**UNIVERSITY OF ECONOMICS AND LAW**

**FACULTY OF INFORMATION SYSTEMS**



**FINAL PROJECT REPORT**

TOPIC: **USING BUSINESS INTELLIGENCE SOLUTIONS TO SOLVE PROBLEMS RELATED TO THE PODUCTION MODULE**

Lecturer: **Nguyen Van Ho, Le Ba Thien**

Group: **Simpsons**

Ho Chi Minh City, May 09, 2023

**Members of Group Simpsons**

|  |  |  |
| --- | --- | --- |
| **No.** | **NAME** | **STUDENT ID** |
| 1 | Nguyễn Thị Mỹ Dung | K204061390 |
| 2 | Nguyễn Thị Bảo Hà | K204061394 |
| 3 | Nguyễn Thị Cẩm Giang | K204060282 |
| 4 | Cù Thị Mỹ Duy | K204061392 |
| 5 | Nguyễn Mỹ Dung | K204061389 |

**Table of Contents**

[CHAPTER 1: INTRODUCTION 11](#_Toc134561106)

[1.1. Business case for the project 11](#_Toc134561107)

[1.2. Objectives of the project 11](#_Toc134561108)

[1.2.1. General Objective 11](#_Toc134561109)

[1.2.2. Specific Objective 11](#_Toc134561110)

[1.3. Research Objects 11](#_Toc134561111)

[1.4. Scope of the project 12](#_Toc134561112)

[1.5. Value and desired outcome of the project 12](#_Toc134561113)

[1.6. Structure of project 12](#_Toc134561114)

[CHAPTER 2: THEORETICAL BASIS 13](#_Toc134561115)

[2.1. Overview about BI 13](#_Toc134561116)

[2.1.1. What is BI? 13](#_Toc134561117)

[2.1.2. BI Architecture 13](#_Toc134561118)

[Figure 2 1. BI Architecture (Source: ResearchGate.net) 14](#_Toc134561119)

[2.1.3. Advantage of BI in enterprises 15](#_Toc134561120)

[2.1.4. BI Strategy for Business 15](#_Toc134561121)

[2.2. ETL Process 15](#_Toc134561122)

[2.2.1. What is ETL? 15](#_Toc134561123)

[2.2.2. Why do we need ETL? 16](#_Toc134561124)

[2.2.3. ETL Process 16](#_Toc134561125)

[2.3. Data warehouse and Data mart 16](#_Toc134561126)

[2.3.1. What are Data warehouse and Data mart? 16](#_Toc134561127)

[2.3.2. Who needs Data warehouse and Data mart? 17](#_Toc134561128)

[2.3.3. Advantages and disadvantages of Data warehouse 17](#_Toc134561129)

[2.3.4. Snowflake and Star schemas 18](#_Toc134561130)

[Figure 2 2. Snowlake (Source: ierp service JSC) 18](#_Toc134561131)

[2.3.5. Star schema 19](#_Toc134561132)

[2.4. KPIs 19](#_Toc134561133)

[2.4.1. KPIs Definition 19](#_Toc134561134)

[2.4.2. The advantages and disadvantages of KPIs 20](#_Toc134561135)

[2.4.3. Categories of KPIs (Production) 21](#_Toc134561136)

[Table 2 1. KPIs ( source: author) 21](#_Toc134561137)

[2.5. Azure Analysis Service (Optional) 22](#_Toc134561138)

[Figure 2 4. Azure Analysis Service (Optional) 22](#_Toc134561139)

[CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION 24](#_Toc134561140)

[3.1. Business processes 24](#_Toc134561141)

[3.1.1. Production department 24](#_Toc134561142)

[3.1.2 The purpose of production process 24](#_Toc134561143)

[3.1.3 The stages of production process 24](#_Toc134561144)

[3.2. Data source and challenges 25](#_Toc134561145)

[3.2.1. Data source 25](#_Toc134561146)

[Table 3 1. Shows the list of tables stored in the production module of Adventure Works 2014.(Source: AdventureWorks) 25](#_Toc134561147)

[Figure 3 2. Data source 28](#_Toc134561148)

[Table 3 2. Table Production.ProductCategory(Source: AdventureWorks) 28](#_Toc134561149)

[Table 3 3. Table Production.ProductModelIllustration (Source: AdventureWorks) 29](#_Toc134561150)

[Table 3 4. Table Production.Illustration(Source: AdventureWorks) 29](#_Toc134561151)

[Table 3 5. Table Production.ProductSubcategory(Source: AdventureWorks) 30](#_Toc134561152)

[Table 3 6. Table Production.ProductModel(Source: AdventureWorks) 30](#_Toc134561153)

[Table 3 7. Table Production.ProductModelProductDescriptionCulture(Source: AdventureWorks) 31](#_Toc134561154)

[Table 3 8. Table Production.ProductDescription(Source: AdventureWorks) 32](#_Toc134561155)

[Table 3 9. Table Production.UnitMeasureCode(Source: AdventureWorks) 32](#_Toc134561156)

[Table 3 10. Table Production.Culture(Source: AdventureWorks) 33](#_Toc134561157)

[Table 3 11. Table Production.ProductReview(Source: AdventureWorks) 33](#_Toc134561158)

[Table 3 12. Table Production.Productdocument(Source: AdventureWorks) 34](#_Toc134561159)

[Table 3 13. Table Production.ProductPhoto(Source: AdventureWorks) 34](#_Toc134561160)

[Table 3 14. Table Production.ProductDocument(Source: AdventureWorks) 35](#_Toc134561161)

[Table 3 15. Table Production.ProductPhoto(Source: AdventureWorks) 36](#_Toc134561162)

[Table 3 16. Table Production.Product(Source: AdventureWorks) 36](#_Toc134561163)

[3.2.2. Challenge 38](#_Toc134561164)

[3.3. IT requirements Analysis (IT & Infrastructure) 38](#_Toc134561165)

[CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA 40](#_Toc134561166)

[4.1. Designing Data Warehouse 40](#_Toc134561167)

[4.1.1 Bus Matrix 40](#_Toc134561168)

[Table 4 1. Bus Matrix (Source: Author) 40](#_Toc134561169)

[4.1.2 Master Data 40](#_Toc134561170)

[Table 4 2. Master Data (Source: Author) 41](#_Toc134561171)

[4.1.3 Transaction Data 41](#_Toc134561172)

[4.1.4. Fact and dimension tables 41](#_Toc134561173)

[Table 4 3. Table of DimLocation 42](#_Toc134561174)

[Table 4 4. Table of DimProduct 42](#_Toc134561175)

[Table 4 5. Table of DimWorkOrder 43](#_Toc134561176)

[Table 4 6. Table of DimTime 44](#_Toc134561177)

[Table 4 7. Table of FactSales 45](#_Toc134561178)

[Table 4 8. Table of FactInventory 45](#_Toc134561179)

[Table 4 9. Table of FactProduction 46](#_Toc134561180)

[4.1.5. Data Warehouse model (Snowflake or Star or Galaxy) 47](#_Toc134561181)

[4.2. ETL processes 48](#_Toc134561182)

[4.2.1. Dimension Table’s ETL Process 48](#_Toc134561183)

[Figure 4 3. SSIS for DimLocation (source:Author) 48](#_Toc134561184)

[Figure 4 4. Result DimLocation (source:Author) 48](#_Toc134561185)

[Figure 4 5. SSIS for DimProduct (source:Author) 49](#_Toc134561186)

[Figure 4 6. Result DimProduct (source:Author) 49](#_Toc134561187)

[Figure 4 7. SSIS for DimWorkOrder (source:Author 50](#_Toc134561188)

[Figure 4 8. Result DimWorkOrder (source:Author) 50](#_Toc134561189)

[Figure 4 9. Result DimTime (source:Author) 50](#_Toc134561190)

[4.2.2. Fact Table’s ETL Process 50](#_Toc134561191)

[Figure 4 10. SSIS for FactInventory (source:Author) 51](#_Toc134561192)

[Figure 4 11. Result FactInventory’s ETL Process (source:Author) 51](#_Toc134561193)

[Figure 4 12. SSIS for FactSales (source:Author) 52](#_Toc134561194)

[Figure 4 13. Result FactSales’s ETL Process (source:Author) 52](#_Toc134561195)

[Figure 4 14. SSIS for FactProduction (source:Author) 52](#_Toc134561196)

[Figure 4 15. Result FactProduction’s ETL Process (source:Author) 53](#_Toc134561197)

[CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION 54](#_Toc134561198)

[5.1. Report and dashboard systems (structure) 54](#_Toc134561199)

[5.1.1. Building the CUBE 54](#_Toc134561200)

[Figure 5 1. Building the CUBE 54](#_Toc134561201)

[Figure 5 2. CUBE Model 55](#_Toc134561202)

[5.1.2. Building KPI System 55](#_Toc134561203)

[Figure 5 3. Building Calculation product defect rate( source: author) 56](#_Toc134561204)

[Figure 5 4. Building Calculation product cost( source: author) 56](#_Toc134561205)

[Figure 5 5. Display KPI ( source: author) 57](#_Toc134561206)

[Figure 5 6. KPI of Product defect rate( source: author) 57](#_Toc134561207)

[Figure 5 7. KPI of Product defect rate( source: author) 58](#_Toc134561208)

[Figure 5 8. Number of Work Order Quantity and Actual Cost of each Location (2011-2014) ( source: author) 58](#_Toc134561209)

[Figure 5 9. KPI of Production Cost( source: author) 59](#_Toc134561210)

[Figure 5 10. Result of KPI of Product defect rate each Products( source: author) 60](#_Toc134561211)

[Figure 5 11. Result of KPI of Product defect rate each Location( source: author) 60](#_Toc134561212)

[5.2. Data analysis with Power BI 60](#_Toc134561213)

[5.2.1 SalesOrder 60](#_Toc134561214)

[Figure 5 12. Dashboard of SalesOrder over the years 61](#_Toc134561215)

[5.2.2. Inventory 62](#_Toc134561216)

[Figure 5 13. Dashboard of Inventory by Product, Category, Subcategory 62](#_Toc134561217)

[5.2.3 Scrapped 63](#_Toc134561218)

[Figure 5 14. Dashboard of Scrapped by Product and ScrapReason 63](#_Toc134561219)

[5.2.4 LateDay 64](#_Toc134561220)

[Figure 5 15. Dashboard of LateDay by OperationSequence and Location 64](#_Toc134561221)

[5.3. Evaluation and Discussion 64](#_Toc134561222)

[CHAPTER 6: CONCLUSION AND FUTURE WORKS 65](#_Toc134561223)

[6.1. Results 65](#_Toc134561224)

[6.2. Limitations 65](#_Toc134561225)

[6.3. Future works 65](#_Toc134561226)

**List of Tables**

[Table 2 1. KPIs ( source: author) 21](#_Toc134556864)

[Table 3 1. Shows the list of tables stored in the production module of Adventure Works 2014.(Source: AdventureWorks) 25](#_Toc134556872)

[Table 3 2. Table Production.ProductCategory(Source: AdventureWorks) 28](#_Toc134556873)

[Table 3 3. Table Production.ProductModelIllustration (Source: AdventureWorks) 29](#_Toc134556874)

[Table 3 4. Table Production.Illustration(Source: AdventureWorks) 29](#_Toc134556875)

[Table 3 5. Table Production.ProductSubcategory(Source: AdventureWorks) 30](#_Toc134556876)

[Table 3 6. Table Production.ProductModel(Source: AdventureWorks) 30](#_Toc134556877)

[Table 3 7. Table Production.ProductModelProductDescriptionCulture(Source: AdventureWorks) 31](#_Toc134556878)

[Table 3 8. Table Production.ProductDescription(Source: AdventureWorks) 32](#_Toc134556879)

[Table 3 9. Table Production.UnitMeasureCode(Source: AdventureWorks) 32](#_Toc134556880)

[Table 3 10. Table Production.Culture(Source: AdventureWorks) 33](#_Toc134556881)

[Table 3 11. Table Production.ProductReview(Source: AdventureWorks) 33](#_Toc134556882)

[Table 3 12. Table Production.Productdocument(Source: AdventureWorks) 34](#_Toc134556883)

[Table 3 13. Table Production.ProductPhoto(Source: AdventureWorks) 34](#_Toc134556884)

[Table 3 14. Table Production.ProductDocument(Source: AdventureWorks) 35](#_Toc134556885)

[Table 3 15. Table Production.ProductPhoto(Source: AdventureWorks) 36](#_Toc134556886)

[Table 3 16. Table Production.Product(Source: AdventureWorks) 36](#_Toc134556887)

[Table 4 1. Bus Matrix (Source: Author) 40](#_Toc134556888)

[Table 4 2. Master Data (Source: Author) 41](#_Toc134556889)

[Table 4 3. Table of DimLocation 42](#_Toc134556890)

[Table 4 4. Table of DimProduct 42](#_Toc134556891)

[Table 4 5. Table of DimWorkOrder 43](#_Toc134556892)

[Table 4 6. Table of DimTime 44](#_Toc134556893)

[Table 4 7. Table of FactSales 45](#_Toc134556894)

[Table 4 8. Table of FactInventory 45](#_Toc134556895)

[Table 4 9. Table of FactProduction 46](#_Toc134556896)

**List of Figures**

[Figure 2 1. BI Architecture (Source: ResearchGate.net) 14](#_Toc134556792)

[Figure 2 2. Snowlake (Source: ierp service JSC) 18](#_Toc134556793)

[Figure 2 3. Star schema ((Source: ierp service JSC) 19](#_Toc134556794)

[Figure 2 4. Azure Analysis Service (Optional) 22](#_Toc134556795)

[Figure 3 1. The stages of production process ( source: author) 24](#_Toc134556796)

[Figure 3 2. Data source 28](#_Toc134556797)

[Figure 4 1. Data Warehouse Schema 47](#_Toc134561227)

[Figure 4 3. SSIS for DimLocation (source:Author) 48](#_Toc134561228)

[Figure 4 4. Result DimLocation (source:Author) 48](#_Toc134561229)

[Figure 4 5. SSIS for DimProduct (source:Author) 49](#_Toc134561230)

[Figure 4 6. Result DimProduct (source:Author) 49](#_Toc134561231)

[Figure 4 7. SSIS for DimWorkOrder (source:Author 50](#_Toc134561232)

[Figure 4 8. Result DimWorkOrder (source:Author) 50](#_Toc134561233)

[Figure 4 9. Result DimTime (source:Author) 50](#_Toc134561234)

[Figure 4 10. SSIS for FactInventory (source:Author) 51](#_Toc134561235)

[Figure 4 11. Result FactInventory’s ETL Process (source:Author) 51](#_Toc134561236)

[Figure 4 12. SSIS for FactSales (source:Author) 52](#_Toc134561237)

[Figure 4 13. Result FactSales’s ETL Process (source:Author) 52](#_Toc134561238)

[Figure 4 14. SSIS for FactProduction (source:Author) 52](#_Toc134561239)

[Figure 4 15. Result FactProduction’s ETL Process (source:Author) 53](#_Toc134561240)

[Figure 5 1. Building the CUBE 54](#_Toc134560372)

[Figure 5 2. CUBE Model 55](#_Toc134560373)

[Figure 5 3. Building Calculation product defect rate( source: author) 56](#_Toc134560374)

[Figure 5 4. Building Calculation product cost( source: author) 56](#_Toc134560375)

[Figure 5 5. Display KPI ( source: author) 57](#_Toc134560376)

[Figure 5 6. KPI of Product defect rate( source: author) 57](#_Toc134560377)

[Figure 5 7. KPI of Product defect rate( source: author) 58](#_Toc134560378)

[Figure 5 8. Number of Work Order Quantity and Actual Cost of each Location (2011-2014) ( source: author) 58](#_Toc134560379)

[Figure 5 9. KPI of Production Cost( source: author) 59](#_Toc134560380)

[Figure 5 10. Result of KPI of Product defect rate each Products( source: author) 60](#_Toc134560381)

[Figure 5 11. Result of KPI of Product defect rate each Location( source: author) 60](#_Toc134560382)

[Figure 5 12. Dashboard of SalesOrder over the years 61](#_Toc134560383)

[Figure 5 13. Dashboard of Inventory by Product, Category, Subcategory 62](#_Toc134560384)

[Figure 5 14. Dashboard of Scrapped by Product and ScrapReason 63](#_Toc134560385)

[Figure 5 15. Dashboard of LateDay by OperationSequence and Location 64](#_Toc134560386)

**LIST OF ACRONYMS**

|  |  |  |
| --- | --- | --- |
| **No.** | **Acronyms** | **Stand for** |
| 1 | BI | Business Intelligence |
| 2 | CRM | Customer Relationship Management |
| 3 | DW | Data Warehouse |
| 4 | ETL | Extracting, Transforming and Loading |
| 5 | JSON | JavaScript Object Notation |
| 6 | KPI | Key Performance Indicator |
| 7 | MDX | Multidimensional Expressions |
| 8 | ODBC | Open Database Connectivity |
| 9 | OLAP | Online Analysis Processing |
| 10 | OLE BD | Object Linking and Embedding Database |
| 11 | OLTP | Online Transaction Processing |
| 12 | RAM | Random Access Memory |
| 13 | SQL | Structured Query Language |
| 14 | SSAS | SQL Server Analysis Services |
| 15 | SSIS | Server Integration Services |

**CHAPTER 1: INTRODUCTION**

*Chapter 1 talk about introduction: Business case for the project, Objectives of the project, Research Objects, Scope of the project, Structure of project,...*

## 1.1. Business case for the project

The AdventureWorks illustration database is built on data from a fictional company called Adventure Works Cycles. This is supposed to be a multinational startup with a global scale. The company manufactures and sells bicycles made of metal and composites. The company's markets include North America, Europe and Asia. The company's headquarters are located in Bothell, Washington and employs 290 employees, in addition to several regional salespeople in the locations where it operates. Although Adventure Works is fictitious, it is treated as a case of reality just like a real company in the industry. Adventure Works provides a database and data warehouse about the entire business process from sales and supply of raw materials to production, supply and human resource training. Therefore, the researcher uses this fictional company as the main basis for building a self-service BI system.1.2. Objectives of the project.

