loaded into the warehouse.

We decided to connect our data warehouse to Tableau for further analysis instead of building cube on SSAS then deploying to SSRS because Tableau supports better, more modern and lively visualization.

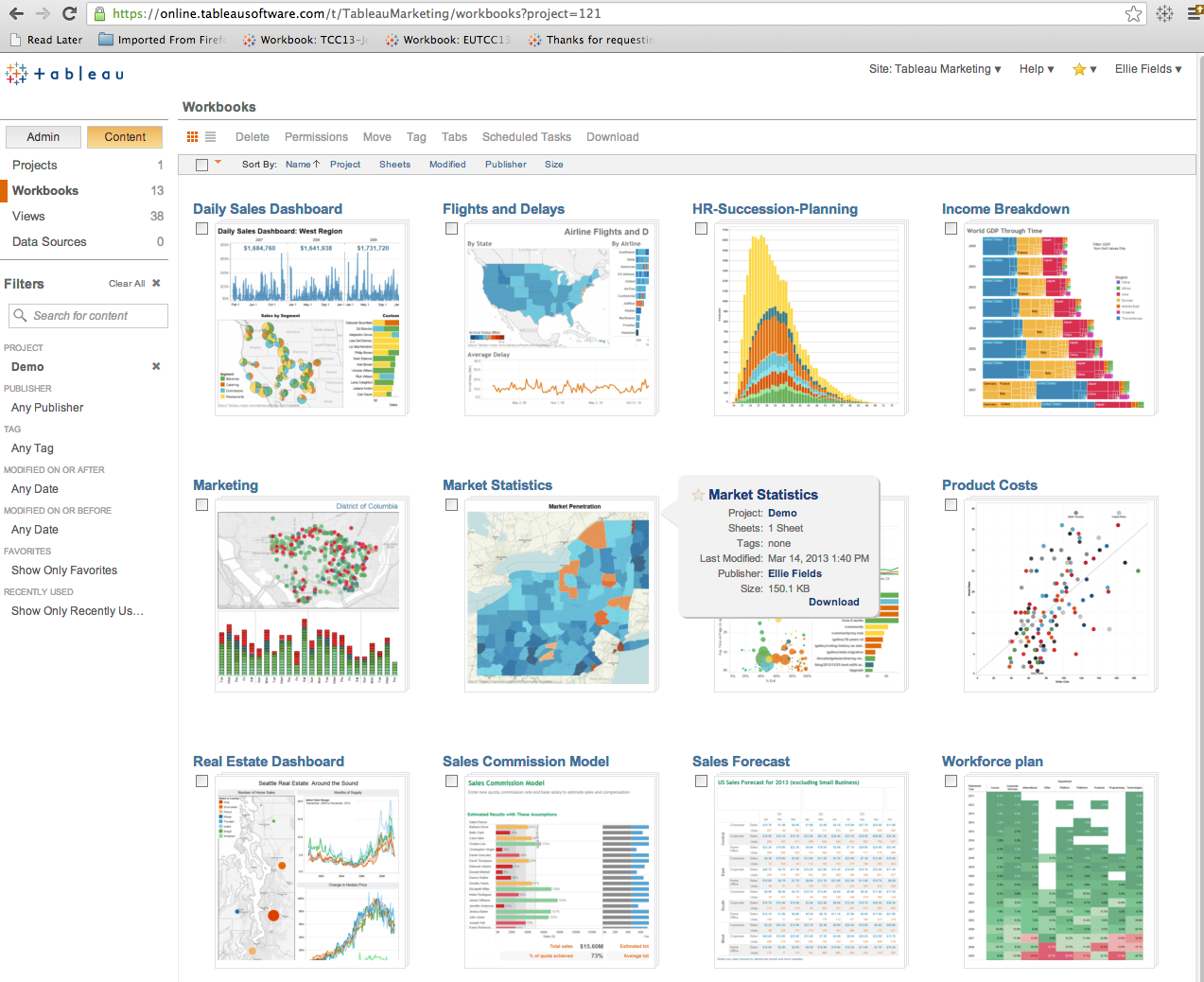


Figure 29: Tableau BI tool

Here are some demo analytics that we got from the built data warehouse:

* Sales amount by countries:

This company is based in the US but has a worldwide business activity.

