UNIVERSITY OF ECONOMICS AND LAW
FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

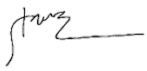
BUSINESS INTELLIGENCE AND DECISION SUPPORT SYSTEMS

TOPIC: WIDE WORLD IMPORTERS BI SOLUTIONS: SALES, WAREHOUSE & PURCHASE DEPARTMENT

Lecturer: Hồ Trung Thành, Nguyễn Văn Hồ and Lê Bá Thiền
Group: *TLLAG*

Ho Chi Minh City, May 8th, 2023

Members of Group TLLAG

STT	Fullname	Student ID	Task	Completed	Signed
1	Hoàng Nguyễn Quỳnh Giang	K204161987	Make Business Requirement, Data Modelling, ETL	100%	
2	Lê Thị Thùy Linh	K204161990	Make Business Requirement, Data Modelling, ETL	100%	
3	Nguyễn Thị Ái Linh	K204161991	Make Business Requirement, Data Modelling, Visualization	100%	
4	Bùi Thị Tri Tâm	K204161999	Make Business Requirement, Data Modelling, Visualization	100%	
5	Đào Phương Anh	K204160660	Make Business Requirement, Data Modelling, Visualization	100%	

ACKNOWLEDGMENTS

My team would like to thank Mr. Ho Trung Thanh, Mr. Nguyen Van Ho and Mr. Le Ba Thien - Lecturer of Business intelligence and decision support systems 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. TLLAG

COMMITMENT

We hereby declare that the above project is the research work of our group under the guidance of lecturer Ho Trung Thanh, lecturer Nguyen Van Ho and 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. TLLAG

CONTENTS

Members of Group TLLAG ..... **i**
ACKNOWLEDGMENTS ..... **ii**
COMMITMENT ..... **iii**
CONTENTS ..... **iv**
CONTENT OF FIGURE ..... **vi**
CONTENT OF TABLE ..... **vii**
CHAPTER 1: INTRODUCTION **1**
1.1 Business case for the project **1**
1.2 Objectives of the project **1**
1.2.1 General Objective ***1***
1.2.2 Specific Objectives ***1***
1.3 Research Objects **1**
1.4 Scope of the project **2**
1.5 Value and desired outcome of the project **2**
1.6 Structure of project **3**
CHAPTER 2: THEORETICAL BASIS **4**
2.1 Overview about BI **4**
2.1.1 What is BI? ***4***
2.1.2 BI Architecture ***4***
2.1.3 Advantage of BI in enterprises ***4***
2.1.4 BI Strategy for Business ***5***
2.2 ETL Process **6**
2.2.1 What is ETL? ***6***
2.2.2 Why do we need ETL? ***6***
2.2.3 ETL Process ***7***
2.3 Data warehouse and Data mart **7**
2.3.1 What are Data warehouse and Data mart? ***7***
2.3.2 Who needs Data warehouse and Data mart? ***8***
2.3.3 Advantages and disadvantages of Data warehouse ***8***
2.3.4 Snowflake and Star schemas ***9***
2.4 KPIs **9**
2.4.1 KPIs Definition ***9***
2.4.2 The advantages and disadvantages of KPIs ***10***
2.4.3 Categories of KPIs (Sales, Warehouse and Purchasing) ***10***
2.5 Microsoft Azure **12**

2.5.1 What is Microsoft Azure? *12*
2.5.2 Advantages and disadvantages *13*
CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION 14
3.1 Business processes **14**
3.1.1 Business department *14*
3.1.2 The purpose of each department *14*
3.1.3 Business process *16*
3.2 Business Requirements Analysis **19**
3.2.1 Sales Department *19*
3.2.2 Warehouse Department *20*
3.2.3 Purchase Department *20*
3.3 Data source and challenges **21**
3.4 IT requirements Analysis (IT & Infrastructure) **21**
CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA **23**
4.1 Designing Data Warehouse **23**
4.1.1 Bus Matrix *23*
4.1.2 Transactional and master data for fact and dimension tables *23*
4.1.3 Data Warehouse model (Snowflake or Star, or galaxy) *36*
4.2 ETL processes **36**
4.2.1 Dimension Table's ETL Process *41*
4.2.2 Fact Table's ETL Process *44*
CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION **46**
5.1 Report and dashboard systems (structure) **46**
5.2 Data analysis with Power BI **46**
5.2.1 Sales *46*
5.2.2 Purchase Department *47*
5.2.3 Warehouse Department *48*
5.3 Evaluation and Discussion **49**
5.3.1 Sales *49*
5.3.2 Purchase Department *51*
5.3.3 Warehouse Department *53*
CHAPTER 6: CONCLUSION AND FUTURE WORKS **56**
6.1 Results **56**
6.2 Limitations **56**
6.3. Future works **56**

CONTENT OF FIGURE

Figure 1: Relational model *36*
Figure 2: Load dimension for Date *37*
Figure 3: Load dimension for Customer *37*
Figure 4: Load dimension for Employee *38*
Figure 5: Load dimension for Location *38*
Figure 6: Load dimension for StockItem *39*
Figure 7: Load dimension for Supplier *39*
Figure 8: Load dimension for Transaction Type *40*
Figure 9: Load Fact Sales *40*
Figure 10: Load Fact Purchase *41*
Figure 11: Load Fact Warehouse *41*
Figure 12: Data in Date Dimension *42*
Figure 13: Data in Customer Dimension *42*
Figure 14: Data in Emloyee Dimension *42*
Figure 15: Data in Location Dimension *43*
Figure 16: Data in StockItem Dimension *43*
Figure 17: Data in Supplier Dimension *43*
Figure 18: Data in Transaction Type Dimension *44*
Figure 19: Data in Fact Sales *44*
Figure 20: Data in Fact Purchase *44*
Figure 21: Data in Fact Warehouse *45*
Figure 22: Dashboard factsales *47*
Figure 23: Dashboard Purchase *48*
Figure 24: Dashboard Warehouse *49*
Figure 25: Calculate relevant parameters *50*
Figure 26: Top 10 best selling product *50*
Figure 27: Total number of products, total list of goods and total cost of goods *51*
Figure 28: Top 6 most ordered products *52*
Figure 29: Monthly Order Trend *52*
Figure 30: Customers and suppliers frequently appear in category name list *53*
Figure 31: Customers and suppliers frequently appear in category name list *54*
Figure 32: Number of shipping days *55*

CONTENT OF TABLE

Table 1: Differences between Data Warehouse and Data Mart.....*8*
Table 2: Bus Matrix.....*23*
Table 3: Transactional and master data for Dim Date.....*24*
Table 4: Transactional and master data for Dim Customer.....*26*
Table 5: Transactional and master data for Dim Employee.....*27*
Table 6: Transactional and master data for Dim Location.....*28*
Table 7: Transactional and master data for Dim StockItem.....*29*
Table 8: Transactional and master data for Dim Supplier.....*30*
Table 9: : Transactional and master data for Dim Transaction Type.....*31*
Table 10: Transactional and master data for Fact Sales.....*31*
Table 11: Transactional and master data for Fact Purchase.....*33*
Table 12: Transactional and master data for Fact Warehouse.....*34*
Table 13: Source & Destination of project......*36*

CHAPTER 1: INTRODUCTION

1.1 Business case for the project

Our group business case is to implement Wide World Importers data warehouse for Business Intelligence. As Wide World Importers (WWI) is a global import and export company that deals with a wide range of products, in order to keep up with the growing competition in the market, WWI needs to make data-driven decisions to improve their business performance. Our group proposes implementing the Wide World Importers data warehouse to support their business intelligence needs (mmohajer9, 2021)

1.2 Objectives of the project

1.2.1 General Objective

To implement the Wide World Importers data warehouse as a centralized data source to support WWI's business intelligence needs.