### **1.2.1. General Objective**

The goal of a BI solution is to help improve product quality in the manufacturing process by addressing factors that affect product quality such as: finding the cause of defective products, helping to control product quality products, inventory control, timely detection of equipment problems, problems from suppliers.

### **1.2.2. Specific Objective**

Firstly, Find out the causes of equipment failures in order to limit interruptions in production or cause product defects in the production process; by checking the number of broken products, the percentage of defective products according to each product type. From there, identify which devices are having problems and find their cause.

Next, determine the cause of the defect from the Production.ProductReview table to collect customer feedback on the product or calculate the defective rate of each product type to easily control product quality.

Check forecasting capabilities and inventory counts to align future inventory levels and production needs with the company's production plan.

## 1.3. Research Objects

Regarding the research object, we mainly focus on the factors affecting the product quality of Adventure works and propose BI solutions to improve product quality in the production process.

## 1.4. Scope of the project

Within the scope of the project, we work to build a product quality management system in the production process using BI solutions, focusing on supporting the production department to analyze data to ensure product quality. products and find solutions to fix defective products.

## 1.5. Value and desired outcome of the project

Determine the percentage of defective products and the cause of the defective product.

Identify product production delays, cost deviations in production.

Consider the deviation between planned and actual production times, planned production costs and actual production costs.

## 1.6. Structure of project

The project has 6 chapters:

* **Chapter 1:** Introduction: This chapter will give an overview of the project through the Business case, Objectives, Research Objects, Scope, Value and desired outcome.
* **Chapter 2:** Theoretical basis: Presenting the concepts of BI, ETL, Data warehouse, Data mart, KPIs.
* **Chapter 3:** Requirements analytics and Introduction to BI solution: Illustrating Business processes, Data source and challenges of the Production department. Besides, analyzing Business requirements, IT requirements and comparing BI and Data Visualization tools to propose the BI solution for the project.
* **Chapter 4:** Building data warehouse and Integrating data: Conducting to design Data Warehouse and doing ETL processes.
* **Chapter 5:** Data analytics and Visualization: Carrying out Data analytics with SSAS technology. Demonstrating the report and dashboard systems and conducting the data analysis with BI and Data Visualization tools.
* **Chapter 6:** Conclusion and Future works: Presenting the results achieved, limitations of the project and future works.

**CHAPTER 2: THEORETICAL BASIS**

*Chapter 2 talk about theoretical basis: Overview about BI, ETL Process, Data warehouse and Data mart, Azure Analysis Service (Optional), KPIs.*

## 2.1. Overview about BI

### **2.1.1**. **What is BI?**

Business Intelligence (BI) is an important area of modern business, making it easier for organizations and businesses to access and use information to make more accurate decisions and increase competitiveness. BI tools and techniques enable the aggregation, processing, and analysis of data from a variety of sources, including internal and external, to generate useful information that helps managers and leaders make informed decisions. strategic decision making. Applications of Business Intelligence are diverse, from monitoring and evaluating business performance, analyzing markets and customers, to forecasting future trends and prospects. This helps organizations make decisions based on a clear and precise understanding of market needs and thereby enhance business interoperability and prosperity.

### **2.1.2. BI Architecture**

Business Intelligence (BI) Architecture is the architectural model for the entire process of building a Business Intelligence system. It includes the basic components of BI such as data sources, data warehouses, data analysis tools, data visualization tools, reporting systems, and smart applications. The BI architecture is designed to ensure that data is efficiently and securely collected, transformed, and stored, analyzes are performed accurately and quickly, and reports and visualizations are provided. Make data available to end users easily and quickly. A good BI architecture must also ensure scalability and flexibility to be able to respond to the growth and change of the business. This architecture needs to be designed so that it can integrate data from a variety of sources, meet security requirements, and align with industry standards and regulations.

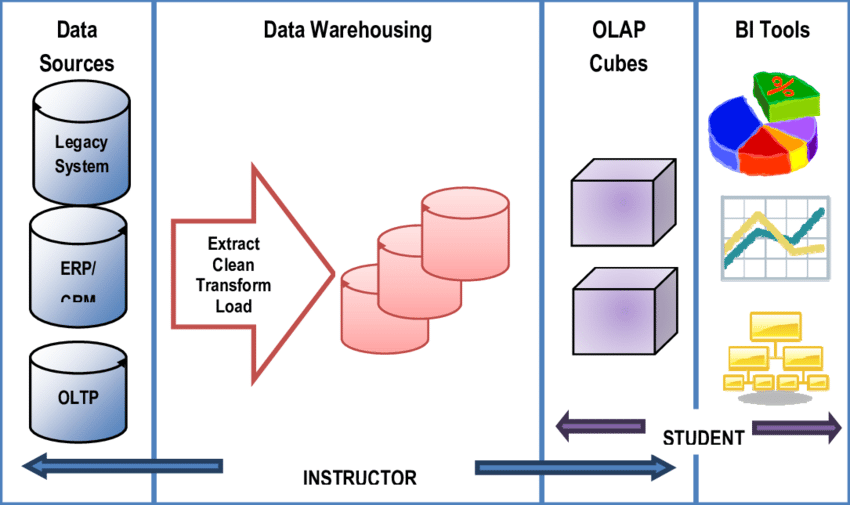


Figure 2 1. BI Architecture (Source: ResearchGate.net)

*Base on figure 2.1 (BI architecture)* consists of 3 main components:

Data Source: a place for integrating and collecting data stored to come from many different sources such as business applications, heterogeneity of origin, and type.

Data Warehouse and Data Mart: Use extraction and conversion tools like ETL (Extract - Transform - Load) to store different data sources into a common Database to support business analysis. It is a database designed in a different model than a relational database and a place to store data for an organization's long-term goals.

BI solution: The data is extracted and finally processed then used for mathematical models and analytical methods to support decision making. In a BI system, several decision support applications can be implemented such as:Multidimensional, Cubes Time series, Data Mining, Optimization model.

The main components of the BI system are:

* Source system
* Data warehouse/Data Mart
* Data exploration: includes tools to perform passive analysis, including reporting and query systems as well as statistical methods.
* Data Mining: consists of proactive BI methods, whose goal is extracting processed data information from Data Warehouse and then combined with algorithms to make or predict profitable decisions for the business. These methods include mathematical models for pattern recognition, data mining techniques
* Data Presentation: Generate reports and charts from the data mining process to serve the needs of users.
* Making Decision: Making choices and practical application of a particular decision

### **2.1.3. Advantage of BI in enterprises**

Improved decision-making: BI provides easily accessible and actionable insights that help executives and managers make more informed decisions.

Increased efficiency: By automating many data-related tasks, BI streamlines processes and allows employees to focus on their core responsibilities.

Competitive advantage: With early access to key business intelligence, enterprises can identify opportunities and stay ahead of competitors.

Customer insights: BI helps organizations understand their customers' needs and preferences to enhance customer satisfaction and loyalty.

Cost savings: By optimizing resources and eliminating inefficiencies, BI can reduce costs across the organization.

Better forecasting: Accurate predictions based on historical data can help organizations better anticipate future trends and plan accordingly.

### **2.1.4. BI Strategy for Business**

The process of developing a BI solution for businesses consists of 6 steps:

* Step 1: Collect and analyze requirements.
* Step 2: Learn, analyze source data and the current state of the system.
* Step 3: Building a Data Warehouse model - Build a Data Warehouse model with the required information from the report with an appropriate structure for data retrieval and analysis. Choosing ETL tools and transfer data from data sources into Data Mart, (we will be using SSIS provided by Microsoft).
* Step 4: Building multidimensional cubes, visualization (use the BI tools: Power BI, Tableau) and predictive analytics - Use SSRS tools to extract data to process or generate reports. Extracting data from Cubes by using the MDX language to query.
* Step 5: Testing and Quality Control (QC).
* Step 6: GO LIVE, deployment, operation, and maintenance.

## 2.2. ETL Process

### **2.2.1. What is ETL?**

ETL (Extract, Transform, Load) is an important process for transforming and normalizing data from various sources into a format suitable for storing and using for data analysis or management purposes. This process consists of three main steps: Extract, Transform, and Load. In this essay, we will explore in more detail the concepts and theoretical basis of the ETL process.

### **2.2.2. Why do we need ETL?**

We need ETL (Extract, Transform, Load) to move and transform data from various sources into a common format that can be used for analysis or other purposes.

ETL allows organizations to integrate data from multiple systems, databases, and formats, making it easier to access, analyze, and use data to make informed decisions.

It also helps in cleaning and validating data, removing duplicates, and preparing it for analysis or reporting. ETL is an essential process for data warehousing, business intelligence, and analytics applications.

### **2.2.3. ETL Process**

Extract: This step involves accessing different data sources to extract the data and move it into temporary storage. For database systems, extraction can be done using SQL or other database management tools to get data from tables or other data structures. For data files, ETL can use file reading tools to read data from CSV, Excel, or other data file types.

Transform: This step involves processing the extracted data to normalize and transform the data into a format suitable for the end purpose. This step includes operations such as removing duplicate data, replacing null or missing values, converting data types, creating calculated columns, and so on.

Load: This step is the process of putting the processed data into the main repository to use for the desired purpose. The repository can be a relational database or other systems such as a data warehouse or data lake. This step includes defining the object and structure of the table or file to enter the data into and performing the data import process.

## 2.3. Data warehouse and Data mart

### **2.3.1. What are Data warehouse and Data mart?**

A data warehouse (DW) or data warehouse is a system that stores data from many different sources and environments such as sales software, accounting, human resources or banking core system, etc. performance of queries for reporting and analysis, has data groups to support decision making Data Warehouse acts as a central repository. Data enters the data warehouse from the transaction system and other related databases. Then, the data is processed and transformed so that users can access these data through Business Intelligence tools, SQL clients or spreadsheets.

A data mart is a subset of a data warehouse focused on a particular line of business, department, or subject area. Data marts make specific data available to a defined group of users, which allows those users to quickly access critical insights without wasting time searching through an entire data warehouse. For example, many companies may have a data mart that aligns with a specific department in the business, such as finance, sales, or marketing.

### **2.3.2. Who needs Data warehouse and Data mart?**

The use of Data warehouse is necessary for: responsible for making decisions based on volume of data Users use complex, customized processes to gather information from a variety of sources People who need to use simple technology to access data People who want a systematized approach to decision making People who want to increase performance with huge amounts of data for reports, grids or charts People who want to discover "hidden patterns" of data flows and groups.

The use of Data Mart is necessary because: Data mart is designed for use by a specific department, unit or group of users within an organization and is a stripped down version of Data Warehouse. For example, Marketing, Sales, Service or finance. It is only used by a single department in an organization.

### **2.3.3. Advantages and disadvantages of Data warehouse**

**Advantages of Data warehouse:**

* Provide a comprehensive view of the business: Although an enterprise has many business segments managed by different operating systems, a Data Warehouse is a place where all information about different business areas is gathered (integrated) to provide a unique perspective. comprehensive view. This integration also provides the ability to cross-assess different business segments to assess the correlation between them.
* Provide complete current and historical information of the business, and ready for exploitation and use for strategic decision support: From the information here, it is shown that the data in the warehouse is not merely raw data taken from operational systems, but it has been aggregated and calculated into measures of analytical significance.
* Capable of providing detailed data on demand without having to access operational systems: This shows that in some cases there may be a need to analyze data at the transaction level, which will also be stored in the Data Warehouse. Make sure the information in the Data Warehouse is consistent.
* For example, a business has many operations managed with different operating systems, but needs to ensure that a customer appears in many systems to a single person. This applies to other data objects. In addition, a data field can be represented in many different ways, for example, Customer Name has a system that separates the family name and first name into 2 different fields, and the system only stores it in 1 field. When aggregated to the Data Warehouse, it will be converted to a unified form.
* As a flexible and interactive source of strategic information:Flexible and interactive text here shows that users can get different information of the same object. The word interactive here indicates that many operations can be performed with data objects instead of returning a static list. For example, you can get aggregated data by day, week, month of the same data object. Interactivity represents the ability of users to apply data analysis operations, an example of which is the SQL Server Analysis Services (SSAS) data analysis platform that allows users to interact to analyze data.

**Disadvantages of Data Warehouse:**

* Not an ideal choice for unstructured data.
* Creating and deploying a data warehouse is time-consuming and can quickly become obsolete.
* Difficult to change data types and ranges, data source schemas, indexes, and queries.
* Data warehouse seems simple, but in reality it is too complicated for the average user who has no knowledge of Data and programming.

So the organization needs to spend a lot of resources on training and implementation.

### **2.3.4. Snowflake and Star schemas**

The snowflake model overcomes the disadvantage of the star model in terms of non-normalized data. It therefore allows dimensional tables to be normalized (depending on the case it can be normalized to the 3 Boyce–Codd norm). Since after normalizing the dimension tables, it is shaped like a snowflake (Figure 3 on the right), this is the reason it got its name.

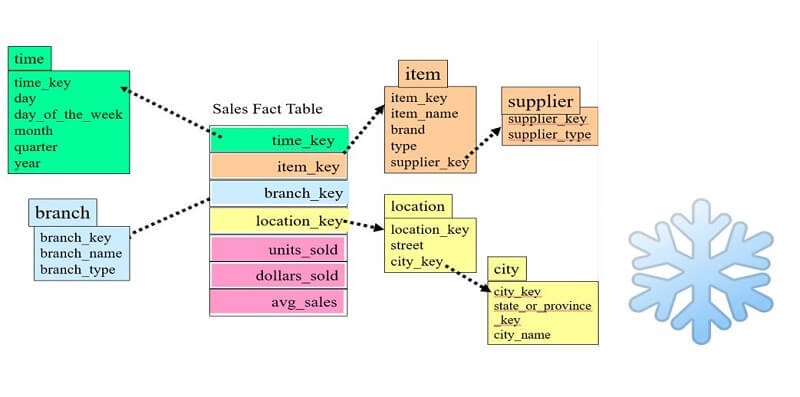
****

Figure 2 2. Snowlake (Source: ierp service JSC)

Figure 2.2 illustrates a snowflake model in which the location table has been normalized, due to lack of space, in this figure, only the city table (which contains the state\_or\_province\_key key) can be drawn to link it to the state\_or\_province table, continue the state\_or\_province table. has a relationship with the country table (if you want to normalize to standard 3). Although it overcomes the disadvantages of the constellation model, it destroys the advantage of the star model, which is the speed of data processing when it has to link many tables together to get data. Similar to the star model, an analytical subject can be represented by one or several snowflakes.

### **2.3.5. Star schema**

In this model, an event table will be in the center and surrounded by dimension tables (Figure 2 on the left), since this image resembles a glowing star, it is named the star model. stars (Figure 2 on the right).

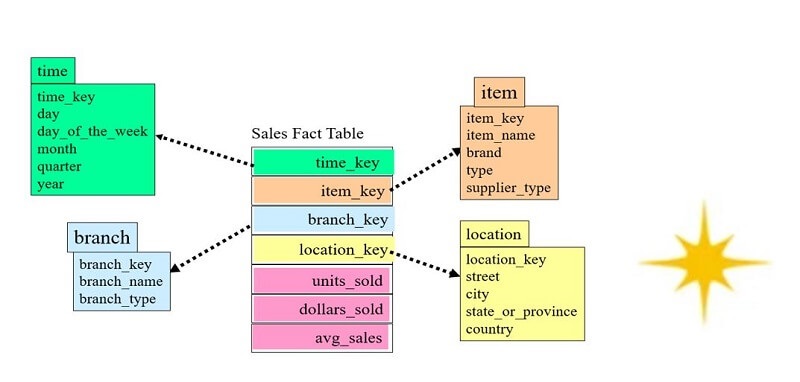


Figure 2 3. Star schema ((Source: ierp service JSC)

In the example in Figure 2.3, the event table contains sales information with the following metrics: unit sold, dollar sold, and average sales. This event table is associated with 4 dimension tables: time (time), store branch (branch), item (item) and store location (location). The feature of the star model is that there is only one level of relationship between the dimensional table and the fact table, so when accessing data, DBMSs will process faster and return results faster. But the disadvantage of this method is that some of the dimensional tables are not normalized. An example is the positional dimensional table, where it is not normalized to the database model. In this table a lot of data is duplicated, for example, all fields city, state\_or\_province, and country will be duplicated on rows that have the same city. Denormalization of data does not guarantee data consistency. When the data changes, for example, people rename the city, it is possible that the update process will be missed because a lot of lines need to be updated. In addition, duplicate data will increase storage space, affecting the backup and data synchronization process. An analytical topic may be represented by one or more “stars”.