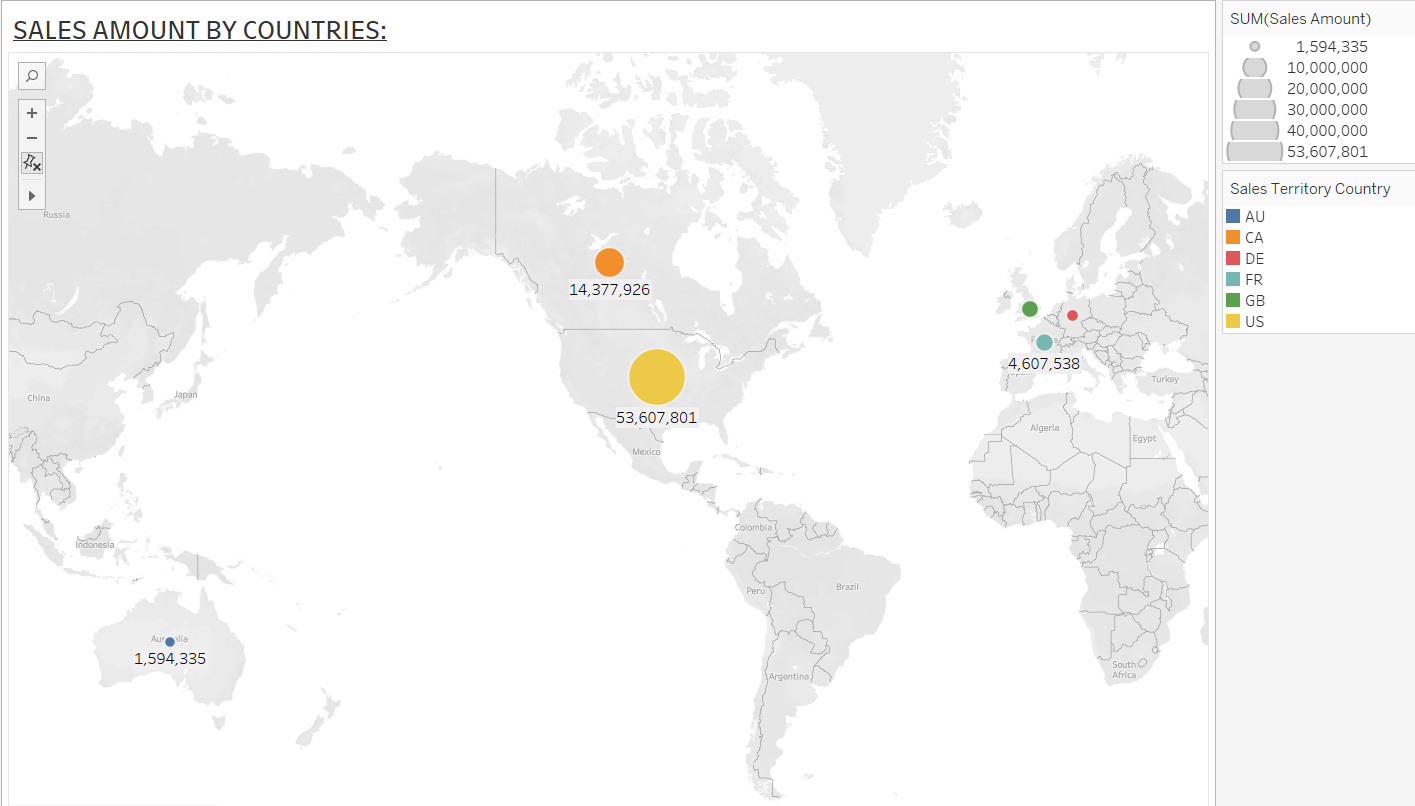


Figure 30: Reseller sales by countries.

As we can see, North America takes the dominance in term of sales amount while Europe, Australia account only a small segment.