1.2.2 Specific Objectives

Firstly is to centralize the data from various business processes and transactions into a single source of truth. Secondly is to provide accurate and timely information for decision-making to WWI's stakeholders. Finally is to enable better forecasting, planning, and performance monitoring for WWI's business operations.

The specific objectives are more focused and measurable, and are directly aligned with the general objective of implementing the data warehouse. These objectives provide a clear roadmap for the project and help to ensure that the project stays on track and meets its intended goals.

1.3 Research Objects

Business processes and transactions: These are the key activities and operations that are carried out by WWI, and are the primary source of data for the data warehouse. The research may focus on identifying the types of processes and transactions that need to be captured in the data warehouse, and how they can be standardized and streamlined to ensure data accuracy and consistency.

Data sources: These are the systems, applications, and other sources that generate data for WWI. The research may focus on identifying the key data sources, assessing the quality and consistency of the data, and developing strategies for integrating the data into the data warehouse.

Business intelligence tools: These are the software applications and tools that are used to analyze and visualize the data in the data warehouse. The research may focus on identifying the appropriate BI tools that can be integrated with the data warehouse, assessing their

capabilities and limitations, and developing strategies for leveraging their functionalities to support WWI's decision-making.

Performance metrics: These are the measures that are used to evaluate the effectiveness and efficiency of WWI's business operations. The research may focus on identifying the key performance metrics, assessing their relevance and validity, and developing strategies for tracking and reporting them using the data in the data warehouse.

1.4 Scope of the project

The scope of this project is to implement the Wide World Importers data warehouse and set up the necessary infrastructure to support the BI solutions. The project will involve the following steps:

- Data warehouse design and development: The database will be designed to support the required data structures and relationships, and populated with data from WWI's various business processes and transactions.
- Data integration: The data warehouse will be integrated with other data sources and systems to ensure that the data is up-to-date and consistent.
- BI solutions: BI solutions will be proposed and visualized in the data warehouse, helping the company have a better overview of the data warehouse built from the WWI dataset.
- Testing and deployment: The data warehouse and the BI solutions will be tested and deployed to ensure that they meet the requirements and perform as expected.

1.5 Value and desired outcome of the project

- Improved decision-making: The data warehouse can provide accurate and up-to-date information to support data-driven decision-making for WWI's stakeholders. This can help them identify opportunities for growth, optimize their resources, and mitigate risks.
- Increased efficiency: By centralizing their data and automating data processing and analysis, WWI can save time and effort for their employees. This can lead to faster response times, better productivity, and improved customer service.
- Better forecasting and planning: With historical and current data, WWI can use the database to forecast future trends and plan their business operations accordingly. This can help them optimize their resources, minimize risks, and stay competitive in the market.
- Enhanced customer service: By having access to customer data, WWI can personalize their services and improve customer satisfaction. This can also help them identify potential opportunities for cross-selling and upselling.
- The desired outcome of the project is to provide WWI with a powerful tool for supporting their business intelligence needs. The project should result in a functional database that is integrated with the necessary infrastructure and BI solutions.

Ultimately, the project should help WWI to make better-informed decisions, optimize their operations, and improve their overall business performance.

1.6 Structure of project

Chapter 1: Introduction

This section should provide an overview of the project, including its background, objectives, and desired outcomes. It should also include a brief description of WWI's business operations and the challenges that the project aims to address.

Chapter 2: Theoretical Basis

This section discusses the theoretical basis of BI solutions, including data warehousing, ETL processes, data modeling, database design, and data analysis and visualization techniques. These concepts and principles provide the foundation for developing a BI solution for WideWorldImporters.

Chapter 3: Requirements Analytics And Introduction To Bi Solution

In this section, business requirements for WideWorldImporters are analyzed, and key performance indicators (KPIs) and metrics for tracking business performance are identified. The BI solution options are discussed, and the appropriate tools and technologies are selected to develop the solution.

Chapter 4: Building Data Warehouse And Integrating Data

This section focuses on the design and development of the data warehouse for WideWorldImporters, as well as the extraction, transformation, and loading (ETL) of data from various source systems. The integration of data into the data warehouse and implementation of data quality and data governance processes are also discussed.

Chapter 5: Results – Data Analytics And Visualization

In this section, data analytics and visualization solutions for WideWorldImporters are developed. Dashboards and reports are created to track KPIs and metrics, and advanced analytics techniques are used to analyze business trends and patterns. Insights and recommendations are presented to business stakeholders based on the results of the data analysis.

Chapter 6: Conclusion And Future Works

The final section summarizes the project outcomes and achievements, discusses lessons learned and challenges faced during the project, and provides suggestions for future improvements and enhancements to the BI solutions. The project concludes with final thoughts and conclusions on the BI solution for WideWorldImporters data warehouse.

CHAPTER 2: THEORETICAL BASIS

2.1 Overview about BI

2.1.1 What is BI?

Business Intelligence "is the process of collection, treatment and diffusion of information that has as an objective, the reduction of uncertainty in the making of all strategic decisions" (Zeng, L., et al, 2007). This definition suggests that BI is more than just a set of tools or technologies, but rather a comprehensive process that involves several stages.

Overall, BI is a complex process that involves multiple stages and requires a range of technical and analytical skills. It is aimed at providing decision-makers with timely and accurate information that can be used to reduce uncertainty, identify opportunities, and optimize business performance.

2.1.2 BI Architecture

Data sources: This involves identifying where the data that will be used in the BI process comes from. Data sources could include databases, spreadsheets, cloud storage, and even social media platforms.

Data integration: This step involves integrating data from multiple sources into a single location or data warehouse. This could involve ETL (extract, transform, load) processes, or ELT (extract, load, transform) processes, depending on the specific architecture.

Data modeling: This involves defining how the data in the data warehouse is structured and organized. This is typically done using a data model, which is a graphical representation of the data and its relationships.