## 2.4. KPIs

### **2.4.1. KPIs Definition**

KPI (Key Performance Indicator) is measurable values used to evaluate how effectively an organization or individual is achieving its objectives. KPIs can be applied to various areas of business, such as finance, marketing, human resources, and customer service.

Some general characteristics of KPIs include:

* They are relevant to the organization's goals and objectives.
* They are quantifiable and can be measured.
* They are easily understood by those who are responsible for achieving them.
* They are actionable, meaning they can be used to drive performance improvements.
* They are regularly monitored and tracked to assess progress towards goals.

Examples of KPIs include revenue growth, customer satisfaction, employee turnover rate, website traffic, and social media engagement. By measuring and tracking these indicators, organizations can make data-driven decisions and continuously improve their performance.

### **2.4.2. The advantages and disadvantages of KPIs**

**Advantages:**

* Focus on the most important metrics: KPIs help organizations focus on the most important metrics that drive success, rather than getting lost in a sea of data.
* Clarity of goals: KPIs can help clarify the goals of an organization or project, making it easier to measure progress and determine whether efforts are aligned with the organization's overall mission.
* Performance monitoring: KPIs provide a way to monitor performance over time, allowing organizations to identify trends and take corrective action before problems become too severe.
* Better decision-making: KPIs can help organizations make better decisions by providing data-driven insights and identifying areas where improvements are needed.
* Accountability: KPIs can create a culture of accountability by making it clear what is expected of individuals and teams.

**Disadvantages:**

* Risk of tunnel vision: Focusing too heavily on a small set of KPIs can lead to tunnel vision, causing organizations to overlook important aspects of their performance that are not being measured.
* Data overload: Collecting and analyzing too many KPIs can lead to information overload, making it difficult to identify the most important metrics.
* Lack of context: KPIs can sometimes lack context, making it difficult to understand the root causes of performance issues.
* Misaligned incentives: KPIs can sometimes create incentives for individuals or teams to prioritize certain metrics at the expense of others, which can lead to suboptimal outcomes.
* Difficulty of measurement: Some aspects of performance can be difficult to measure with KPIs, such as customer satisfaction or employee morale. In these cases, KPIs may not provide a complete picture of organizational performance.

### **2.4.3. Categories of KPIs (Production)**

Categories of KPIs( *base on table 2.1*)

Table 2 1. KPIs ( source: author)

|  |  |  |
| --- | --- | --- |
| **No.** | **KPIs type** | **Including** |
| 1 | Sales and Marketing KPIs | * Sales Revenue * Customer Acquisition Cost * Conversion Rate * Churn Rate * Customer Lifetime Value * Net Promoter Score * Website Traffic * Click - Through Rate * Cost per Lead * Return on Investment |
| 2 | Production KPIs | * Defective products * Order management * Maintenance * Productivity * Overall Equipment Effectiveness * Cycle Time * Yield * Downtime * Inventory Turns * Cost per unit * Throughput * Lead Time |
| 3 | Human Resources KPIs | * Employee Turnover Rate * Employee Engagement Score * Absenteeism Rate * Time to Hire * Cost per Hire * Employee Satisfaction * Training and Development Spending per Employee * Average Employee Tenure * Diversity and Inclusion Metrics |
| 4 | Operations KPIs | * Overall Equipment Effectiveness * Manufacturing Cycle Time * On-Time Delivery * Inventory Levels * Quality Metrics * Lead Time * Supplier Performance * Maintenance Costs * Health and Safety Incidents |

## 2.5. Azure Analysis Service (Optional)

Azure Analysis Services is a fully managed platform as a service (PaaS) that provides enterprise-grade data models in the cloud. Use advanced mashup and modeling features to combine data from multiple data sources, define metrics, and secure your data in a single, trusted tabular semantic data model. The data model provides an easier and faster way for users to perform ad hoc data analysis using tools like Power BI and Excel.

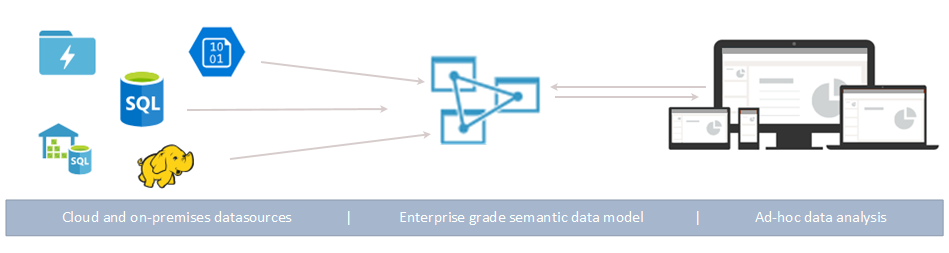


Figure 2 4. Azure Analysis Service (Optional)

Azure Analysis Services ( *figure 2.4* )can be applied in Business Intelligence (BI) solutions as follows:

Create a data model that fits your business needs, including cubes and dimensions.

Connect various data sources to Azure Analysis Services, including SQL Server databases and cloud computing services such as Azure Data Lake Store and Azure Blob Storage.

* Integrate Azure Analysis Services with other BI tools such as Power BI or Excel to access and analyze data.
* Perform data analysis tasks, including multidimensional analysis, row-level analysis, and holistic analysis.
* Grant permissions to users to access and process reports, dashboards, and other visualizations.

In summary, Azure Analysis Services is a tool that serves for data analysis and optimization in your BI solution. It helps users easily access and analyze complex data, making the decision-making process based on data easier.

**CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION**

Chapter 3 talk about :*Business processes, Data source and challenges, Information Tables about ,...*

## 3.1. Business processes

### **3.1.1. Production department**

The manufacturing department is the part of the business that is responsible for the manufacturing processes and activities involved in producing products or services. Specifically, the manufacturing department is responsible for designing, implementing, monitoring, and managing day-to-day production processes, including the management of resources such as materials, human resources, and machinery.

### **3.1.2 The purpose of production process**

The goal of the manufacturing department is to create products or services that ensure quality, quantity and delivery time are consistent with customer and business requirements, and achieve business efficiency and productivity. highest capacity.

### **3.1.3 The stages of production process**

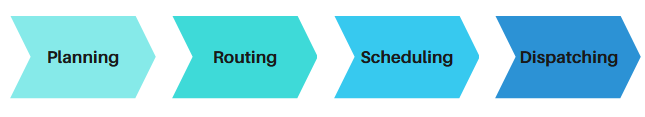


Figure 3 1. The stages of production process ( source: author)

*Base on figure 3.1 (The stages of production process)*, there are four stages to production process: planning, routing, scheduling, and dispatching.

* Planning: To ensure that production stays on track, it is necessary to have a pre-production plan that aligns with existing goals.`
* Routing: The planning process includes determining product details, estimating production costs, sourcing raw materials, establishing the assembly line, creating a production budget, and hiring the necessary workforce.
* Scheduling: The routing stage aims to understand the sequence of activities that need to occur. For instance, the furniture manufacturing process may involve collecting raw materials, shaping them, polishing, finishing, and conducting quality control before distribution. Proper sequencing in this stage minimizes supply chain hiccups.
* Dispatching: Scheduling involves determining work hours and timeframe for each activity. This is accomplished by creating a master schedule that can be broken down into smaller, more specific plans.

## 3.2. Data source and challenges

### **3.2.1. Data source**

AdventureWorks Database is a sample product from Microsoft used for online transaction processing (OLTP) and in support of Adventure Works Cycles. The data is initially stored in Microsoft SQL Server, then imported into Power BI Desktop to perform data warehouse operations using Power Query and the data view in Power BI.

This database consists of 6 modules, of which the Production module has 25 data tables that are used as the main source to create an OLAP cube with many dimensions and a visual analysis data block containing many fact tables. The remaining modules include dbo, Purchase, Person, HumanResources and Sales, with different number of data tables. ( *Base on table 3.1*)

Table 3 1. Shows the list of tables stored in the production module of Adventure Works 2014.(Source: AdventureWorks)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **SchemaName** | **TableName** | **Description** | **Rowcount** |
| 1 | Production | BillOfMaterials | Items required to make bicycles and bicycle subassemblies. It identifies the heirarchical relationship between a parent product and its components. | 2679 |
| 2 | Production | Culture | Lookup table containing the languages in which some AdventureWorks data is stored. | 8 |
| 3 | Production | Document | Product maintenance documents. | 13 |
| 4 | Production | Illustration | Bicycle assembly diagrams. | 5 |
| 5 | Production | Location | Product inventory and manufacturing locations. | 14 |
| 6 | Production | Product | Products sold or used in the manfacturing of sold products. | 504 |
| 7 | Production | ProductCategory | High-level product categorization. | 4 |
| 8 | Production | ProductCostHistory | Changes in the cost of a product over time. | 395 |
| 9 | Production | ProductDescription | Product descriptions in several languages. | 762 |
| 10 | Production | ProductDocument | Cross-reference table mapping products to related product documents. | 32 |
| 11 | Production | ProductInventory | Product inventory information. | 1069 |
| 12 | Production | ProductListPriceHistory | Changes in the list price of a product over time. | 395 |
| 13 | Production | ProductModel | Product model classification. | 128 |
| 14 | Production | ProductModelIllustration | Cross-reference table mapping product models and illustrations. | 7 |
| 15 | Production | ProductModelProductDescriptionCulture | Cross-reference table mapping product descriptions and the language the description is written in. | 762 |
| 16 | Production | ProductPhoto | Product images. | 101 |
| 17 | Production | ProductProductPhoto | Cross-reference table mapping products and product photos. | 504 |
| 18 | Production | ProductReview | Customer reviews of products they have purchased. | 4 |
| 19 | Production | ProductSubcategory | Product subcategories. See ProductCategory table. | 37 |
| 20 | Production | ScrapReason | Manufacturing failure reasons lookup table. | 16 |
| 21 | Production | TransactionHistory | Record of each purchase order, sales order, or work order transaction year to date. | 113443 |
| 22 | Production | TransactionHistoryArchive | Transactions for previous years. | 89253 |
| 23 | Production | UnitMeasure | Unit of measure lookup table. | 38 |
| 24 | Production | WorkOrder | Manufacturing work orders. | 72591 |
| 25 | Production | WorkOrderRouting | Work order details. | 67131 |

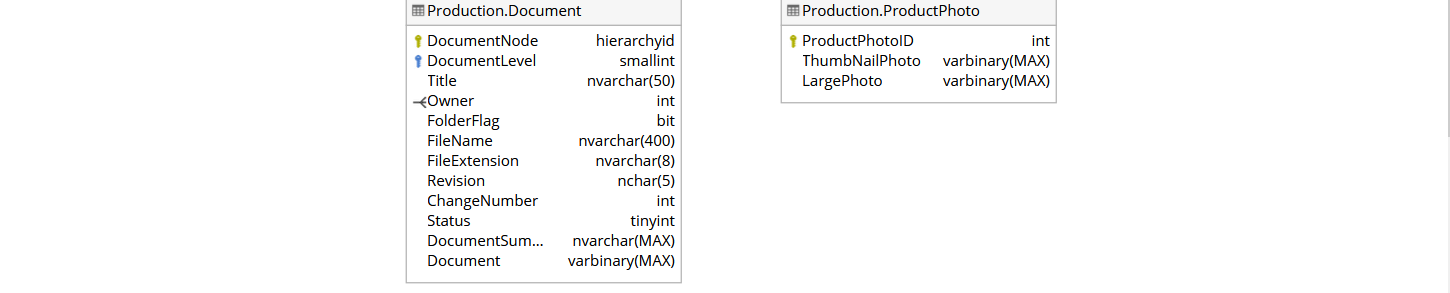
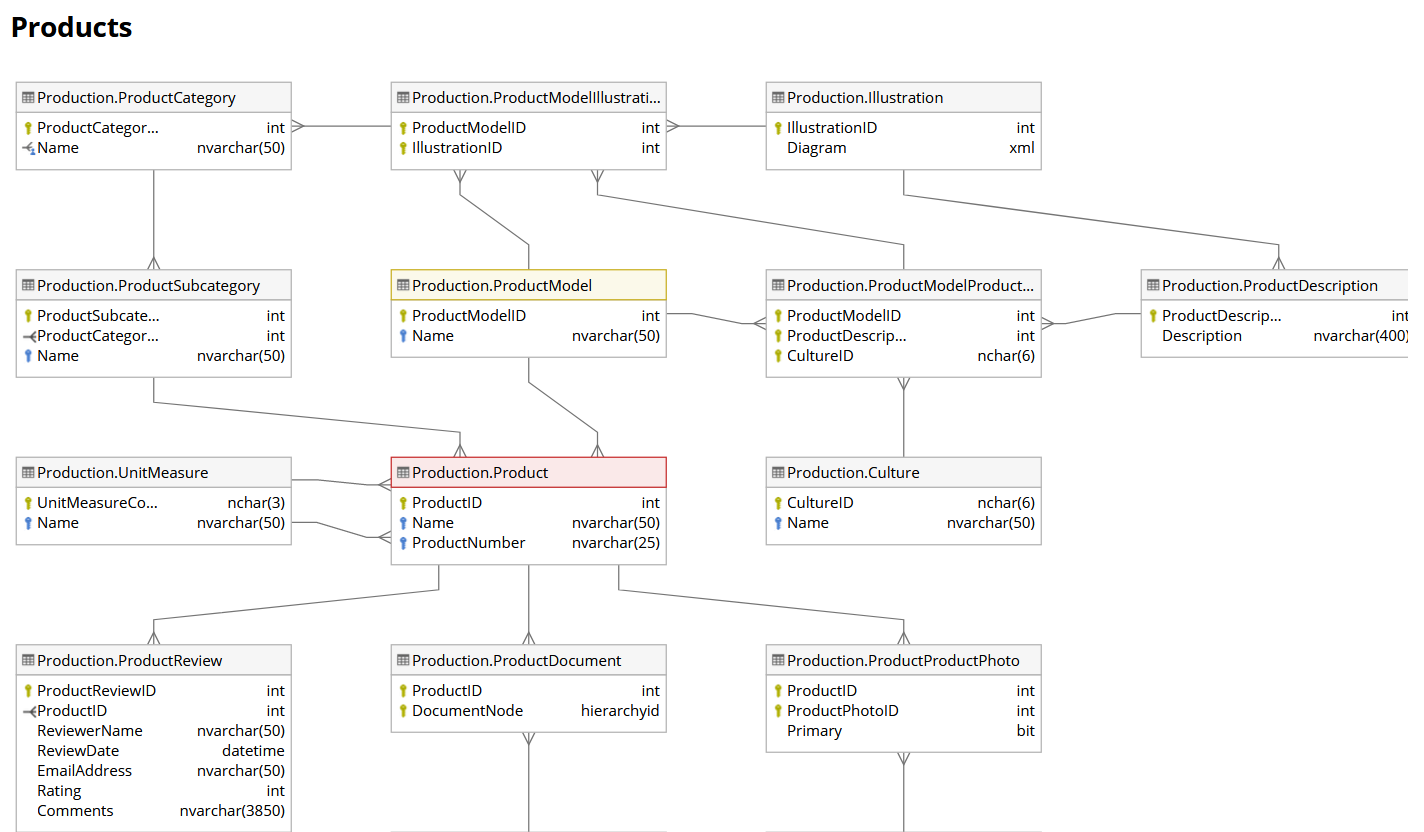


Figure 3 2. Data source

In the original database, the Production process of Adventure Work has 15 different tables ( *figure 3.2: Data source*)

Table 3.2 describes key, name, datatype, null, attributes, description, columns: ProductCategoryId, Name, rowguid, ModifiedDate.

Table 3 2. Table Production.ProductCategory(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductCategoryID | int |  | Identity | Primary key for ProductCategory records. |
| 2 |  | Name | nvarchar(50) |  |  | Category description. |
| 3 |  | rowguid | uniqueidentifier |  | Default: newid() | ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample. |
| 4 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.3 describes key, name, datatype, null, attributes, description, columns: ProductModelID, IllustrationID, ModifiedDate

Table 3 3. Table Production.ProductModelIllustration (Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductModelID | int |  |  | Primary key. Foreign key to ProductModel.ProductModelID. |
| 2 | PK | IllustrationID | int |  |  | Primary key. Foreign key to Illustration.IllustrationID. |
| 3 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.4 describes key, name, datatype, null, attributes, description, columns: IllustrationId, Diagram, ModifiedDate.

Table 3 4. Table Production.Illustration(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | Illustration  ID | int |  | Identity | Primary key for Illustration records. |
| 2 |  | Diagram | xml | X |  | Illustrations used in manufacturing instructions. Stored as XML. |
| 3 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.5 describes key, name, datatype, null, attributes, description, columns:ProductionSubcategoryID, ProductCategoryID, Name, rowguid, ModifiedDate.