* Sales over time:

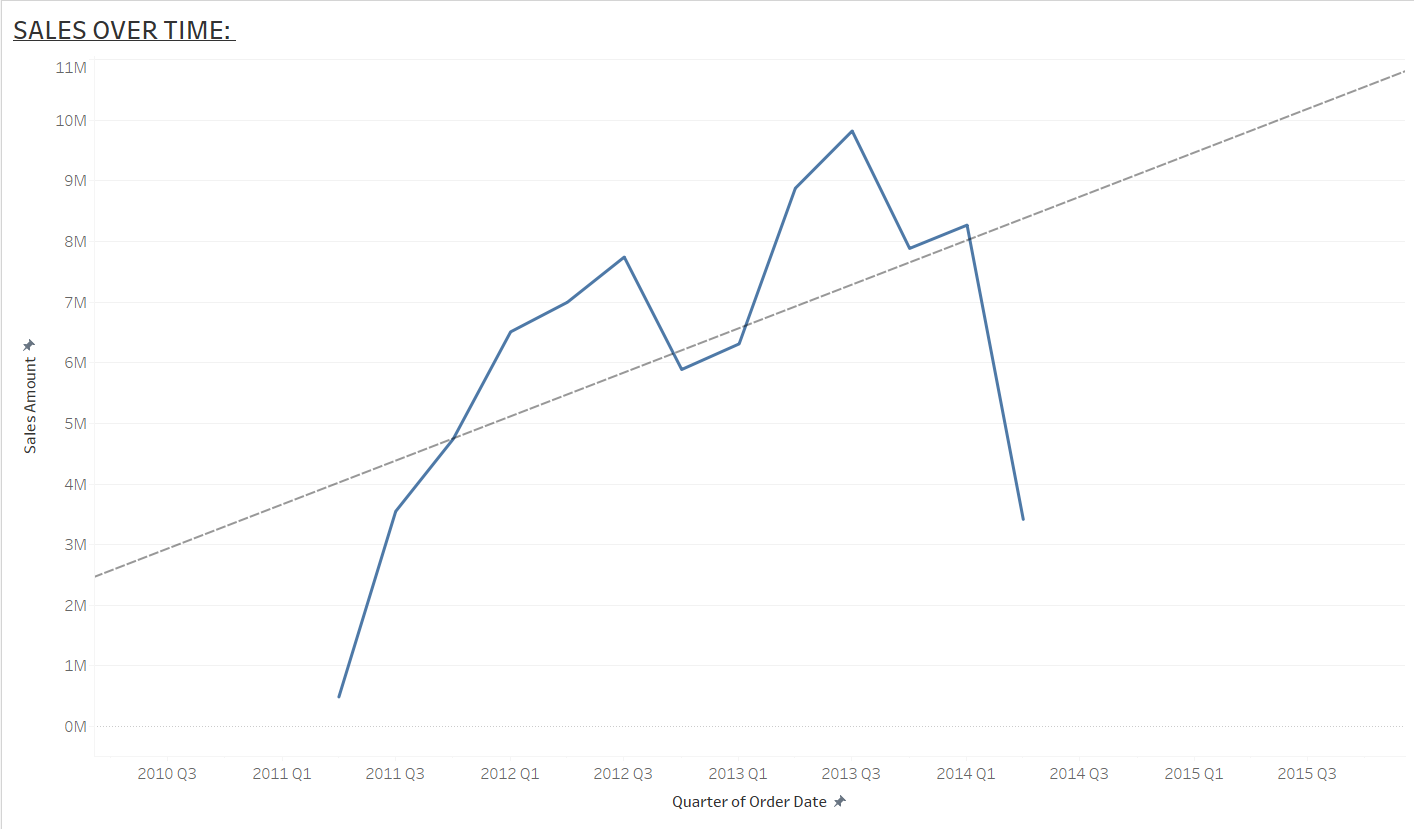
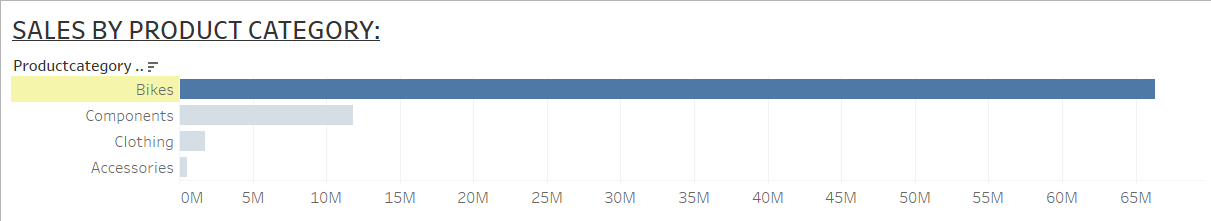
**

Figure 31: Sales over time (quarterly).