BI tools: This involves selecting and configuring the right software tools for data analysis, visualization, and reporting. This could include tools like Power BI, Tableau, QlikView, or other similar tools.

Data analysis: This step involves analyzing the data using the BI tools to uncover insights and trends that can support business decision-making.

Visualization and reporting: This involves presenting the data analysis in a visual and understandable format using charts, graphs, and dashboards. Reports are created to communicate the insights to stakeholders, enabling them to make informed decisions based on the data.

2.1.3 Advantage of BI in enterprises

Business Intelligence (BI) offers numerous advantages to organizations that adopt it. With BI, companies can reduce guesswork and enhance communication and coordination across departments, leading to faster responses to changes in financial conditions, customer

preferences, and supply chain operations. By leveraging timely and accurate information, companies can enhance their overall performance, as information is often considered the second most critical resource after employees. Speedy and well-informed decision-making is crucial in today's competitive business landscape, where prompt and accurate information can help achieve superior performance. BI also enables companies to improve their customer experience by responding to customer issues and priorities promptly and appropriately.

- Better decision-making: BI provides businesses with the ability to collect and analyze large volumes of data from multiple sources. This data can be transformed into insights that can be used to make informed and data-driven decisions.
- Improved operational efficiency: By analyzing operational data, BI can help identify inefficiencies and bottlenecks in processes, enabling organizations to optimize their operations for greater efficiency.
- Increased revenue: BI can help identify sales trends and customer behavior, enabling businesses to make adjustments to their strategies and offerings to maximize revenue potential.
- Better customer insights: By analyzing customer data, BI can help businesses gain insights into customer behavior, preferences, and needs. This information can be used to create targeted marketing campaigns and improve customer service.
- Competitive advantage: BI can give businesses a competitive edge by allowing them to make faster and more accurate data-driven decisions than their competitors. This is because BI can extract insights and trends from large amounts of data, helping businesses identify opportunities, optimize processes, and make informed strategic decisions that keep them ahead of the competition.
- Predictive analytics: BI can enable businesses to use historical data to make predictions about future trends and outcomes. This can help businesses make proactive decisions and identify opportunities before they become apparent.

2.1.4 BI Strategy for Business

Developing a BI strategy for business is a multi-step process that can significantly improve an organization's decision-making capabilities and overall performance. The first step involves defining clear objectives that align with the organization's overall strategy and goals. These objectives serve as a guide throughout the development process and ensure that the BI strategy meets the needs of the organization.

The second step developing a business intelligence (BI) strategy involves identifying the data sources that will be utilized to inform decision-making. This includes both internal data, such as sales and customer data, as well as external data, such as market trends and economic indicators. After identifying the necessary data sources, the next step is to establish a data warehouse (DW) to store and manage the data. The DW serves as a central repository that enables the integration of data from various sources into a structured format, making it easier for organizations to perform complex queries and analysis on the data.

The third step is to perform data modeling, which involves identifying the relationships between data sets and developing a structure that allows for efficient and effective analysis. This step is critical for ensuring that the data is accurately and appropriately analyzed.

The fourth step is to perform ETL/ELT (extract, transform, load) processes to move and prepare data for analysis. This step may involve using cloud-based tools to facilitate data movement and analysis.

The fifth step is to visualize and report on the data to support decision-making. This may include creating dashboards, reports, and other visualizations that allow stakeholders to easily understand and interpret the data.

A BI strategy that is properly developed can enhance inter-departmental communication, enable businesses to quickly respond to changes in financial conditions, supply chain operations, and customer preferences. Additionally, it can expedite the decision-making process by utilizing timely and accurate information. By leveraging these advantages, companies can ultimately improve their overall performance and gain a competitive advantage over their rivals.

2.2 ETL Process

2.2.1 What is ETL?

ETL refers to the process of Extracting, Transforming, and Loading data from various sources into a data warehouse or a data analytics platform. ETL is an essential component that enables visual analytics to function. Without ETL, the data would be restricted to the layout and format of its original source. As a user progresses in visual analytics, they may need to merge different data sources, combine fields within the sources, or even break down individual fields into multiple ones. Additionally, users may need to replace codes and abbreviations with their actual values, and they would want to analyze the structure of the data to ensure its accuracy and reliability (Tim Costello & Lori Blackshear, 2020).

2.2.2 Why do we need ETL?

ETL is crucial because it extracts data from multiple sources, transforms it into an analyzable format, and loads it into a data warehouse or data mart. This integration process facilitates better querying and analysis of data, which helps organizations make informed, data-driven decisions. This process enables organizations to combine data from disparate sources and gain valuable insights into their business operations.

ETL allows organizations to clean, standardize, and consolidate data from different sources, making it easier to analyze and report on. It also helps to ensure data quality, which is essential for making informed business decisions. ETL workflows can be automated, saving time and reducing the risk of errors that can occur with manual data handling. Without ETL, organizations would struggle to extract meaningful insights from their data, leading to inaccurate or incomplete decision-making.

Example, when a company wants to analyze data from various systems such as sales systems, customer management systems, and inventory management systems. These systems may store data in different formats and structures, making it difficult to combine and analyze the data without first transforming it through ETL processes. By using ETL tools, the company can extract the relevant data from each system, transform it into a standardized format, and load it into a data warehouse or data lake. This enables the company to perform comprehensive analyses across different business areas and make data-driven decisions to improve their operations and strategy. Without ETL, the data may be incomplete or inconsistent, leading to inaccurate analyses and potentially costly business decisions.

2.2.3 ETL Process

The ETL (Extract, Transform, Load) process is a crucial component of data warehousing and business intelligence. It involves extracting data from multiple sources, such as databases, flat files, and web services, transforming it into a consistent format, and loading it into a data warehouse or data mart for analysis.

During the extraction phase, raw data is copied or exported from various source locations to a staging area. In the transform phase, the data is processed and transformed into a consolidated format for its intended use case. This includes filtering, cleaning, validating, and authenticating the data, formatting it into tables, and performing calculations or summarizations. The load phase involves moving the transformed data from the staging area to a target data warehouse or data mart.

The ETL process enables organizations to integrate data from multiple sources, transform it into a consistent format, and load it into a centralized repository for analysis. By doing so, businesses can obtain valuable insights and make informed, data-driven decisions.

2.3 Data warehouse and Data mart

2.3.1 What are Data warehouse and Data mart?

A data warehouse is a large, centralized repository of integrated data that supports business decision-making activities by enabling analysis and reporting across multiple sources of data. It is typically used to store historical data, and it is designed to support complex queries, data analysis, and data mining. A data warehouse is optimized for read-heavy workloads, and it provides a single source of truth for the organization's data.