Table 3 5. Table Production.ProductSubcategory(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductSubcategoryID | int |  | Identity | Primary key for ProductSubcategory records. |
| 2 |  | ProductCategoryID | int |  |  | Product category identification number. Foreign key to ProductCategory.ProductCategoryID. |
| 3 |  | Name | nvarchar(50) |  |  | Subcategory description. |
| 4 |  | rowguid | uniqueidentifier |  | Default: newid() | ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample. |
| 5 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

‘hnm

Table 3.6 describes key, name, datatype, null, attributes, description, columns:ProductModelID, Name, CatalogDescription, Instructions.

Table 3 6. Table Production.ProductModel(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductModelID | int |  | Identity | Primary key for ProductModel records. |
| 2 |  | Name | nvarchar(50) |  |  | Product model description. |
| 3 |  | CatalogDescription | xml | X |  | Detailed product catalog information in xml format. |
| 4 |  | Instructions | xml | X |  | Manufacturing instructions in xml format. |
| 5 |  | rowguid | uniqueidentifier |  | Default: newid() | ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample. |
| 6 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.7 describes key, name, datatype, null, attributes, description, columns:ProductModelID, ProductDescriptionID, CultureId, ModifiedDate.

Table 3 7. Table Production.ProductModelProductDescriptionCulture(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductModelID | int |  |  | Primary key. Foreign key to ProductModel.ProductModelID. |
| 2 | PK | ProductDescriptionID | int |  |  | Primary key. Foreign key to ProductDescription.ProductDescriptionID. |
| 3 | PK | CultureID | nchar(6) |  |  | Culture identification number. Foreign key to Culture.CultureID. |
| 4 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.8 describes key, name, datatype, null, attributes, description, columns:ProductDescriptionID, Description, rowguid, ModifiedDate.

Table 3 8. Table Production.ProductDescription(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductDescriptionID | int |  | Identity | Primary key for ProductDescription records. |
| 2 |  | Description | nvarchar(400) |  |  | Description of the product. |
| 3 |  | rowguid | uniqueidentifier |  | Default: newid() | ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample. |
| 4 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.9 describes key, name, datatype, null, attributes, description, columns:UnitMeasureCode, Name, ModifiedDate.

Table 3 9. Table Production.UnitMeasureCode(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | UnitMeasureCode | nchar(3) |  |  | Primary key. |
| 2 |  | Name | nvarchar(50) |  |  | Unit of measure description. |
| 3 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.10 describes key, name, datatype, null, attributes, description, columns:CultureID, Name, ModifiedDate.

Table 3 10. Table Production.Culture(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | CultureID | nchar(6) |  |  | Primary key for Culture records. |
| 2 |  | Name | nvarchar(50) |  |  | Culture description. |
| 3 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.11 describes key, name, datatype, null, attributes, description, columns:ProductReviewID, ProductID, ReviewerName, ReviewDate,EmailAddress,..

Table 3 11. Table Production.ProductReview(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductReviewID | int |  | Identity | Primary key for ProductReview records. |
| 2 |  | ProductID | int |  |  | Product identification number. Foreign key to Product.ProductID. |
| 3 |  | ReviewerName | nvarchar(50) |  |  | Name of the reviewer. |
| 4 |  | ReviewDate | datetime |  | Default: getdate() | Date review was submitted. |
| 5 |  | EmailAddress | nvarchar(50) |  |  | Reviewer's e-mail address. |
| 6 |  | Rating | int |  |  | Product rating given by the reviewer. Scale is 1 to 5 with 5 as the highest rating. |
| 7 |  | Comments | nvarchar(3850) |  |  | Reviewer's comments |
| 8 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.12 describes key, name, datatype, null, attributes, description, columns:DocumentNode, ProductID, ModifiedDate.

Table 3 12. Table Production.Productdocument(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductID | int |  |  | Product identification number. Foreign key to Product.ProductID. |
| 2 | PK | DocumentNode | hierarchyid |  |  | Document identification number. Foreign key to Document.DocumentNode. |
| 3 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.13 describes key, name, datatype, null, attributes, description, columns:ProductPhotoID, ProductID,Primary.

Table 3 13. Table Production.ProductPhoto(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 |  | ProductID | int |  |  | Product identification number. Foreign key to Product.ProductID. |
| 2 | PK | ProductPhotoID | int |  |  | Product photo identification number. Foreign key to ProductPhoto.ProductPhotoID. |
| 3 |  | Primary | bit |  | Default: 0 | 0 = Photo is not the principal image. 1 = Photo is the principal image. |
| 4 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.14 describes key, name, datatype, null, attributes, description, columns:DocumentNode,DocumentLevel, Title,Owner, FolderFlag,...

Table 3 14. Table Production.ProductDocument(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | DocumentNode | hierarchyid |  |  | Primary key for Document records. |
| 2 |  | DocumentLevel | smallint |  | Computed: [DocumentNode].[GetLevel]() | Depth in the document hierarchy. |
| 3 |  | Title | nvarchar(50) |  |  | Title of the document. |
| 4 |  | Owner | int |  |  | Employee who controls the document. Foreign key to Employee.BusinessEntityID |
| 5 |  | FolderFlag | bit |  | Default: 0 | 0 = This is a folder, 1 = This is a document. |
| 6 |  | FileName | nvarchar(400) |  |  | File name of the document |
| 7 |  | FileExtension | nvarchar(8) |  |  | File extension indicating the document type. For example, .doc or .txt. |
| 8 |  | Revision | nchar(5) |  |  | Revision number of the document. |
| 9 |  | ChangeNumber | int |  | Default: 0 | Engineering change approval number. |
| 10 |  | Status | tinyint |  |  | 1 = Pending approval, 2 = Approved, 3 = Obsolete |
| 11 |  | DocumentSummary | nvarchar(MAX) |  |  | Document abstract. |
| 12 |  | Document | varbinary(MAX) |  |  | Complete document. |
| 13 |  | rowguid | uniqueidentifier |  | Default: newid() | ROWGUIDCOL number uniquely identifying the record. Required for FileStream. |
| 14 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.15 describes key, name, datatype, null, attributes, description, columns:ProductPhotoID, ThumbNailPhoto, ThumbnailPhotoFileName, LargePhoto, LargePhotoFileName...

Table 3 15. Table Production.ProductPhoto(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductPhotoID | int |  | Identity | Primary key for ProductPhoto records. |
| 2 |  | ThumbNailPhoto | varbinary(MAX) | X |  | Small image of the product. |
| 3 |  | ThumbnailPhotoFileName | nvarchar(50) | X |  | Small image file name. |
| 4 |  | LargePhoto | varbinary(MAX) | X |  | Large image of the product. |
| 5 |  | LargePhotoFileName | nvarchar(50) | X |  | Large image file name. |
| 6 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

Table 3.16 describes key, name, datatype, null, attributes, description, columns:ProductID, Name, ProductNumber, MakeFlag, FinishedGoodsFlag, …

Table 3 16. Table Production.Product(Source: AdventureWorks)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Key** | **Name** | **Data type** | **Null** | **Attributes** | **Description** |
| 1 | PK | ProductID | int |  | Identity | Primary key for Product records. |
| 2 |  | Name | nvarchar(50) |  |  | Name of the product. |
| 3 |  | ProductNumber | nvarchar(25) |  |  | Unique product identification number. |
| 4 |  | MakeFlag | bit |  | Default: 1 | 0 = Product is purchased, 1 = Product is manufactured in-house. |
| 5 |  | FinishedGoodsFlag | bit |  | Default: 1 | 0 = Product is not a salable item. 1 = Product is salable. |
| 6 |  | Color | nvarchar(15) | X |  | Product color. |
| 7 |  | SafetyStockLevel | smallint |  |  | Minimum inventory quantity. |
| 8 |  | ReorderPoint | smallint |  |  | Inventory level that triggers a purchase order or work order. |
| 9 |  | StandardCost | money |  |  | Standard cost of the product. |
| 10 |  | ListPrice | money |  |  | Selling price. |
| 11 |  | Size | nvarchar(5) | X |  | Product size. |
| 12 |  | SizeUnitMeasureCode | nchar(3) | X |  | Unit of measure for Size column. |
| 13 |  | WeightUnitMeasureCode | nchar(3) | X |  | Unit of measure for Weight column. |
| 14 |  | Weight | decimal(8, 2) | X |  | Product weight. |
| 15 |  | DaysToManufacture | int |  |  | Number of days required to manufacture the product. |
| 16 |  | ProductLine | nchar(2) | X |  | R = Road, M = Mountain, T = Touring, S = Standard |
| 17 |  | Class | nchar(2) | X |  | H = High, M = Medium, L = Low |
| 18 |  | Style | nchar(2) | X |  | W = Womens, M = Mens, U = Universal |
| 19 |  | ProductSubcategoryID | int | X |  | Product is a member of this product subcategory. Foreign key to ProductSubCategory.ProductSubCategoryID. |
| 20 |  | ProductModelID | int | X |  | Product is a member of this product model. Foreign key to ProductModel.ProductModelID. |
| 21 |  | SellStartDate | datetime |  |  | Date the product was available for sale. |
| 22 |  | SellEndDate | datetime | X |  | Date the product was no longer available for sale. |
| 23 |  | DiscontinuedDate | datetime | X |  | Date the product was discontinued. |
| 24 |  | rowguid | uniqueidentifier |  | Default: newid() | ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample. |
| 25 |  | ModifiedDate | datetime |  | Default: getdate() | Date and time the record was last updated. |

### **3.2.2. Challenge**

The AdventureWorks2019 database provides all the tables needed to solve problems: determine the percentage of defective products and cause of defects, the number of products sold, inventory, time and cost in production process

Some tables have NULL values, but these NULL values are well defined and correspond to some special case - for example, NULL in the WorkOrder table could mean that the WorkOrder has not been completed, or NULL in the Product table could mean no product description.

## 3.3. IT requirements Analysis (IT & Infrastructure)

In enterprises, these BI solutions will be essential to the production department in solving problems of managing employees. The application of these solutions will help businesses save a lot of costs, resources. Data management, Data query, and Data analysis are technical requirements that are very necessary to provide a warning about the situation. Data management allows users to monitor, ensure that all analytical processes and insights are in compliance with business policies and procedures, ensure data integrity and minimize risks. Besides that, it must mention "Extract, Transform, Load (ETL) Tools", OLAP and Multidimensional Analysis. Structured Query Language (SQL), from the database extracts information and formats it for consumption and analysis.production.

Data queries can perform deeper calculations, automate tasks, or dig through data mining,helping to uncover hidden trends and relationships between data points.

Data analysis turns raw information into useful insights, helping businesses maximize the value of their data for better business decisions.

Data analytics helps users extract value from operational information and gives them insight into their business.

There are four main types of analysis: descriptive, predictive, indicative and diagnostic. It includes Ad Hoc analysis, Analysis by groups and segments, Cluster analysis, Analyze the scenario and what if it happens, Statistical analysis and regression, Time series analysis and prediction, …

In addition, Power BI is also a necessary technical requirement in order to visualize the employee data, views reports and dashboards, and supports report data that is automatically updated & processed in real time.

This report helps management see how performance impacts business. It also answers the question of whether the hierarchy of positions in the enterprise, salary structure, qualifications, and capabilities of employees has responded.

**CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA**

## 4.1. Designing Data Warehouse

### **4.1.1 Bus Matrix**

Table 4 1. Bus Matrix (Source: Author)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Business Process** |  | **COMMON DIMENSIONS** | | | |
| **DimProduct** | **DimWorkOrder** | **DimLocation** | **DimTime** |
| The amount of goods sold in the past for each item type. | X |  |  | X |
|
| Inventory situation, from which to consider whether to add or reduce. | X | X | X | X |
|
|
| Events that slow down production | X |  | X | X |
| Capture the situation of defective products | X |  | X | X |
| Analysis of product shortage situation | X | X | X | X |
|
|

### **4.1.2 Master Data**

In the Data Warehouse, tables represent data Master Data is the data used to build the dimensions.

Table 4 2. Master Data (Source: Author)

|  |  |  |  |
| --- | --- | --- | --- |
| No | Schema | Subject | Description |
| 1 | Production | ProductInventory | Product inventory information. |
| 2 | Product | Products sold or used in the manufacturing of sold products. |
| 3 | ProductCategory | High-level product categorization. |
| 4 | ProductSubCategory | Product subcategories. See ProductCategory table. |
| 5 | Location | Product inventory and manufacturing locations. |
| 6 | ScrapReason | Manufacturing failure reasons lookup table. |
| 7 | WorkOrder | Manufacturing work orders. |
| 8 | WorkOrderRouting | Work order details. |
| 9 | Sales | SalesOrderDetail | Individual products associated with a specific sales order. See SalesOrderHeader. |
| 10 | SalesOrderHeader | General sales order information. |

### **4.1.3 Transaction Data**

In the Data Warehouse, tables represent data Transaction Data is the transaction data used to construct fact tables

|  |  |  |
| --- | --- | --- |
| **No** | **Subject** | **Description** |
| 1 | WorkOrder | Detailed information on manufacturing work orders. |
| f | Product | Information of products sold or used in the manufacturing of sold products. |

### **4.1.4. Fact and dimension tables**

**4.1.4.1. DimLocation**

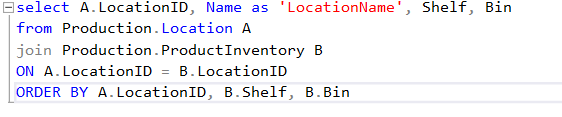


Table of DimLocation

Table 4 3. Table of DimLocation

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow Nulls |
| LocationKey(PK) | smallint |  |
| LocationID | smallint | ☑️ |
| LocationName | nvarchar(50) | ☑️ |
| Shelf | nvarchar(10) | ☑️ |
| Bin | tinyint | ☑️ |

**4.1.4.2. DimProduct**

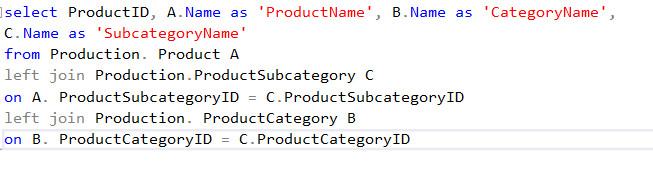


Table of DimProduct:

Table 4 4. Table of DimProduct

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow Nulls |
| ProductKey(PK) | int |  |
| ProductID | int | ☑️ |
| ProductName | nvarchar(50) | ☑️ |
| CategoryName | nvarchar(50) | ☑️ |
| SubcategoryName | nvarchar(50) | ☑️ |
| LastUpdate | datetime | ☑️ |

**4.1.4.3. DimWorkOrder**

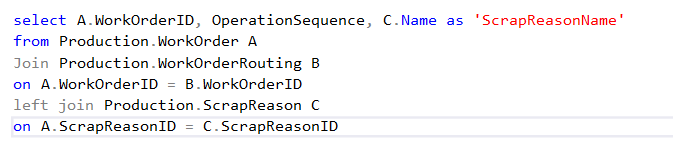


Table of DimProduct:

Table 4 5. Table of DimWorkOrder

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow Nulls |
| WorkOrderKey(PK) | int |  |
| WorkOrderID | int |  |
| OperationSequence | smallint |  |
| ScrapReasonName | nvarchar(50) | ☑️ |

**4.1.4.4. DimTime**

CREATE TABLE DimTime

(

TimeKey int NOT NULL PRIMARY KEY,

Date date NOT NULL,

Week int NOT NULL,

Month int NOT NULL,

Quarter int NOT NULL,

Year int NOT NULL,

DateName nvarchar(15) NOT NULL,

MonthName nvarchar(15) NOT NULL,

QuarterName nvarchar(15) NOT NULL

);

DECLARE @StartDate date = '2000-01-01';

DECLARE @EndDate date = '2025-12-31';

WHILE (@StartDate <= @EndDate)

BEGIN

INSERT INTO DimTime (TimeKey, Date, Week, Month, Quarter, Year, DateName, MonthName, QuarterName)

VALUES (

CONVERT(int, CONVERT(char(8), @StartDate, 112)),

@StartDate,

DATEPART(WEEK, @StartDate),

DATEPART(MONTH, @StartDate)

DATEPART(QUARTER, @StartDate),

DATEPART(YEAR, @StartDate),

DATENAME(WEEKDAY, @StartDate),

DATENAME(MONTH, @StartDate),

'Q' + CAST(DATEPART(QUARTER, @StartDate) AS AST(NVARCHAR) + ' ' + CDATEPART(YEAR, @StartDate) AS NVARCHAR)

);

SET @StartDate = DATEADD(day, 1, @StartDate);

END;

Table of DimTime:

Table 4 6. Table of DimTime

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow NullsNulls |
| TimeKey(PK) | int |  |
| Date | date |  |
| Week | int |  |
| Month | int |  |
| Quarter | int |  |
| Year | int |  |
| DateName | nvarchar(15) |  |
| MonthName | nvarchar(15) |  |
| QuarterName | nvarchar(15) |  |

**4.1.4.5. FactSales**

T-SQL:

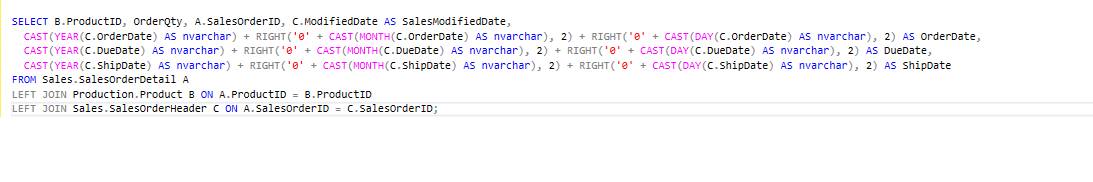


Table of FactSales:

Table 4 7. Table of FactSales

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow Nulls |
| FactSalesKey(PK) | int |  |
| ProductKey | int | ☑️ |
| OrderQty | smallint | ☑️ |
| SalesOrderID | int | ☑️ |
| SalesModifiedDate | datetime | ☑️ |
| OrderDate | int | ☑️ |
| DueDate | int | ☑️ |
| ShipDate | int | ☑️ |

**4.1.4.6. FactInventory**

T-SQL:

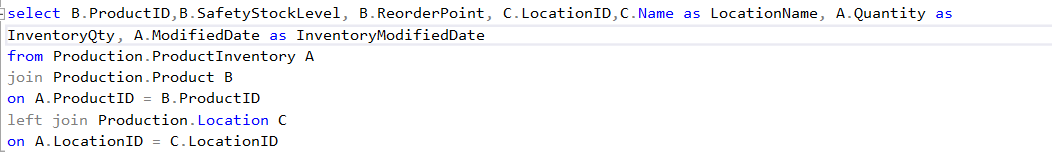


Table of FactInventory:

Table 4 8. Table of FactInventory

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow Nulls |
| FactInventoryKey(PK) | int |  |
| ProductKey | *int* | ☑️ |
| LocationKey | *smallint* | ☑️ |
| LocationName | *nvarchar(50)* | ☑️ |
| InventoryQty | *smallint* | ☑️ |
| InventoryModifiedDate | *datetime* | ☑️ |

**4.1.4.7. FactProduction**

T-SQL:

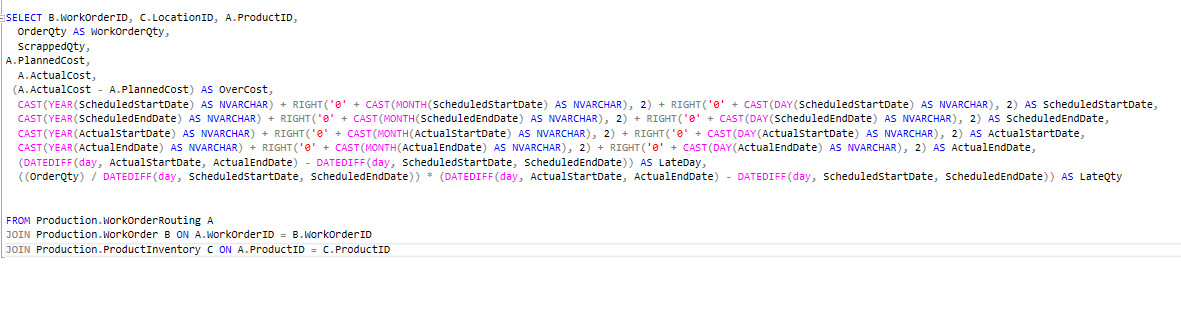


Table of FactProduction:

Table 4 9. Table of FactProduction

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Allow Nulls |
| FactProductionKey(PK) | int |  |
| LocationKey | smallint | ☑️ |
| ProductKey | int | ☑️ |
| WorkOrderKey | int | ☑️ |
| WorkOrderQty | int | ☑️ |
| StockedQty | int | ☑️ |
| ScrappedQty | smallint | ☑️ |
| InventoryQty | smallint | ☑️ |
| SafetyStockLevel | smallint | ☑️ |
| ReorderPoint | smallint | ☑️ |
| ScheduledStartDate | int | ☑️ |
| ScheduledEndDate | int | ☑️ |
| ActualStartDate | int | ☑️ |
| ActualEndDate | int | ☑️ |
| LateDate | int | ☑️ |
| LateQty | int | ☑️ |
| PiannedCost | money | ☑️ |
| ActualCost | money | ☑️ |
| OverCost | money | ☑️ |

### **4.1.5. Data Warehouse model (Snowflake or Star or Galaxy)**

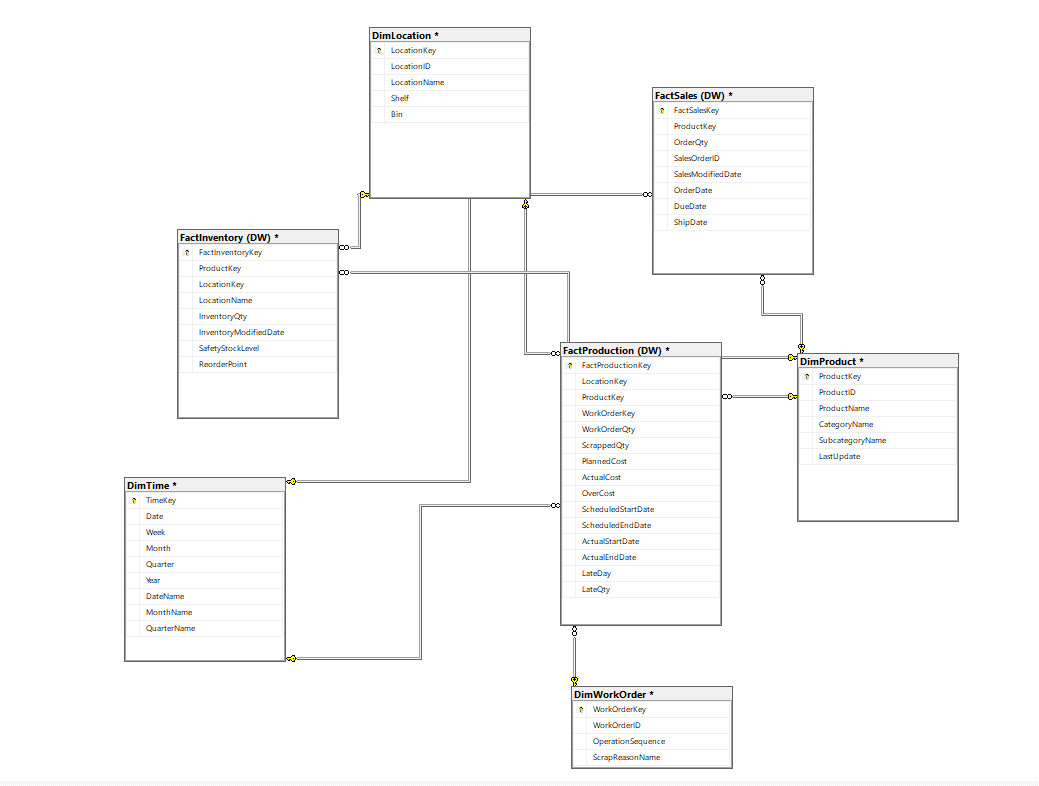


Figure 4 1. Data Warehouse Schema

## 4.2. ETL processes

### **4.2.1. Dimension Table’s ETL Process**

**4.2.1.1. DimLocation’s ETL Process**

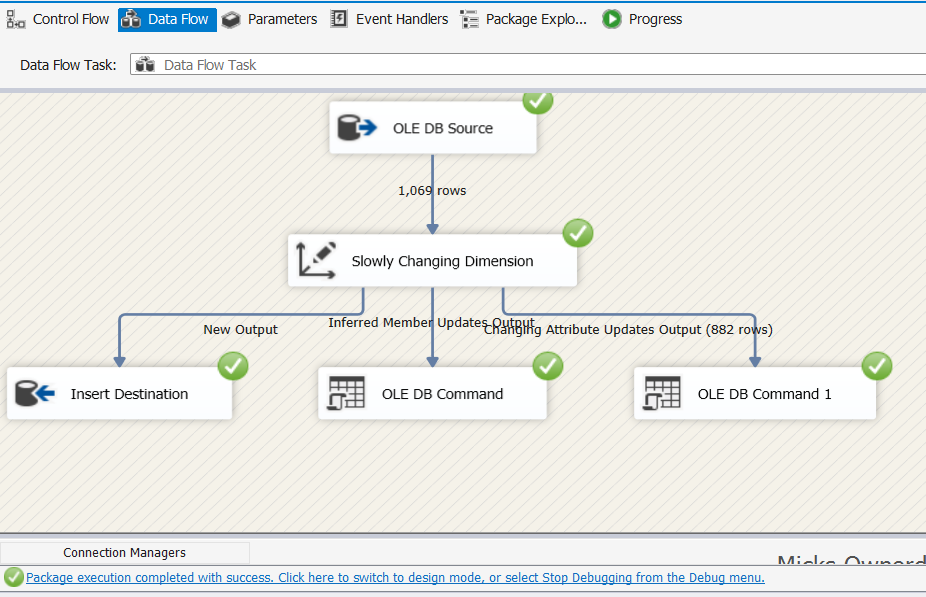


Figure 4 2. SSIS for DimLocation (source:Author)

Figure 4.2 Describes the ETL process for the DimLocation table.

To create the DimLocation table, we used T-SQL to aggregate the data from the Data Source and used Slowly Changing to load the data into the Destination

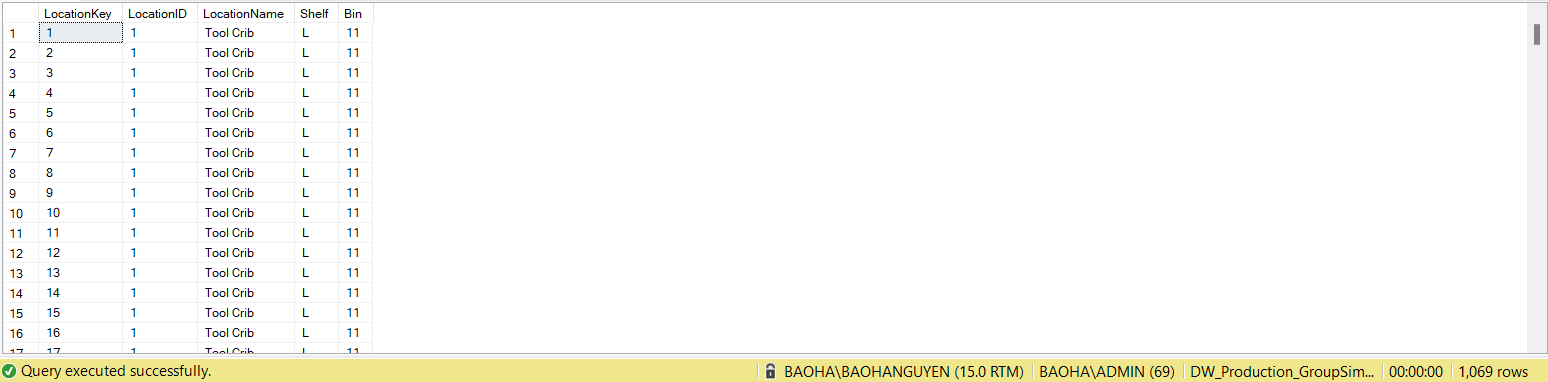


Figure 4 3. Result DimLocation (source:Author)

Figure 4.3 shows the results of the above process, DimLocation Table has 1,069 rows, including 4 tables: LocationKey, LocationID, LocationName,Sheft,Bin to fully represent the data dimension of Location

**4.2.1.2. DimProduct’s ETL Process**

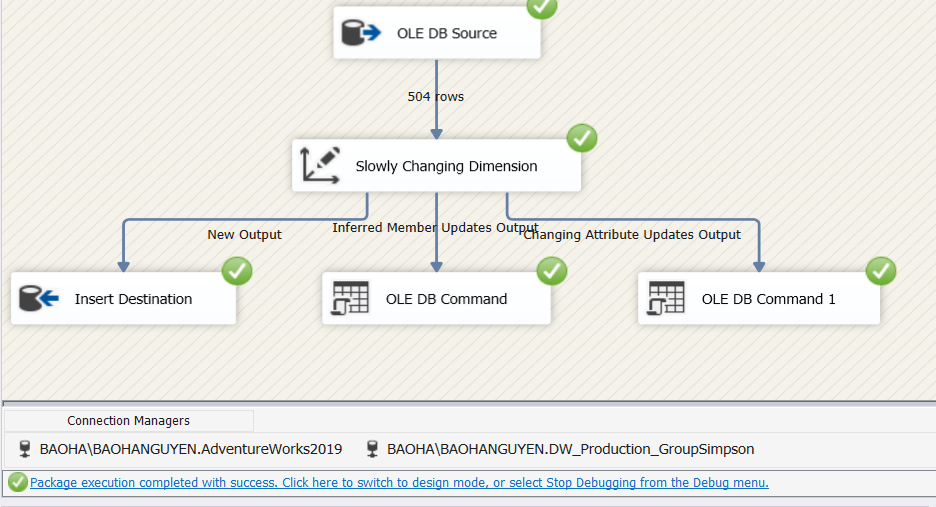


Figure 4 4. SSIS for DimProduct (source:Author)

Similar to DimLocation, Figure 4.4 Describes the ETL process for the DimProduct table. To create the DimProduct table, we used T-SQL to aggregate the data from the Data Source and used Slowly Changing to load the data into the Destination

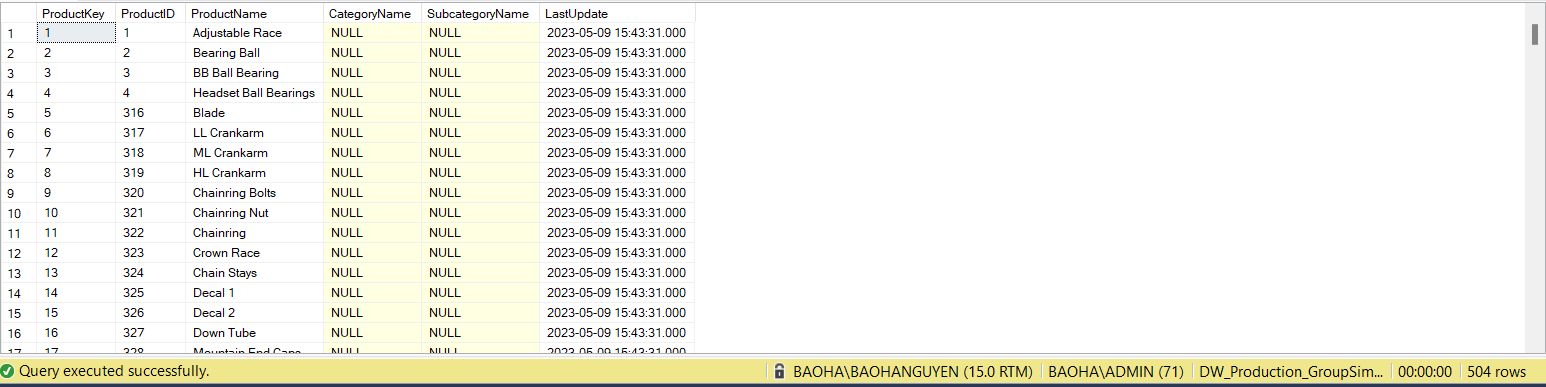


Figure 4 5. Result DimProduct (source:Author)

Figure 4.5 shows the results of the above process, DimProduct Table has 504 rows, including 6 tables: ProductKey, ProductID, ProductName, SubCategoryName, CategoryName and LastUpdate, to fully represent the data dimension of Product

**4.2.1.3. DimWorkOrder’s ETL Process**

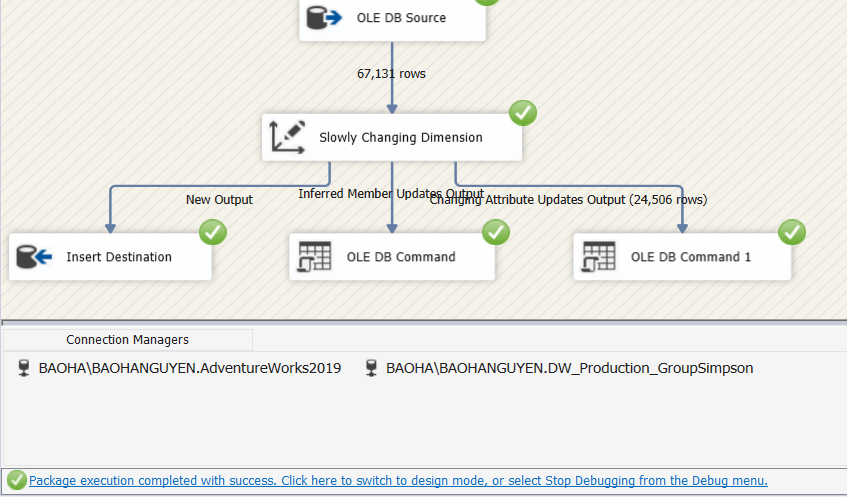


Figure 4 6. SSIS for DimWorkOrder (source:Author)

Similar, Figure 4.6 Describes the ETL process for the DimWorkOrder table. To create the DimWorkOrder table, we used T-SQL to aggregate the data from the Data Source and used Slowly Changing to load the data into the Destination. We want to query the steps in the production process as well as the reason for the product failure from this table.