* Comparing sales amount through time between Reseller and Internet sales

**

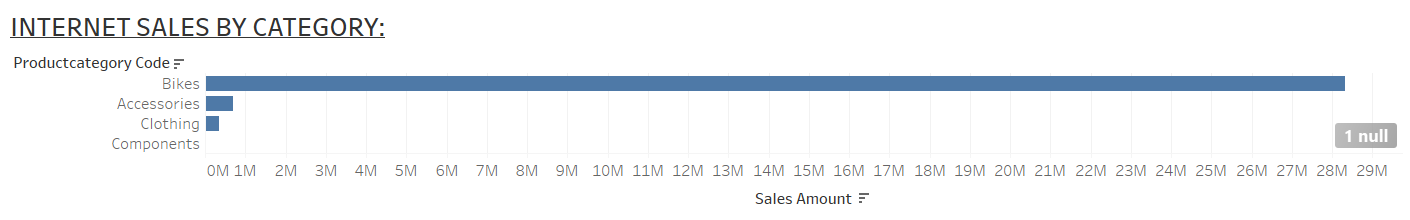
**

Figure 32: Compare sales amount of Reseller and Internet by product category.

* Figure out top 10 salesman who have best performance in the company:

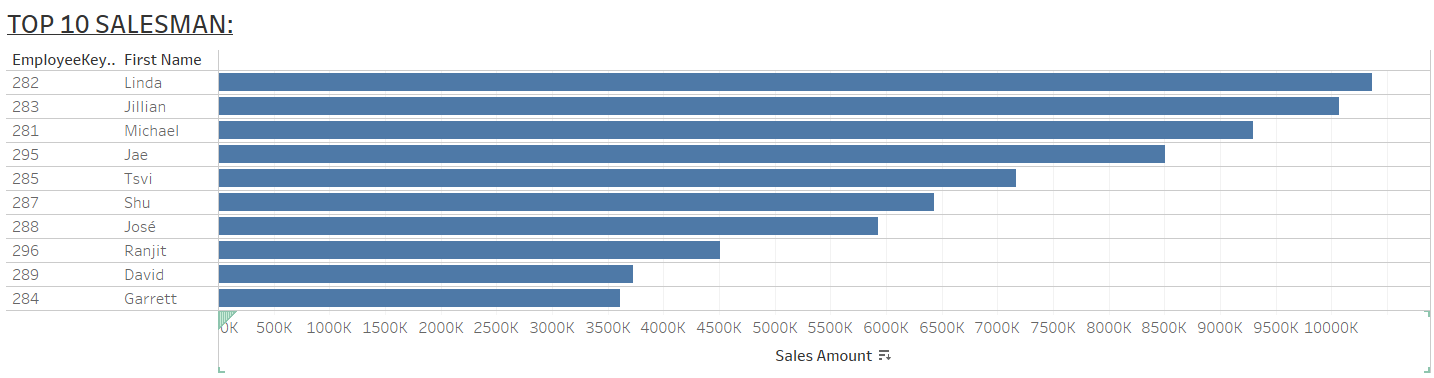
**

Figure 33: Top 10 salesman

## Limitations

One of the biggest challenges is that we do not have a Tableau Prep Builder for a thorough ETL process as Tableau Prep Builder provides users with a strong automatic ETL schedule tool. So that we have to ETL with SSIS first, then connect manually to Tableau Desktop to analyze data.