On the other hand, a data mart is a subset of a data warehouse that is focused on a particular business function or department. It is a smaller, more focused version of a data warehouse that contains data relevant to a specific area of the business. Data marts are often created to support specific business needs, such as sales analysis, marketing analysis, or customer analysis. They are designed to provide fast, easy access to data for a particular group of users.

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

2.3.2 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

2.3.3 Advantages and disadvantages of Data warehouse

- Advantages

- Save funds 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.

2.3.4 Snowflake and Star schemas

Snowflake schema (Figure xx) 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.

Star schema is a type of relational schema used to represent a multidimensional data model in a data warehouse. It is called a "star schema" because its entity-relationship diagram looks like a star with points extending from a center table. The central table of the schema is the fact table, while the tables that connect to it are the dimension tables. This schema is explicitly designed for data warehousing, and it is optimized for fast querying and reporting. It is a popular choice for business intelligence applications and decision support systems.

2.4 KPIs

2.4.1 KPIs Definition

Key Performance Indicators are measurable values that are used to evaluate the success of an organization or a particular activity. In Business Intelligence, KPIs are metrics that are used to measure the performance of a company against its goals. KPIs are used to help businesses make informed decisions based on real-time data, and they are used to track progress and identify areas where improvements can be made. KPIs can be used to measure a variety of factors, such as financial performance, customer satisfaction, employee performance, and more. By identifying and tracking KPIs, businesses can gain insights into their performance and make data-driven decisions to improve their operations. KPIs can also be used to set targets and goals for employees and departments, and to monitor progress towards those targets over

time. In any case, a key performance indicator has a goal value, evaluative intervals, or both, and is used to measure the improvement or deterioration of business results. KPIs are usually defined for a specific activity and are strictly goal oriented. It is important to make sure that all key performance indicators can be measured and quantified. If defined and calculated correctly, KPIs help organizations to understand how well or poorly they perform and how they actually are able to fulfill the set goals. Therefore, the right set of KPIs should not only shed light on the results of an organization, but also help to identify areas where additional attention is needed (Vigintas Sakys et al, 2013). At the end of the day, "what gets measured gets done" and "without the right KPIs managers are sailing blind" (Marr, 2012). One of the goals of good managers and decision makers is to understand the main efficiency measurements, and therefore it is these people who should point out the key performance indicators.

2.4.2 The advantages and disadvantages of KPIs

Advantages of KPIs:

- Improved focus and alignment: KPIs help organizations to focus on what is important and align their activities with the overall goals and objectives. This ensures that all efforts are directed towards achieving the desired outcomes.
- Better decision making: With KPIs, organizations have access to real-time data and insights, enabling them to make informed decisions based on accurate information.
- Increased accountability: KPIs provide a clear way to measure and track performance, which creates a sense of accountability among employees and helps to drive performance improvement.
- Enhanced communication: KPIs facilitate communication between different levels of the organization, enabling everyone to work towards the same objectives and understand how their work contributes to the overall success of the organization.

Disadvantages of KPIs:

- Over-reliance on numbers: KPIs are often quantitative measures that can lead to an over-reliance on numbers and metrics, which can result in overlooking other important factors such as employee satisfaction or customer feedback.
- Difficulty in defining and selecting relevant KPIs: Choosing the right KPIs can be challenging, as they must be relevant to the organization's objectives and provide meaningful insights.
- Time-consuming: Implementing and monitoring KPIs can be time-consuming, requiring significant resources and efforts.
- Inappropriate incentives: In some cases, KPIs can create inappropriate incentives that encourage employees to prioritize achieving targets over quality or ethical considerations.

2.4.3 Categories of KPIs (Sales, Warehouse and Purchasing)

- Sales KPIs:
- Revenue growth rate: This KPI measures the percentage increase in revenue over a period of time. It is an important KPI to track as it shows how successful the Sales department has been in generating revenue for the company.

- Sales growth rate: Similar to the revenue growth rate, this KPI measures the percentage increase in sales over a period of time. It is a good indicator of how well the Sales department is performing in terms of generating new business.
- Sales conversion rate: This KPI measures the percentage of leads that are converted into actual sales. It is important for the Sales department to keep this KPI high as it shows how effective they are in turning leads into paying customers.
- Average order value: This KPI measures the average value of each sale. It can be used to identify trends in customer behavior and to help the Sales department identify areas where they can improve sales
- Customer lifetime value: This KPI measures the total value of a customer to the company over the course of their relationship. It is an important KPI to track as it shows how profitable each customer is over time.
- Customer acquisition cost: This KPI measures the cost of acquiring a new customer. It is important for the Sales department to keep this KPI as low as possible to maximize profitability.
- Repeat customer rate: This KPI measures the percentage of customers who make repeat purchases. It is important for the Sales department to keep this KPI high as it shows how successful they are in retaining customers.
- Sales per employee: This KPI measures the amount of revenue generated per employee. It is an important KPI to track as it shows how productive the Sales department is overall.
- Sales by region: This KPI measures the amount of revenue generated by the Sales department in each geographic region. It can be used to identify areas where the Sales department is performing well and areas where they may need to focus more attention.
- Sales by product category: This KPI measures the amount of revenue generated by the Sales department in each product category. It can be used to identify trends in customer behavior and to help the Sales department identify areas where they can improve sales.

● Warehouse KPIs:

- Inventory accuracy rate: This KPI measures the accuracy of inventory data by comparing physical inventory counts to the recorded amounts in the system. A high accuracy rate indicates effective inventory management and minimizes the risk of stockouts or overstocking.
- Order fulfillment rate: This KPI measures the percentage of customer orders that are fulfilled on time and in full. A high fulfillment rate indicates efficient order processing and shipping operations.
- Stockout rate: This KPI measures the percentage of time that inventory is not available when it is needed. A high stockout rate indicates poor inventory management and can lead to lost sales and dissatisfied customers.
- Time to process orders: This KPI measures the time it takes to process an order from the time it is received to the time it is shipped. A short processing time indicates efficient order fulfillment processes.
- Time to pick and pack orders: This KPI measures the time it takes to pick and pack items for an order. A short pick and pack time indicates efficient warehouse operations and reduces order processing time.

- On-time delivery rate: This KPI measures the percentage of orders that are delivered to customers on time. A high on-time delivery rate indicates effective logistics and transportation management.
- Order accuracy rate: This KPI measures the percentage of orders that are fulfilled accurately, without errors or missing items. A high accuracy rate indicates effective order processing and reduces the need for returns and exchanges.
- Warehouse capacity utilization rate: This KPI measures the percentage of warehouse capacity that is being used. A high utilization rate indicates effective warehouse space management and can lead to cost savings.