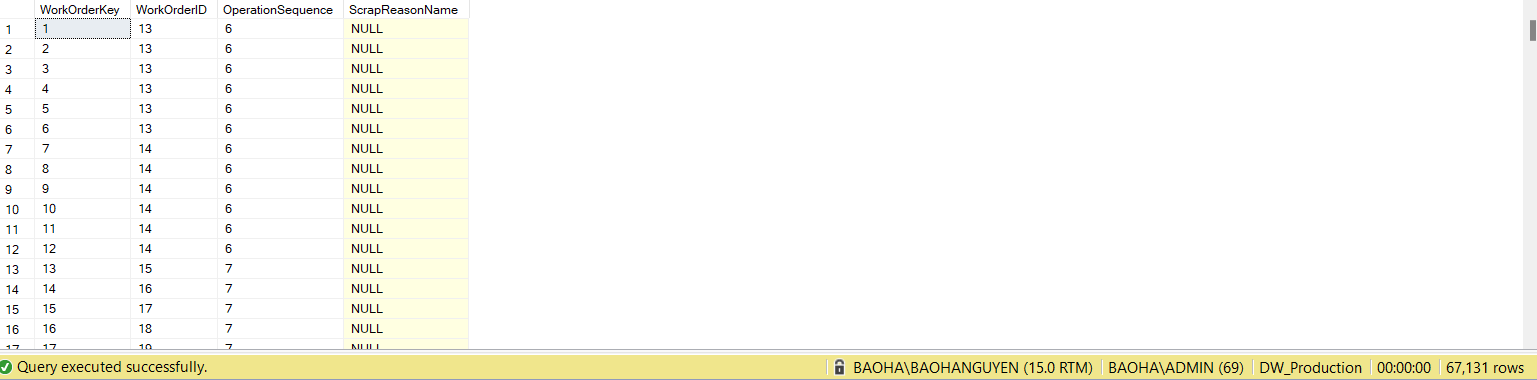


Figure 4 7. Result DimWorkOrder (source:Author)

Figure 4.7 shows the results of the above process, DimWorkOrder Table has 67.131 rows, including 4 tables: WorkOrderKey, WorkOrderID, OperationSequence,ScrapReasonName, to fully represent the data dimension of WorkOrder

**4.2.1.4: DimTime’s ETL Process**

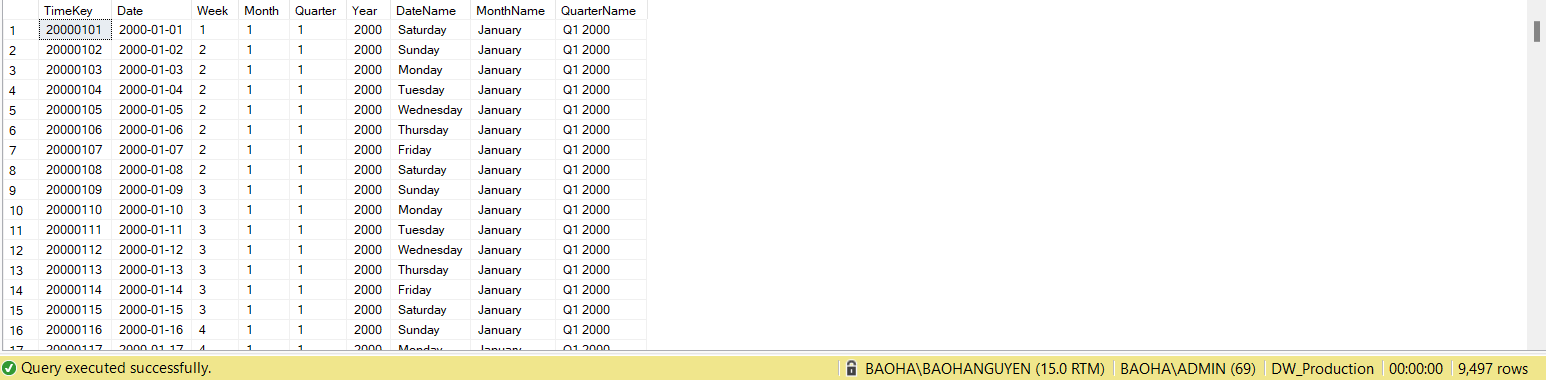


Figure 4 8. Result DimTime (source:Author)

For DimTime, we only use T-SQL to query the time dimension, from 2000 to 2025, including the years of operation of the company. Figure 4.8 shows the results of both DimTime, detailed query by Year, Quarter, Month, Day

### **4.2.2. Fact Table’s ETL Process**

**4.2.2.1. FactInventory’s ETL Process**

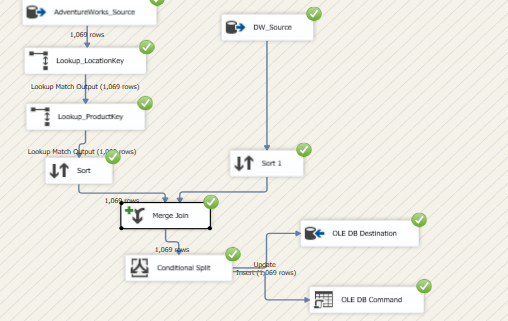
****

Figure 4 9. SSIS for FactInventory (source:Author)

Figure 4.9 shows the FactInventory table ETL process.

- First, we will use T-SQL to query the data columns in the Data Source

- Lockup with Dim: Lockup LocationKey by mapping with LocationID, with ProductKey by mapping with ProductID

- Sort the data of the table by InventoryQty: quantity of inventory

- Create SQL Server FactInventory table to Merge Join 2 tables together

- Set conditions to create 2 tables: Insert and Update

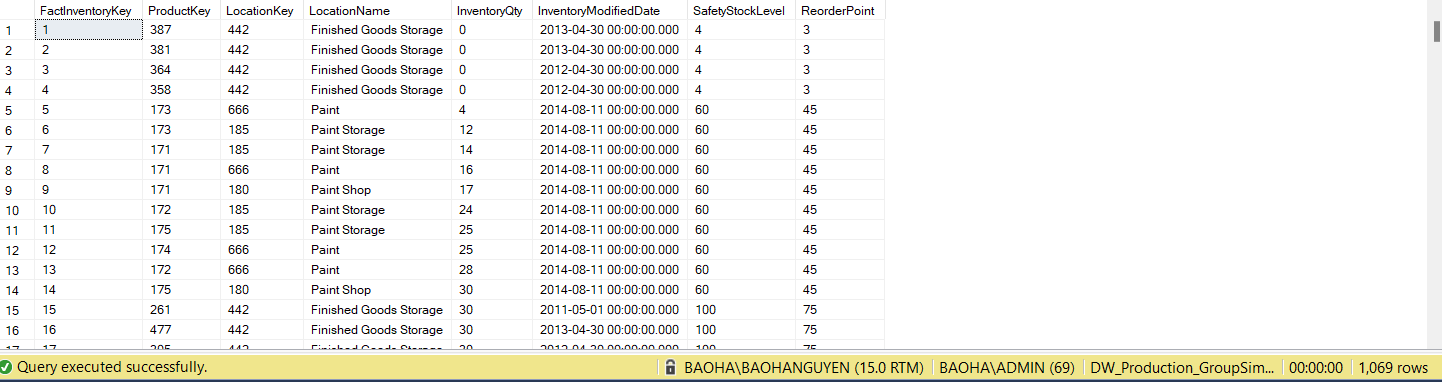


Figure 4 10. Result FactInventory’s ETL Process (source:Author)

Figure 4.10 shows the result of the ETL process. The data entered into the table is 1,069 rows

**4.2.2.1. FactSales’s ETL Process**

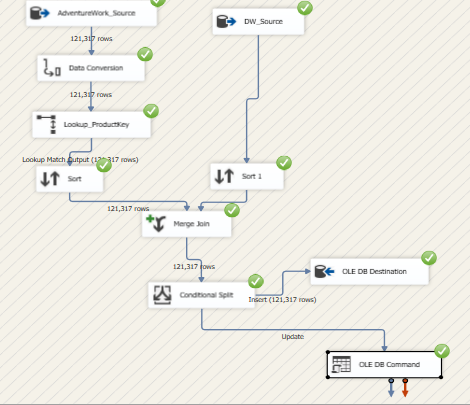
****

Figure 4 11. SSIS for FactSales (source:Author)

Figure 4.11 shows the FactSales table ETL process.

- First, we will use T-SQL to query the data columns in the Data Source

- Convert columns: OrderDate,DueDate,ShipDate from nvarchar to int data type

- Lockup with Dim: Lockup ProductKey by mapping with ProductID

- Sort the data of the table by OrderQty

- Create SQL Server FactSales table and Merge Join 2 tables together

- Set conditions to create 2 tables: Insert and Update

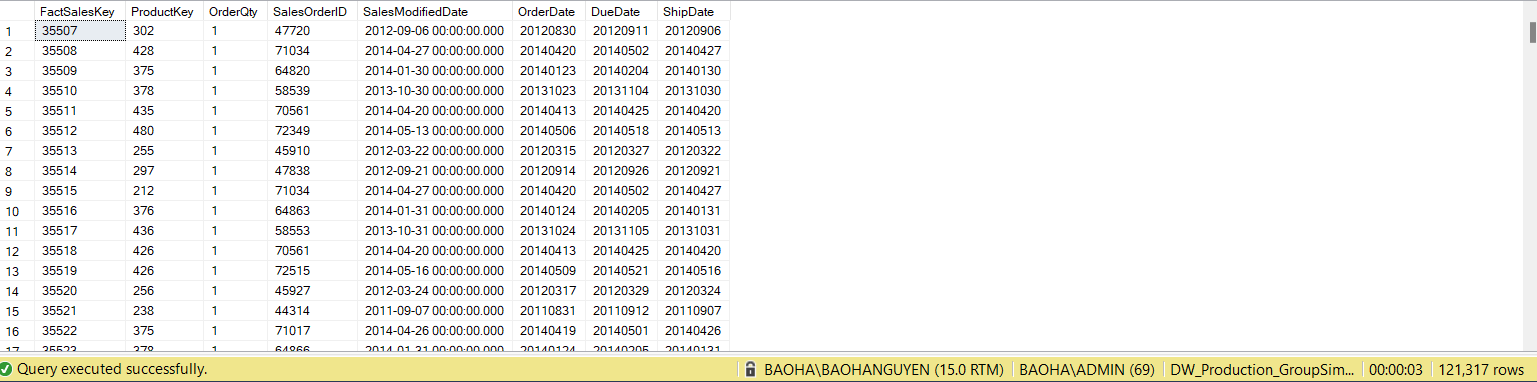


Figure 4 12. Result FactSales’s ETL Process (source:Author)

Figure 4.12 shows the result of the ETL process. The data entered into the table is 12.317 rows contains the necessary measures related to Sales

**4.2.2.1. FactProduction’s ETL Process**

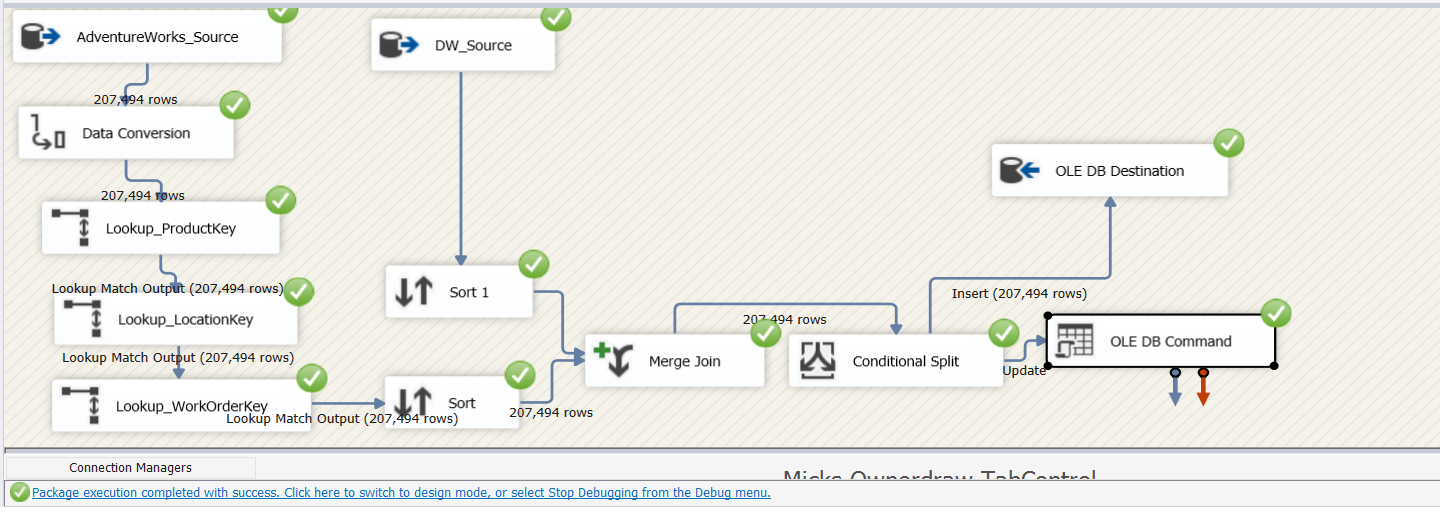
****

Figure 4 13. SSIS for FactProduction (source:Author)

Figure 4.13 shows the FactProduction table ETL process.

- First, we will use T-SQL to query the data columns in the Data Source

- Lockup with Dim: Lockup LocationKey by mapping with LocationID, with ProductKey by mapping with ProductID, with WorkOrderKey by mapping with WorkOrderID.

- Sort the data of the table by LateDay

- Create SQL Server FactInventory table to Merge Join 2 tables together

- Set conditions to create 2 tables: Insert and Update

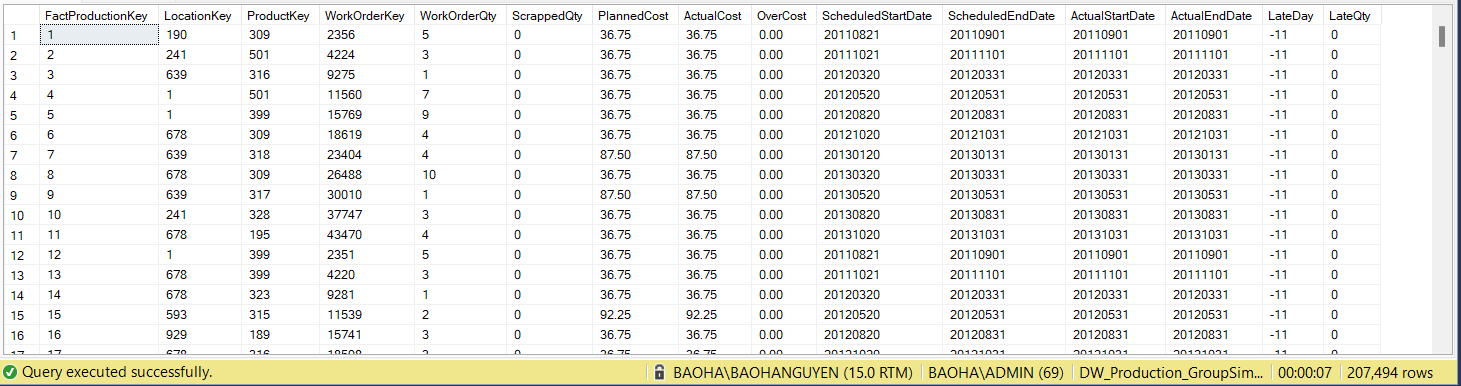


Figure 4 14. Result FactProduction’s ETL Process (source:Author)

Figure 4.114 shows the result of the ETL process. The data entered into the table is 207.494 rows contains the necessary measures related to Production

**CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION**

## 5.1. Report and dashboard systems (structure)

### **5.1.1. Building the CUBE**

Overall structure of SSAS solution:

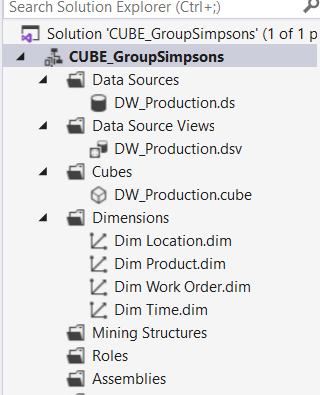


Figure 5 1. Building the CUBE

To analyze data from DW data warehouse and build intelligent reports to support decision making, the thesis uses SSAS tool. SSAS is integrated in the BIDS (Business Intelligence Development Studio) tool that is included with SQL Server.