Tableau does not allow joining the second Fact with already joined Dim tables like figure below:

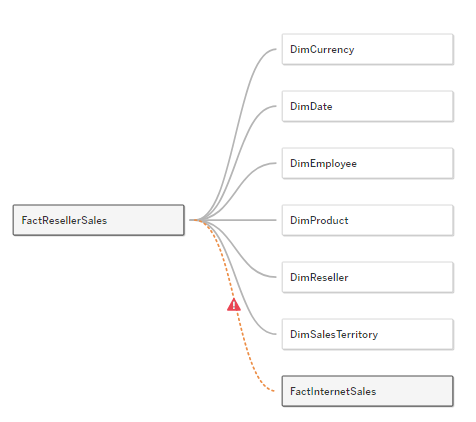


Figure 34: Fact - Dim connection in Tableau example.

So our choice is to create different data sources. We considered this not a big problem. Instead, we find it a little convenient to analyze data.

Last of all is our shortage of knowledge and experience in implementing a complete data warehouse so mistakes are inevitable.

## Future works

In the short future, we would like to rebuild our project and scale up the data warehouse for not only Sales, HR but also other departments with professor instruction.

Our intention is to work with Purchasing and Production schema. Similar to Sales and Hr, Purchasing and Production plays an important role in business. That’s the reason why we choose those schemes for the next work.

# REFERENCES

[1] [Mouhib Alnoukari](https://siust.academia.edu/MouhibAlnoukari?swp=rr-ac-190451) (2008), Using Data Mining for Predicting Future Car Market Demand: DCX Case Study,

[2] Dejen Sarka & Matija Lah & Grega Jerkic (2012), Implementing a Data Warehouse with Microsoft SQL, Training Kit.

[3] Nguyễn Văn Trung (2008, Trường Đại học Công nghệ) Thiết kế và triển khai kho dữ liệu khách hàng sử dụng dịch vụ viễn thông của Tổng Công ty Bưu chính Viễn thông Việt Nam (VNPT)

[4] Piotr Muryjas & Monika Wawer & Magdalena Rzemieniak (2021, European Research Studies JournalVolume XXIV, Special Issue 2, 2021), Managing the Process of Evaluation of the Academic Teachers with the Use of Data Mart and Business Intelligence.

[5] Ralph Kimball & Margy Ross (2013), The Data Warehouse Toolkit, The Dei nitive Guide to Dimensional Modeling – Third Edition.

[6] Harco Leslie Hendrie Spits Warners; Richard Randriatoamanana

Datawarehouser: A data warehouse artist who have ability to understand data warehouse schema pictures.

# APPENDIX

[1] Education, I. (2020). What is a Data Warehouse?. Retrieved 21 December 2022, from <https://www.ibm.com/cloud/learn/data-warehouse>

[2] What is a Data Warehouse?. (2022). Retrieved 21 December 2022, from <https://www.oracle.com/database/what-is-a-data-warehouse>

[3] Montgomery, L. (2019). The Benefits of Business Intelligence in Enterprise Environments. Retrieved 21 December 2022, from <https://mytechdecisions.com/compliance/benefits-business-intelligence-enterprise/>

[4] (2022). Retrieved 21 December 2022, from <https://www.talend.com/resources/what-is-data-mart/>

[5] Star Schema vs Snowflake Schema - javatpoint. (2022). Retrieved 21 December 2022, from <https://www.javatpoint.com/data-warehouse-star-schema-vs-snowflake-schema>

[6] Education, I. (2020). What is ETL (Extract, Transform, Load)?. Retrieved 21 December 2022, from <https://www.ibm.com/cloud/learn/et>

[7] Tuấn Thanh, Sales Department là gì? Tìm hiểu về Sales Department, tino.org, <https://wiki.tino.org/sales-department-la-gi/>

[8] Misa.amis (2021), HR department là gì? Vai trò của người làm HR trong thời đại 4.0, <https://amis.misa.vn/30711/hr-department-la-gi/>

[9] Analysis Services Tutorial Scenario, <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/analysis-services-tutorial-scenario?view=asallproducts-allversions>

[10] AdventureWorks Data Dictionary (2017) <https://dataedo.com/download/AdventureWorks.pdf>

[11] Bus matrix definition,

<https://docs.genesys.com/Glossary:Bus_Matrix>