● Purchasing

- Cost savings: This KPI measures the amount of money saved through effective purchasing strategies such as negotiating better prices, finding alternate suppliers, or reducing wastage.
- Supplier performance: This KPI measures the performance of suppliers based on factors such as on-time delivery, quality of goods, and adherence to contract terms.
- Purchase order cycle time: This KPI measures the time taken to process purchase orders from initiation to receipt of goods. A shorter cycle time indicates a more efficient purchasing process.
- Inventory turnover: This KPI measures the number of times inventory is sold and replenished within a given time period. A higher turnover rate indicates that inventory is being managed efficiently and is not being overstocked.
- Purchasing lead time: This KPI measures the time it takes to procure goods from the time of order placement to delivery. A shorter lead time indicates a more efficient purchasing process.
- Cost per purchase order: This KPI measures the total cost incurred in processing a single purchase order. A lower cost indicates that the purchasing process is more efficient and cost-effective.
- Compliance: This KPI measures the extent to which purchasing activities adhere to legal and ethical requirements, such as environmental regulations or supplier diversity policies.
- Contract compliance: This KPI measures the extent to which purchasing activities adhere to the terms of contracts, such as pricing, delivery, and quality requirements.

2.5 Microsoft Azure

2.5.1 What is Microsoft Azure?

Microsoft Azure, which was formerly referred to as Windows Azure, is a cloud computing platform provided by Microsoft for public use. Its services range from analytics, storage, compute to networking, and users have the flexibility to choose which of these services they need to develop and scale either new or existing applications in the public cloud.

Azure is a cloud platform designed to help organizations address their challenges and meet their business objectives. It provides a range of tools suitable for various industries, including finance, e-commerce, and Fortune 500 companies. The platform also supports open-

source technologies, allowing users to choose the tools and technologies that best suit their needs (Stephen-J-Bigelow, 2022). Azure offers four cloud computing services, including infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and serverless functions. These services provide businesses with flexibility and scalability in managing their operations.

2.5.2 Advantages and disadvantages

- Advantages of Microsoft Azure:
- Scalability: Azure provides an easy way to scale up or down based on business needs, allowing users to pay only for what they use.
- Flexibility: Azure supports a wide range of operating systems, programming languages, and tools, making it easy for users to integrate with existing systems and applications.
- Security: Azure provides a secure platform with features such as threat detection, identity management, and access controls to safeguard data and applications.
- Hybrid Capabilities: Azure offers the ability to integrate with on-premises infrastructure and other public cloud providers.
- Cost-effective: Azure's pay-as-you-go model and flexibility in choosing services allows users to manage their costs efficiently.
- Disadvantages of Microsoft Azure:
- Complexity: Azure can be complex to navigate and requires expertise to set up and manage effectively.
- Cost: While Azure can be cost-effective for small businesses, it can become expensive for larger organizations or those with complex needs.
- Downtime: Like any other cloud service, Azure can experience downtime, which can impact business operations and productivity.
- Learning Curve: It can take time for users to learn how to use Azure effectively, and the process can be challenging for those without a strong technical background.
- Data Transfer Costs: Azure charges data transfer fees for data transferred between different regions or between Azure and other cloud providers.

CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION

3.1 Business processes

3.1.1 Business department

- Purchase Department

The Purchase department is in charge of tracking down and buying raw materials and finished goods for the firm. This department has to ensure that it obtains items from reputable vendors at the most competitive cost. The department must also maintain the optimum quantity of inventory and minimize stockouts, which may have an influence on the sales department's success. To guarantee that its aims are met, the Purchase Department must constantly review its performance and adjust its sourcing tactics.

- Warehouse Department

The warehouse department is in the role of inventory management and ensuring that items are kept and dispatched effectively. This department must verify that inventory levels are accurately recorded and that items are stored in proper conditions to avoid harm. In addition, the department must guarantee that the items are delivered to clients in a timely and efficient way. To guarantee that its aims are met, the warehouse department must regularly review its performance and change its methods.

- Sales Department

The sales department is in charge of generating more revenue by selling imported goods to clients. This department collaborates closely with marketing to produce leads, which are ultimately converted into sales by customers. The sales department must also guarantee that the items are delivered on time and that the consumers' expectations are satisfied. To fulfill its goals, the department must constantly assess its performance and adapt its strategy.

3.1.2 The purpose of each department

- Purchase Department

In the case of Wide World Importers, the Purchase Department is responsible for procuring the inventory items needed for the company's retail and wholesale operations. The department is responsible for identifying the products and materials needed, creating purchase requisitions, identifying potential suppliers, negotiating contracts, and placing purchase orders. The Purchase Department at Wide World Importers also plays a key role in maintaining relationships with suppliers, monitoring order status, and approving invoices and payments.

One of the challenges faced by the Purchase Department at Wide World Importers is managing the complexities of the global supply chain. The company sources products from a wide range of suppliers located in different parts of the world. This requires the Purchase

Department to stay up-to-date with the latest trends and developments in global trade, and to develop relationships with reliable suppliers that can deliver quality products at competitive prices.

In addition to these responsibilities, the Purchase Department at Wide World Importers also plays a key role in implementing and enforcing ethical sourcing policies. The company is committed to ensuring that its suppliers adhere to strict ethical standards, and the Purchase Department is responsible for conducting regular audits to ensure that suppliers are in compliance with these standards. This helps to ensure that the products and materials sourced by the company are not only of high quality, but also meet the company's values and standards.

- Warehouse Department

The Warehouse Department at Wide World Importers is responsible for receiving incoming products from suppliers, storing them in the warehouse, and preparing them for distribution to customers. This process involves a range of tasks, including unpacking, inspecting, and cataloging products, as well as storing them in the appropriate locations within the warehouse. The department also works closely with the Purchase Department to ensure that incoming products are accurately recorded and tracked in the company's inventory management system.

One of the biggest challenges faced by the Warehouse Department at Wide World Importers is managing inventory levels. The department must carefully balance the need to have sufficient inventory on hand to meet customer demand while also avoiding overstocking, which can lead to inefficiencies and increased costs.

The Warehouse Department also plays a key role in managing the company's logistics operations. This includes coordinating with transportation providers to ensure that products are delivered to customers on time and in good condition. The department is responsible for managing the shipping process, including selecting carriers, scheduling deliveries, and tracking shipments to ensure that they arrive at their destination as planned.

Another important function of the Warehouse Department at Wide World Importers is managing the physical layout of the warehouse. This includes optimizing the use of available space, ensuring that products are stored in the correct locations, and implementing best practices for warehouse safety and security.

- Sale Department

The Sales Department at Wide World Importers employs a range of sales strategies to drive revenue growth, including building relationships with existing customers, acquiring new customers, and expanding the company's product offerings. The department is responsible for identifying potential customers, developing sales proposals, negotiating contracts, and closing deals.

One of the key challenges faced by the Sales Department at Wide World Importers is the competitive nature of the retail and wholesale markets. The company operates in a highly competitive environment where it must constantly innovate and adapt to stay ahead of the competition. The Sales Department is responsible for understanding market trends, analyzing customer needs, and developing sales strategies that meet the changing demands of the market.