Building the cube includes 4 steps:

* Step 1: Create Project in SSAS
* Step 2: Create Data sources
* Step 3: Create Data source Views combined with creating Measures
* Step 4: Create Cube combined with creating Dimension

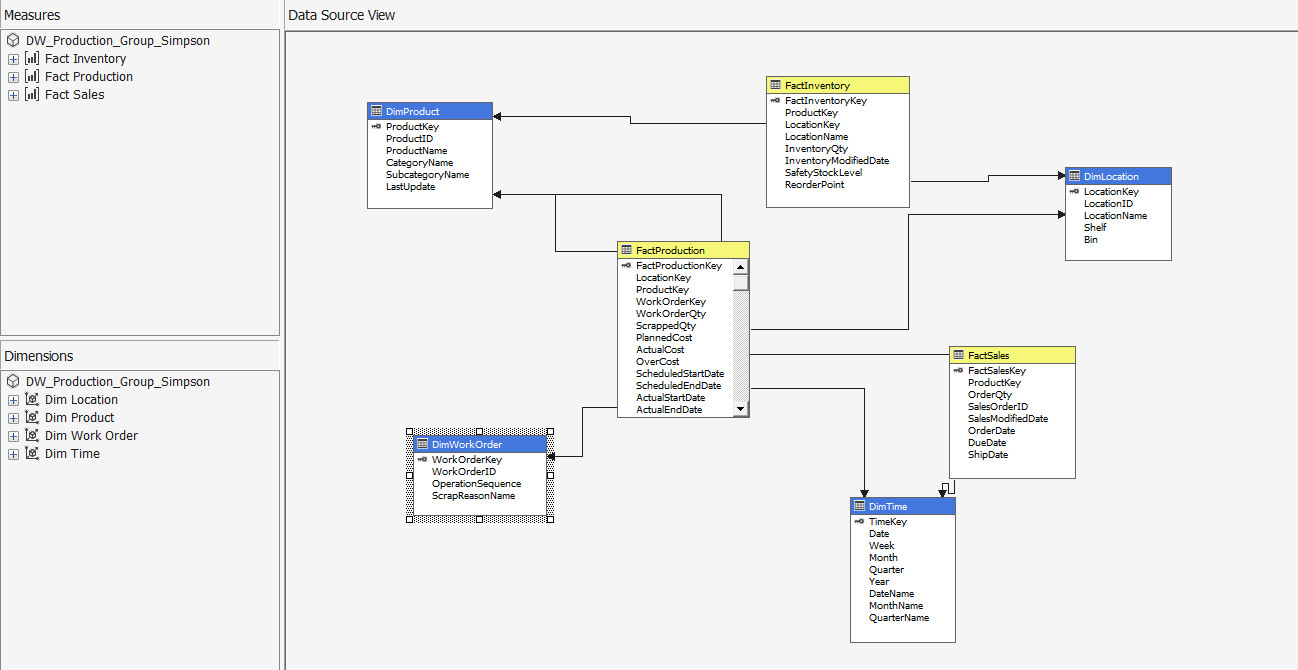


Figure 5 2. CUBE Model

### **5.1.2. Building KPI System**

**5.1.2.1. Building Calculation**

Product defect rate:

* The product defect rate is the percentage or proportion of defective products produced by a manufacturing process.
* It is calculated by dividing the number of defective products by the total number of products produced.

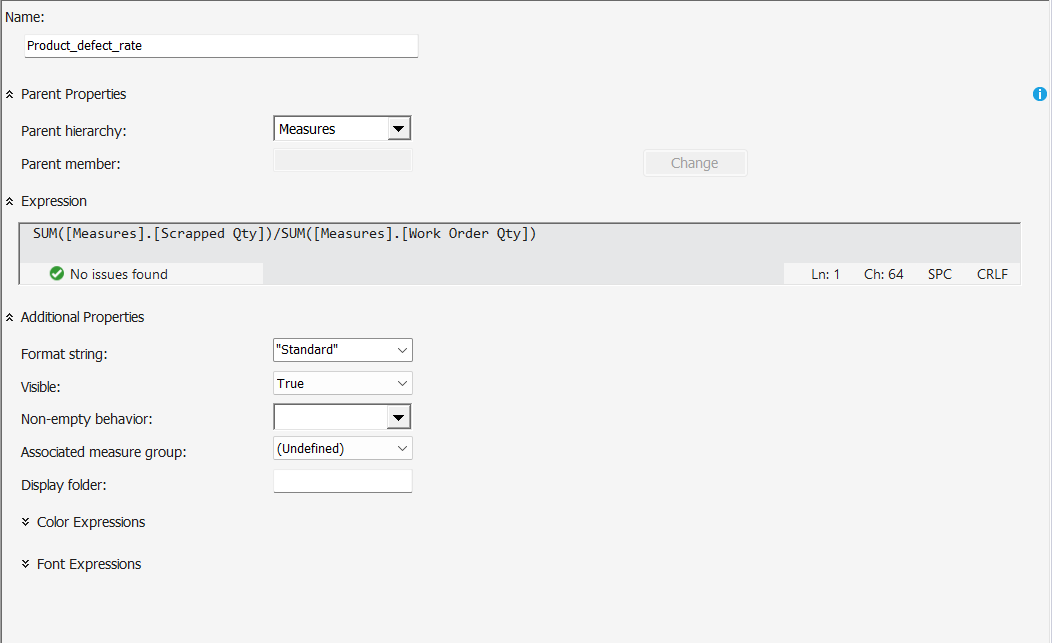


Figure 5 3. Building Calculation product defect rate( source: author)

Production Cost:

* Production cost refers to the total expenses incurred in manufacturing a product or providing a service. These costs include direct materials, labor, and overhead expenses such as rent, utilities, and other operational expenses.

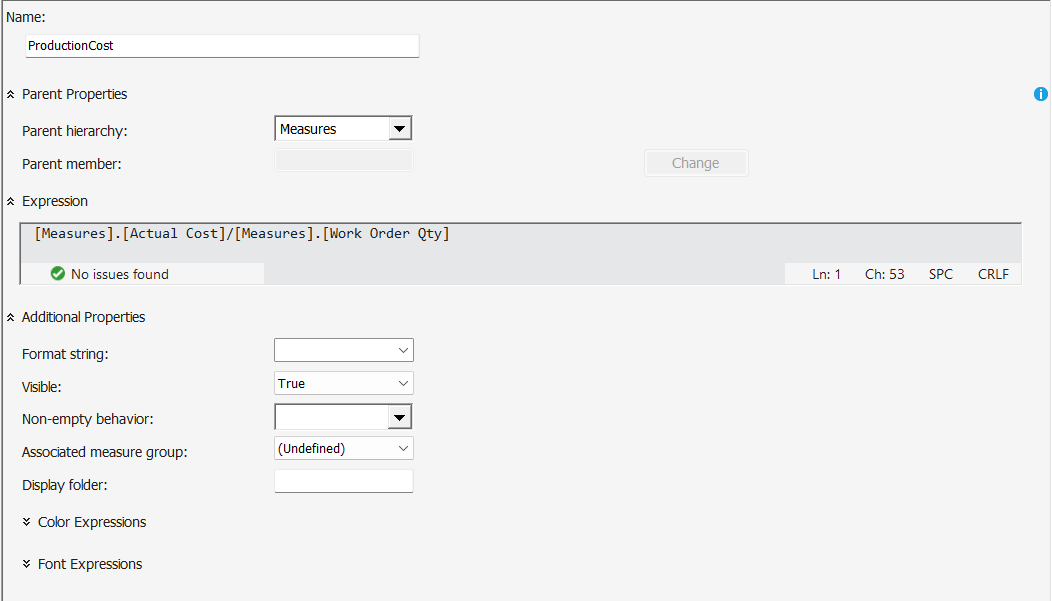


Figure 5 4. Building Calculation product cost( source: author)

**5.1.2.2. KPI System**

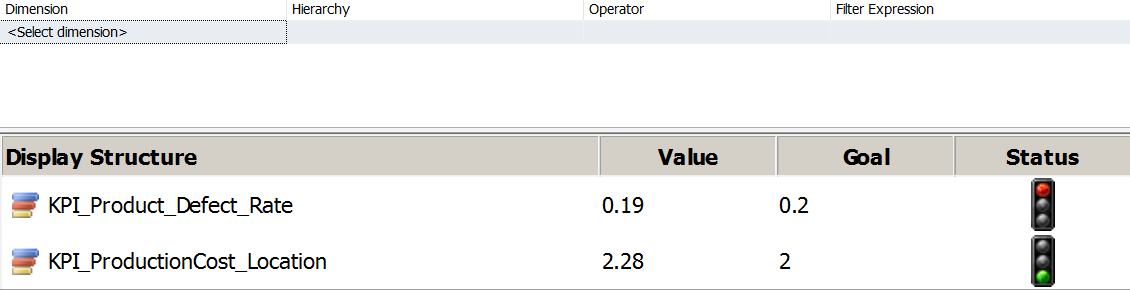


Figure 5 5. Display KPI ( source: author)

Status light:

* Green color: shows that the target has been reached.
* Yellow color: shows the actual value has not reached the target but is still at acceptable level receive.
* Red color: not acceptable, the actual result is too low compared to the target.

KPI of Product defect rate: Track the defective production rate of each item. Set a KPI for this ratio (0.2) to control production quality.

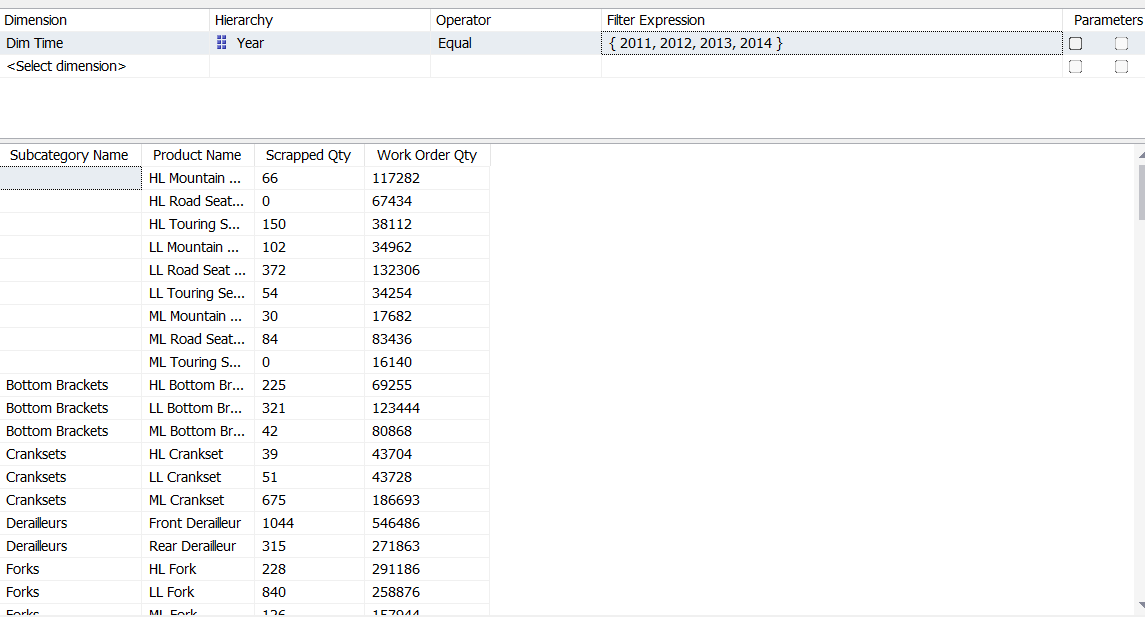


Figure 5 6. KPI of Product defect rate( source: author)

Figure: Number of scrapped quantity and products produced of each product (2011 - 2014).

Scrapped Quantity of each product and number of products produced of each product.Here the reference group by the number of years of operation of the company (from 2011-2014).



Figure 5 7. KPI of Product defect rate( source: author)

KPI of Production Cost: Track production costs of each production unit, set KPIs for each location.

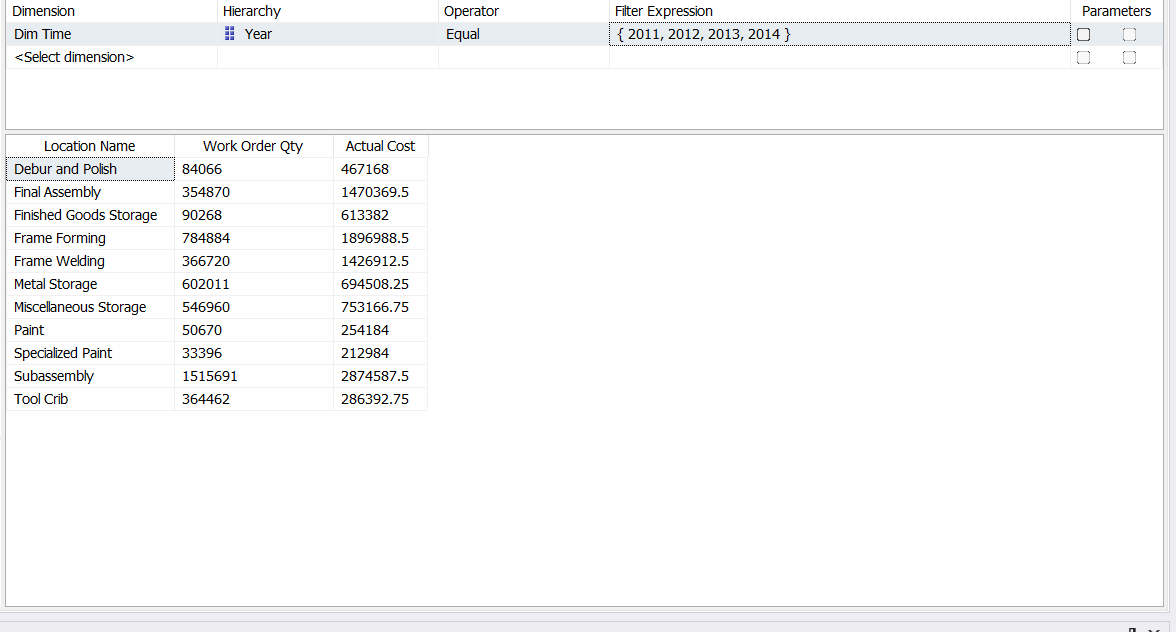


Figure 5 8. Number of Work Order Quantity and Actual Cost of each Location (2011-2014) ( source: author)

Actual production cost and production quantity of each Location.

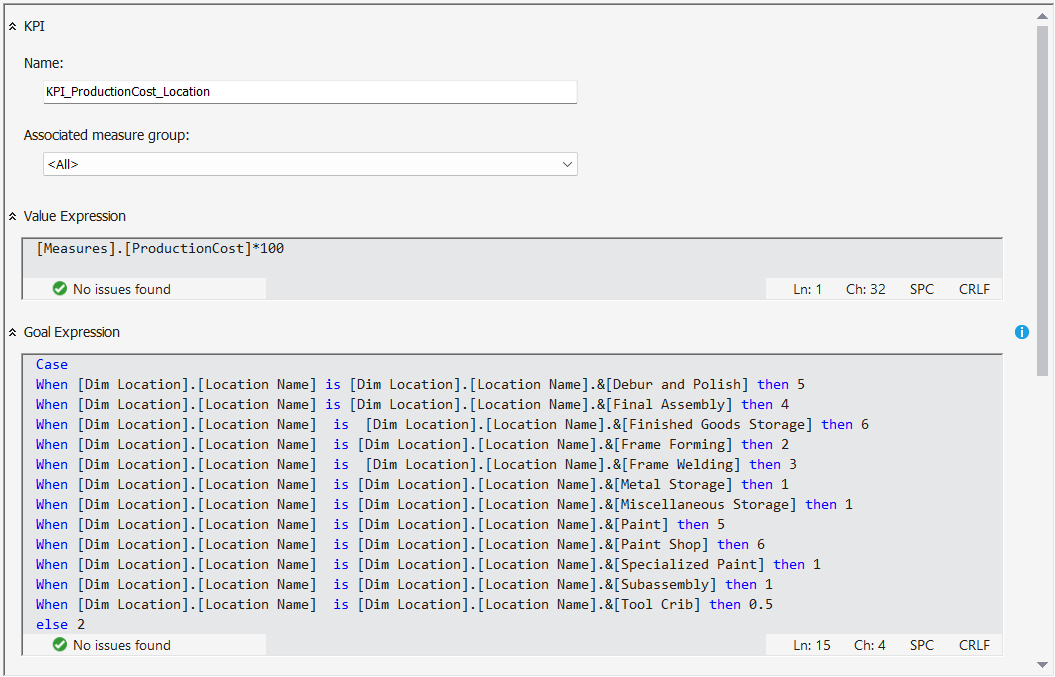


Figure 5 9. KPI of Production Cost( source: author)

Result:

* KPI of Product defect rate:

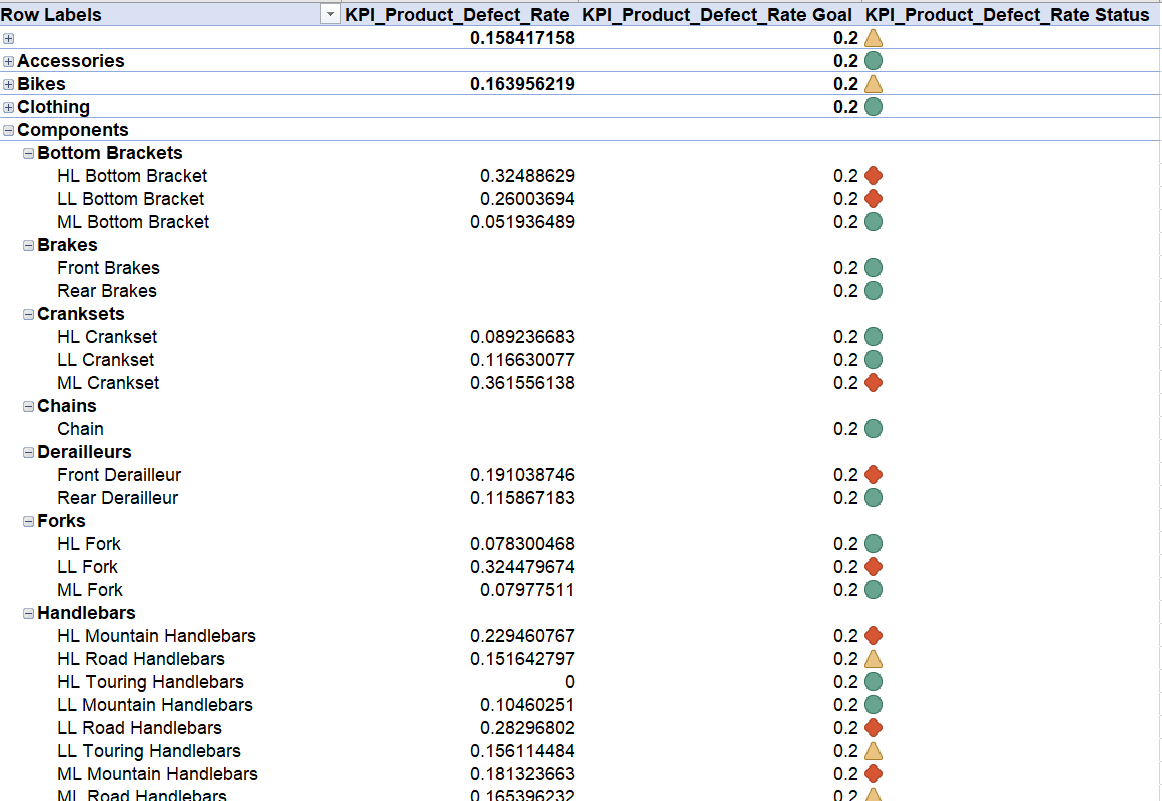


Figure 5 10. Result of KPI of Product defect rate each Products( source: author)

* KPI of Production Cost:

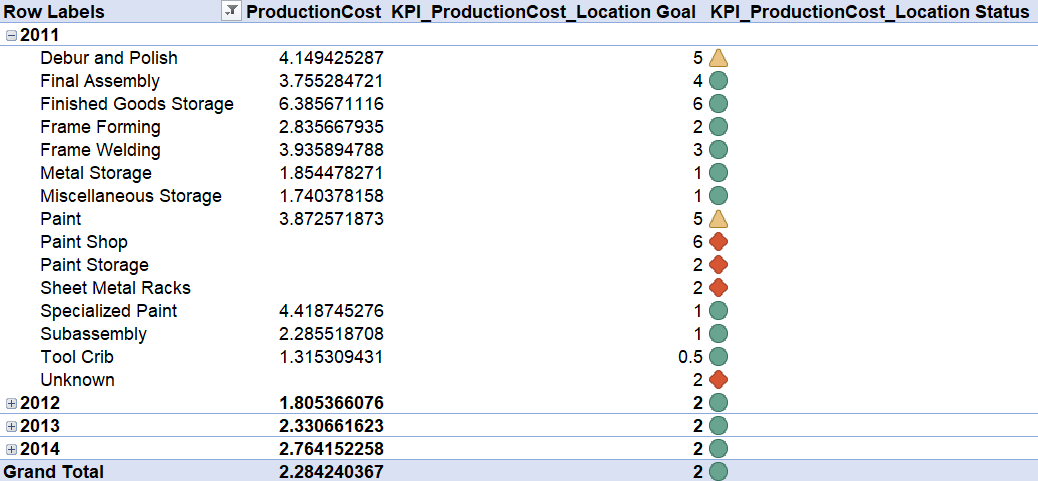


Figure 5 11. Result of KPI of Product defect rate each Location( source: author)

## 5.2. Data analysis with Power BI

With PowerBI, our group analyze and design the dashboard based on the data from the Production DataWarehouse in the following:

### **5.2.1 SalesOrder**



Figure 5 12. Dashboard of SalesOrder over the years

From Dashboard in Figure 5.12, the total number of orders is 3,774, the total number of products sold is 63,000 with 504 products.

The 5 Best-selling products are: Full-Finger Gloves, L (2,100 products), Long-Sleeve Logo Jersey, L (2.00 products), AWC Logo Cap (1,900 products), Sport-100 Helmet, Blue (1,400 products), Full-Finger Gloves, M (1,400 products).

5 products sold at Least include: HL Road Frame - Black, 48(48 products), HL Road Frame - Red, 48(48 products), ML Road Frame - Red, 52(35 products), Mountain Bike Socks, L(24 products), HL Mountain Frame - Black, 44 (8 products)

The team analyzed the increase and decrease in the number of products over the years from 2011 to 201, giving the results: Subcategory Road Bikes in 2012 sold the most with 17,598 products, an increase of 12,290 products compared to the previous year.

### **5.2.2. Inventory**

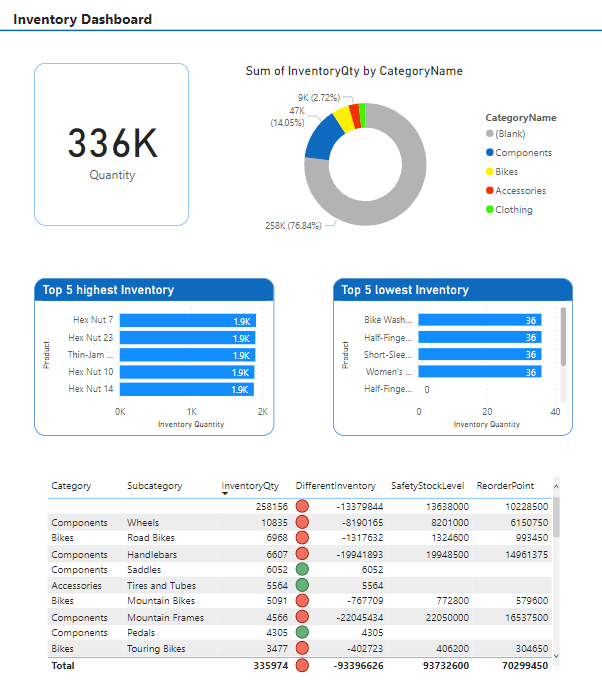


Figure 5 13. Dashboard of Inventory by Product, Category, Subcategory

In this dashboard about *Figure 5.13*, Total InventoryQty is about 336,000 products, including Component: 47,214 products (accounting for 14.05%); Bike: 15,536 products (accounting for 4.62%); Accessories: 9,128 products (accounting for 2.72%); Clothing: 5,940 products (accounting for 1.77%); the rest Blank accounted for 258,156 products (accounting for 76.84%).

The 5 products with the Highest inventory are: Hex Nut 7 (1911 products), Hex Nut 23 (products), Thin-Jam Hex Nut 3 (1901 products), Hex Nut 10 (1888 products), Hex Nut 14 (1880 products).

The 5 products with the Lowest inventory are: Bike Wash - Dissolver (36 products), Half-Finger Gloves, L (36 products), Short-Sleeve Classic Jersey, L (36 products), Women's Tights, L (36 products).

At the same time, our team compared InventoryQty with SafetyStockLevel to check which items were missing by how much through the DifferentInventory column, the results showed:

Subcategory Mountain Frames with InventoryQty: 4,566, SafetyStockLevel: 22,050,000 causing a shortage of 22,045,434 products; Subcategory Handlebars with InventoryQty: 6,607, SafetyStockLevel: 19,941,893 causing a shortage of 19,948,500 products; Subcategory Wheels with InventoryQty: 10,835, SafetyStockLevel: 8,201,000 caused a shortage of 8,190.165 products.

### **5.2.3 Scrapped**

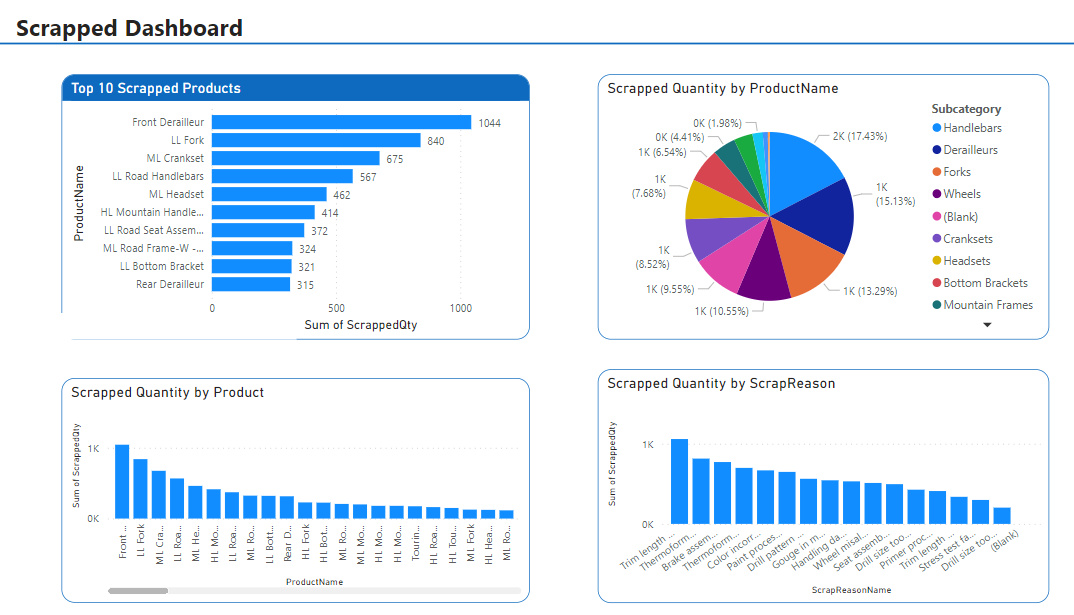


Figure 5 14. Dashboard of Scrapped by Product and ScrapReason

From Dashboard in *Figure 5.14*, Top 10 products with the highest number of scraps with Front Derailleur with 1044.

The group also analyzed Scrapped Quantity by Subcategory with Subcategory Handlebars having the highest number of scraps at 1566, accounting for 17.43%.

Next, the team analyzes Scrapped Quantity by Product in descending order and Scrapped Quantity by ScrapReason.

### **5.2.4 LateDay**

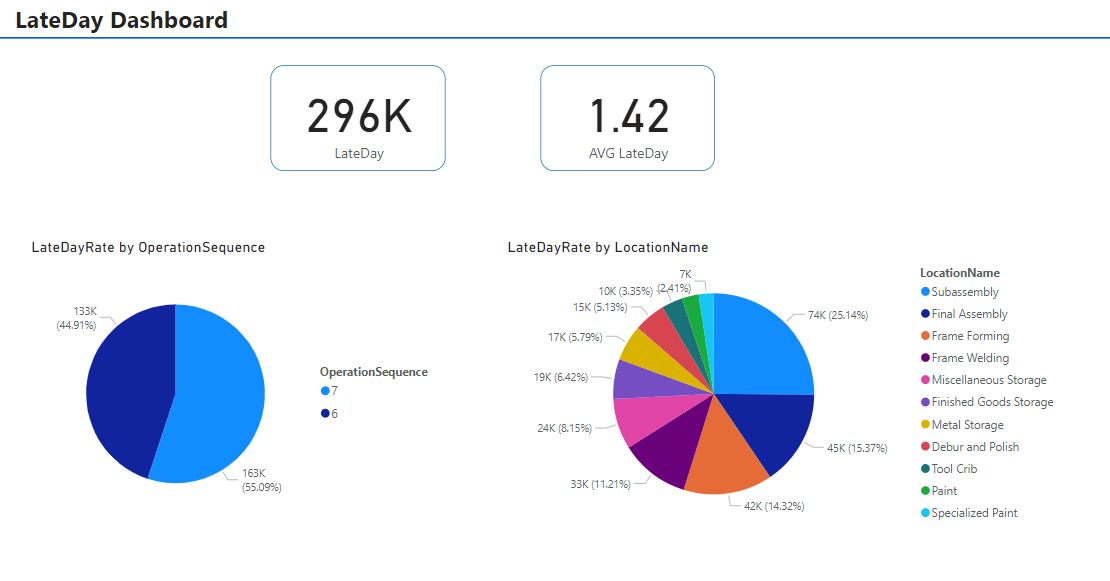


Figure 5 15. Dashboard of LateDay by OperationSequence and Location

In this dashboard about *Figure 5.15*,The total number of days behind schedule is about 296,000 products.

The average number of days late is 1.42 days.

The LateDayRate by OperationSequence pie chart shows that the percentage of late days in OperationSequence 6: 132,749 days equivalent to 44.91%, OperationSequence 7: 162,810 days equivalent to 55.09%.

The location with the most product delays is Subassembly with 74,298 sản phẩm products of late quantity.

## 5.3. Evaluation and Discussion

**CHAPTER 6: CONCLUSION AND FUTURE WORKS**

## 6.1. Results

Solved the questions posed at the beginning of the report, including:

* Predict future production quantity based on:
* Past purchase history to know the number of products sold of each item over the years, thereby seeing the increase or decrease in the number of products.
* Review the amount of inventory from which to plan future production.
* Determine the percentage of defective products and the cause of defective products.
* Analyze the number of products that are overdue, find out the LocationNames with the most products that are late.

From the above analysis, it is possible to predict future production plans to help optimize the production process, reduce the amount of redundant products, and meet the required output for the business.

## 6.2. Limitations

Due to the limited time, the report is flawed such as not solving more problems and not giving many insights, Time series and Forecasting or Predictive model (Tableau, R, or Python) have not yet been applied to the project.

## 6.3. Future works

In the future, we will collect and analyze more factors in the production module related to people, machines and equipment to have a full assessment in more aspects.

Learn more Time Series and Forecasting or Predictive Modeling (Tableau, R or Python).

**PROJECT MEMBER EVALUATION (Signed by all members)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Full name** | **Task** | **Contribution** | **Signed** |
| 1 | Nguyễn Thị Mỹ Dung | 1. Analyze topics, select topics, identify problems  2. Identify data source  3. Business processes, Objectives of the project  4. Building data warehouse  5. Visualization  6. Word | 100% |  |
| 2 | Nguyễn Thị Bảo Hà | 1. Analyze topics, select topics, identify problems  2. Identify data source  3. Value and desired outcome of the project  4. ETL Table  5. SSAS, KPI  6. Slides | 100% |  |
| 3 | Nguyễn Mỹ Dung | 1. Analyze topics, select topics, identify problems  2. Identify data source  3. Theoretical Basis  4. ETL Table  4. Visualization  5. Slides | 100% |  |
| 4 | Cù Thị Mỹ Duy | 1. Analyze topics, select topics, identify problems  2. Theoretical Basis  3. Building data warehouse: Identify Master Data  4. SSAS, KPI  5. Word | 100% |  |
| 5 | Nguyễn Thị Cẩm Giang | 1. Analyze topics, select topics, identify problems  2. Identify data source  3. Building data warehouse  4. SSAS, KPI  5. Word | 100% |  |

**REFERENCES**

[1]Datadictionary.AdventureWorks.(n.d.).<https://dataedo.com/samples/html/AdventureWorks/doc/AdventureWorks_2/home.html>

[2]  Stedman, C., & Burns, E. (2023, February 10). What is Business Intelligence (BI)?: Definition from TechTarget. Business Analytics.<https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI>

[3]  What is ETL (extract, transform, load)?. IBM. (n.d.). <https://www.ibm.com/topics/etl>

[4]  Dive into anything. Reddit. (n.d.). <https://www.reddit.com/r/AZURE/comments/x3zcym/what_is_azure_analysis_services/>

[5]  Strauss, L. (2022, November 4). What is business intelligence? methods + examples. Zapier.\<https://zapier.com/blog/what-is-business-intelligence/>Dive into anything.

[6]  eddit. (n.d.). <https://www.reddit.com/r/AZURE/comments/x3zcym/what_is_azure_analysis_services/>

[7]  Josh Fruhlinger and Mary K. Pratt, "What is business intelligence? Transforming data into business insights," 16 October 2019

[8]  "SQL Server 2019 Pricing," Microsoft, [Online]. Available: https://www.microsoft.com/en-us/sql-server/sql-server-2019- pricing?fbclid=IwAR1e6NSJfP1dgJm3keQ6fzMMsIyWcoVcc8ukNjgO1KSBGG8QgfC4j2j\_I0. [Accessed 25 May 2022].

[9]  Hannah Wilken, "5 reasons you should measure employee performance".Performance Management.

[10]  Bergen Adair, "Ultimate Guide to the Best Visual Analytics Tools".Business Intelligence.

[11] Josh Fruhlinger and Mary K. Pratt, "What is business intelligence? Transforming data into business insights," 16 October 2019.

 [12] M. K. Pratt, "TechTarget," Business Intelligence Architecture, July 2020. [Online]. Available: https://www.techtarget.com/searchbusinessanalytics/definition/businessintelligence-architecture. [Accessed 6 May 2022].

[13]  *What is business intelligence? transforming data into business insights*. CIO. (2023, January 20).https://www.cio.com/article/272364/business-intelligence-definition-and-solutions.html