To manage these challenges, the Sales Department at Wide World Importers uses RFM for Customer Segmentation software that helps to streamline the sales process, automate routine tasks, and track customer interactions. This also helps the department to identify potential sales opportunities, prioritize leads, and track the performance of sales representatives.

In addition to these responsibilities, the Sales Department at Wide World Importers also plays a key role in developing and implementing promotional campaigns. These campaigns may include advertising, sales promotions, or other marketing activities designed to increase brand awareness and drive sales revenue. The Sales Department works closely with the Marketing Department to develop and execute these campaigns, ensuring that they are aligned with the company's overall business strategy.

3.1.3 Business process

- Purchase Department
    - Identify purchase requirements: The Purchase Department identifies the company's requirements for products and materials, based on factors such as sales forecasts, production plans, inventory levels and place an order at the request of a customer when there is a customer order but there is not enough in stock.
    - Create purchase requisitions: The Purchase Department creates purchase requisitions that detail the products or materials needed, their quantities, and any other relevant specifications or requirements.
    - Identify potential suppliers: The Purchase Department identifies potential suppliers for the products or materials needed, based on factors such as price, quality, reliability, and delivery time.
    - Send requests for quotations (RFQs): The Purchase Department sends RFQs to the selected suppliers, requesting quotes for the products or materials needed.
    - Evaluate supplier proposals: The Purchase Department evaluates the proposals received from the suppliers, taking into account factors such as price, quality, delivery time, and terms and conditions.
    - Negotiate contracts: The Purchase Department negotiates contracts with the selected suppliers, including terms and conditions such as price, delivery time, payment terms, and warranties.
    - Place purchase orders: The Purchase Department places purchase orders with the selected suppliers, specifying the products or materials ordered, their quantities, and the delivery dates.

- Approve invoices and payments: The Purchase Department approves the invoices received from the suppliers and authorizes payment for the products or materials received.
- Manage supplier relationships: The Purchase Department manages the relationships with the suppliers, maintaining regular communication to ensure that the company's requirements are being met and addressing any issues or concerns that arise.
- Conduct supplier evaluations: The Purchase Department conducts periodic evaluations of the suppliers, assessing their performance based on factors such as quality, delivery time, and responsiveness.
- Maintain purchasing records: The Purchase Department maintains records of all purchasing activities, including purchase requisitions, RFQs, purchase orders, invoices, and supplier evaluations.
- Develop and implement purchasing policies and procedures: The Purchase Department develops and implements policies and procedures related to purchasing activities, ensuring that they are in compliance with relevant laws and regulations and consistent with the company's goals and values.

● Warehouse Department
  - Receive purchase orders: The Warehouse Department receives purchase orders from the Purchase Department that detail the products, quantities, and expected delivery dates.
  - Check stock availability: The Warehouse Department checks the inventory records to ensure that the ordered products are in stock and available for delivery.
  - Prepare for incoming shipments: The Warehouse Department prepares the warehouse for the incoming shipments by ensuring that there is sufficient space and resources for storing and handling the products.
  - Receive and inspect shipments: When the shipments arrive, the Warehouse Department receives and inspects them to ensure that the products are in good condition and match the specifications in the purchase order.
  - Record incoming products: The Warehouse Department records the incoming products in the inventory system, including their location in the warehouse.
  - Store and organize products: The Warehouse Department stores and organizes the products in the appropriate locations in the warehouse, taking into account factors such as their size, weight, and storage requirements (e.g. chiller sections for perishable items).
  - Pick products for orders: When orders are received from the Sales Department, the Warehouse Department picks the products from their locations in the warehouse and prepares them for shipment
  - Pack products for shipment: The Warehouse Department packs the products for shipment, taking care to use appropriate packaging materials to ensure that the products are protected during transportation.

- Arrange for shipment: The Warehouse Department arranges for the shipment of the products, either through the company's own fleet or through third-party logistics providers.
- Update inventory records: The Warehouse Department updates the inventory records to reflect the changes in stock levels and locations.
- Monitor inventory levels: The Warehouse Department monitors the inventory levels to ensure that there is sufficient stock to meet demand and to avoid stockouts.
- Conduct inventory counts: The Warehouse Department conducts periodic inventory counts to ensure that the inventory records are accurate and to identify any discrepancies that need to be resolved.
- Maintain warehouse cleanliness and safety: The Warehouse Department maintains the cleanliness and safety of the warehouse, including conducting regular inspections and addressing any hazards or issues that arise.
- Provide support to other departments: The Warehouse Department provides support to other departments, such as the Sales and Customer Service Departments, by providing information about stock availability and shipment status.

● Sales Department
  - Receive customer inquiries: The Sales Department receives inquiries from customers, either directly or through the company's website or sales channels.
  - Qualify leads: The Sales Department qualifies the leads based on factors such as the customer's budget, timeline, and product requirements.
  - Provide product information: The Sales Department provides customers with information about the company's products, features, and pricing, as well as any promotions or discounts available.
  - Create quotes: Based on the customer's requirements, the Sales Department creates quotes that detail the products, quantities, and pricing.
  - Negotiate terms: Sales negotiate with customers on terms of sale, including price, delivery, and payment options (for reseller customers).
  - Finalize orders: Once the terms are agreed upon, the Sales Department finalizes the orders and creates sales orders that detail the products, quantities, pricing, and delivery dates.
  - Verify stock availability: The Sales Department checks the inventory records to ensure that the ordered products are in stock and available for delivery.
  - Confirm orders with customers: Sales department confirms orders with customers including detailed information about products, delivery dates, prices and products to be delivered later if there is not enough stock when customers order.
  - Prepare and ship products: The Sales Department moves orders to the Warehouse Department to pack and arrange for delivery to customers.
  - Update sales records: The Sales Department updates the sales records to reflect the new orders and associated revenues.

- Follow up with customers: The Sales Department follows up with customers after the sale to ensure their satisfaction and to solicit feedback on the products and services.
- Provide customer support: The Sales Department provides ongoing support to customers, including handling complaints, providing technical assistance, and processing returns or exchanges.

When implementing a BI solution, each department needs to needs to:

- Define the business requirements for the BI solution, including the types of reports, dashboards, and analytics that are needed to support sales operations.
- Identify and gather relevant data from various sources, such as customer data, sales data, and marketing data, to populate the BI solution.
- Develop a data model that represents the relationships between the data sources and supports the desired BI outputs.
- Transform and load the data into the data model using ETL (extract, transform, load) processes and tools.
- Develop the reports, dashboards, and analytics that support the sales team's needs, using BI tools such as Microsoft Power BI or Tableau.
- Provide training and support to the sales team to ensure that they can effectively use the BI solution to inform their decision-making processes.

3.2 Business Requirements Analysis

The first step in building a BI solution for the Sales and Warehouse departments at Wide World Importers (WWI) is to conduct a business requirements analysis for each department.

3.2.1 Sales Department

What are the revenues and expenses for each period?

Who are the customers with the highest spending and employees with the best sales?

What are the department's revenue targets, and how are these targets currently being tracked?

What are the department's key performance indicators (KPIs), such as number of leads generated, conversion rates, sales growth and sales revenue?

What are the department's current sales processes, and are there any bottlenecks or inefficiencies that could be addressed through data analysis?

What customer data is currently being collected, and how is it being used to inform sales strategies?

For Sales department, the important KPIs are (mentioned in section 2.4.3):

- Sales Revenue
- Sales per employee
- Sales by region
- Sales by product category

3.2.2 Warehouse Department

Based on the number of goods in the warehouse, is it possible to determine whether the warehouse has just imported or exported goods?

Is the shipping of the warehouse effective? Based on what point? If not, suggestions for improvement?

Is it possible to identify suppliers and loyal customers from the import and export of warehouses from those units?

Are the delivery and receipt documents up to date?

For Warehouse department, the important KPIs are (mentioned in section 2.4.3):

- Order accuracy rate
- On-time delivery rate
- Time to pick and pack orders
- Time to process orders
- Stockout rate
- Order fulfillment rate

3.2.3 Purchase Department

What is the trend of importing goods in each period (month/quarter/year)?

Who are the best suppliers?

What are the department's procurement goals, and how are these goals currently being tracked?

What are the department's key performance indicators (KPIs), such as cost savings, on-time delivery rates, and supplier performance?

What are the department's current procurement processes, and are there any bottlenecks or inefficiencies that could be addressed through data analysis?

What supplier data is currently being collected, and how is it being used to inform procurement strategies?

For Purchase department, the important KPIs are (mentioned in section 2.4.3):

- Purchasing lead time
- Supplier performance
- Cost savings

3.3 Data source and challenges

The data source for a BI solution for the Sale and Warehouse departments using the Wide World Importers database would be the various tables within the database that contain information about customers, orders, products, and inventory.

The main challenge in building a BI solution for the Sale and Warehouse departments would be to effectively analyze and present the data in a meaningful way. This would require creating a data model that accurately reflects the relationships between the different tables, and building out a set of reports and dashboards that provide insights into the key metrics and trends for the business.

Some of the specific challenges that may arise in building a BI solution for these departments could include:

- Data quality issues: Ensure the quality and completeness of the data might be difficult, especially if the source data is generated from numerous systems or sources.
- Complex data relationships: The Wide World Importers database, as previously stated, comprises a vast number of tables with intricate interactions between them. It might be difficult to construct queries and reports that effectively exploit a data model that appropriately reflects these connections.
- Performance optimization: Maintaining excellent query and report performance levels might become increasingly challenging as the number of data rises. Tuning the data model and optimizing queries can help with this problem.
- Cost management: Managing the cost of using Azure services for data storage, processing, and analysis is important to ensure that the BI solution is cost-effective.

3.4 IT requirements Analysis (IT & Infrastructure)

IT Requirements Analysis for Sale Department:

- Sales Dashboard: The sales department needs a dashboard that provides real-time sales data, including sales volume, revenue, and customer insights. The dashboard should be easy to use, customizable, and accessible from any device.
- RFM analysis for Customer Segmentation (RFM): The sales department requires a customer segmentation that can manage customer data, track customer interactions, and provide insights for targeted marketing campaigns.
- Sales Order Management System: The sales department requires an order management system that can track sales orders, inventory levels, and delivery schedules. The system should be integrated with the warehouse and logistics systems to ensure timely delivery and accurate inventory management.

IT Requirements Analysis for Purchase Department:

- Purchase Order Management System: The purchase department requires a purchase order management system that can manage purchase orders, vendor information, and inventory levels. The system should be integrated with the accounting system to facilitate payment processing.
- Vendor Management System: The purchase department requires a vendor management system that can manage vendor data, track vendor performance, and facilitate vendor communication.
- Purchase Request System: The purchase department requires a purchase request system that can manage purchase requests, approvals, and budget tracking. The system should be integrated with the financial and accounting systems to ensure accurate budget tracking.

IT Requirements Analysis for Warehouse Department:

- Warehouse Management System: The warehouse department requires a warehouse management system that can manage inventory levels, track inventory movements, and facilitate picking and packing operations.
- Logistics Management System: The warehouse department requires a logistics management system that can manage shipping schedules, track shipments, and provide real-time delivery updates. The system should be integrated with the sales and purchase order systems to ensure accurate order fulfillment.

Overall IT Requirements Analysis:

- Data Integration: The various IT systems across different departments should be integrated to ensure smooth data flow and facilitate accurate reporting and analysis.
- Data Security: All IT systems should have robust security measures in place to protect sensitive data, prevent unauthorized access, and ensure compliance with regulations and standards.
- Scalability: All IT systems should be scalable to accommodate future growth and changing business needs.
- Cloud-based Infrastructure: The IT infrastructure should be cloud-based to facilitate remote access, reduce maintenance costs, and improve scalability and reliability.

CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

4.1 Designing Data Warehouse

4.1.1 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.

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

Table 2: Bus Matrix

	Common Dimension						
Business process	Date	Customer	Employee	Location	StockItem	Supplier	Transaction Type
Sales							
Sales Invoice Management	**x**	**x**	**x**	**x**	**x**		
Purchase							
Manage orders from suppliers	**x**				**x**	**x**	
Warehouse							
Movements of stock items	**x**	**x**			**x**	**x**	**x**

4.1.2 Transactional and master data for fact and dimension tables

Dim Date

Table 3: Transactional and master data for Dim Date

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
DateKey	A unique identifier for each date in the dataset.	date			
Date	The actual date in the format yyyy-mm-dd.	datetime	x	x	
FullDate	A combination of the date, month, and year in the format yyyy-mm-dd.	char(10)	x	x	
DayOfMonth	The day of the month, represented as a numeric value.	varchar(2)	x	x	
DayName	The name of the day of the week (e.g. Monday, Tuesday), represented as a text value.	varchar(9)	x	x	
DayOfWeek	The day of the week, represented as a numeric value (e.g. 1 for Monday, 2 for Tuesday).	char(1)	x	x	
DayOfWeekInMonth	The position of the day in the week within the month, represented as a numeric value (e.g. the first Monday in a month would have a value of 1).	varchar(2)	x	x	
DayOfWeekInYear	The position of the day in the week within the year, represented as a numeric value (e.g. the first Monday of the year would have a value of 1).	varchar(2)	x	x	

DayOfQuarter	The day of the quarter, represented as a numeric value.	varchar(3)	x	x	
DayOfYear	The day of the year, represented as a numeric value.	varchar(3)	x	x	
WeekOfMonth	The week of the month, represented as a numeric value.	varchar(1)	x	x	
WeekOfQuarter	The week of the quarter, represented as a numeric value.	varchar(2)	x	x	
WeekOfYear	The week of the year, represented as a numeric value.	varchar(2)	x	x	
Month	The month of the year, represented as a numeric value (e.g. 1 for January, 2 for February).	varchar(2)	x	x	
MonthName	The name of the month (e.g. January, February), represented as a text value.	varchar(9)	x	x	
MonthOfQuarter	The month of the quarter, represented as a numeric value.	varchar(2)	x	x	
Quarter	The quarter of the year, represented as a numeric value (e.g. 1 for Q1, 2 for Q2).	char(1)	x	x	
QuarterName	The name of the quarter (e.g. Q1, Q2), represented as a text value.	varchar(9)	x	x	
Year	The year, represented as a numeric value.	char(4)	x	x	
YearName	The year, represented as a text value (e.g. "2023").	char(7)	x	x	

MonthYear	A combination of the month and year in the format yyyy-mm.	char(10)	x	x	
MMYYYY	A combination of the month and year in the format mmyyyy.	char(6)	x	x	
FirstDayOfMonth	The first day of the month, represented as a date in the format yyyy-mm-dd.	date	x	x	
LastDayOfMonth	The last day of the month, represented as a date in the format yyyy-mm-dd.	date	x	x	
FirstDayOfQuarter	The first day of the quarter, represented as a date in the format yyyy-mm-dd.	date	x	x	
LastDayOfQuarter	The last day of the quarter, represented as a date in the format yyyy-mm-dd.	date	x	x	
FirstDayOfYear	The first day of the year, represented as a date in the format yyyy-mm-dd.	date	x	x	
LastDayOfYear	The last day of the year, represented as a date in the format yyyy-mm-dd.	date	x	x	

Dim Customer

Table 4: Transactional and master data for Dim Customer

Field Name	**Descriptive**	**Type**	**Null**	**Master Data**	**Transaction Data**
CustomerKey	DW key for the customer dimension	int			
CustomerId	Numeric ID used for reference to a	int	x	x	

	customer within the WWI database				
CustomerName	Customer's full name (usually a trading name)	nvarchar(100)	x	x	
CategoryName	Customer's category	nvarchar(50)	x	x	
AccountOpened Date	Date of Customer's account opened	date	x	x	
PrimaryContact PersonID	Primary contact	int	x	x	
PotalCode	Delivery postal code for the customer	nvarchar(10)	x	x	
PhoneNumber	Phone number of customer	nvarchar(20)	x	x	

Dim Employee

Table 5: Transactional and master data for Dim Employee

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
EmployeeKey	DW key for the employee dimension	int			
EmployeeId	Numeric ID (PersonID) in the WWI database	int	x	x	
FullName	Full name for this person	nvarchar(50)	x	x	
PreferredName	Name that this person prefers to be called	nvarchar(50)	x	x	

IsSalesperson	Is this person a staff salesperson?	bit	x	x	
PhoneNumber	Phone number of employee	nvarchar(20)	x	x	
FaxNumber	Fax number of employee	nvarchar(20)	x	x	
EmailAddress	Email of employee	nvarchar(256)	x	x	

Dim Location

Table 6: Transactional and master data for Dim Location

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
LocationKey	A unique identifier for each location in the dataset.	int			
CityID	A unique identifier for each city in the dataset.	int	x	x	
CityName	The name of the city where the location is located.	nvarchar(50)	x	x	
StateProvinceName	The name of the state or province where the location is located.	nvarchar(50)	x	x	
SalesTerritory	The name of the sales territory where the location is located.	nvarchar(50)	x	x	

Dim StockItem

Table 7: Transactional and master data for Dim StockItem

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
StockItemKey	Color (optional) for this stock item	int			
StockItemID	Numeric ID used for reference to a stock item within the WWI database	int	x	x	
StockItemName	Full name of a stock item (but not a full description)	nvarchar(100)	x	x	
UnitPrice	Selling price (ex-tax) for one unit of this product	decimal(18, 2)	x	x	
Brand	Brand for the stock item (if the item is branded)	nvarchar(50)	x	x	
Size	Size of this item (eg: 100mm)	nvarchar(20)	x	x	
IsChillerStock	Does this stock item need to be in a chiller?	bit	x	x	
LeadTimeDays	Number of days typically taken from order to receipt of this stock item	int	x	xx	
ColorName	Color (optional) for this stock item	nvarchar(20)	x	x	

QuantityPerOuter	Quantity of the stock item in an outer package	int	x	x	
Barcode	Barcode for this stock item	nvarchar(50)	x	xx	

Dim Supplier

Table 8: Transactional and master data for Dim Supplier

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
SupplierKey	DW key for the supplier dimension	int			
SupplierId	Numeric ID used for reference to a supplier within the WWI database	int	x	x	
SupplierName	Supplier's full name (usually a trading name)	nvarchar(100)	x	x	
SupplierCategory Name	Supplier's category	nvarchar(100)	x	x	
PhoneNumber	Phone number of supplier	nvarchar(20)	x	x	
FaxNumber	Fax number of supplier	nvarchar(20)	x	x	
SupplierReference	Supplier reference for our organization (might be our account number at the supplier)	nvarchar(20)	x	x	

Postalcode	Delivery postal code for the supplier	nvarchar(10)	x	x	

Dim Transaction Type

Table 9: : Transactional and master data for Dim Transaction Type

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
TransactionTypeKey	DW key for the transaction type dimension	int			
TransactionTypeId	Numeric ID used for reference to a transaction type within the WWI database	int	x	x	
TransactionType	Full name of the transaction type	nvarchar(50)	x	x	

Fact Sales

Table 10: Transactional and master data for Fact Sales

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
SaleKey	DW key for a row in the Sale fact	int			
EmployeeKey	Emloyee for this invoice	int	x		
LocationKey	Location for this invoice	int	x		
CustomerKey	Customer for this invoice	int	x		

StockItemKey	Stock item for this invoice	int	x		
DateKey	Date for this invoice	date	x		
InvoiceID	InvoiceID in source system	int	x		
Quantity	Quantity supplied	int	x	x	
UnitPrice	Unit price charged	decimal(18, 2)	x	x	
TaxRate	Tax rate applied	decimal(18, 3)	x	x	
TaxAmount	Total amount of tax	decimal(18, 2)	x	x	
LineProfit	The profit earned from a single line item in a transaction, calculated as ExtendedPrice - TotalCost.	decimal(18, 2)	x	x	
TotalCost	Total cost of a line item in a transaction (including taxes, fees, etc.).	decimal(18, 2)	x	x	
ExtendedPrice	Extended price of a line item in a transaction: Quantity * UnitPrice. This value represents the item price before any discounts or adjustments.	decimal(18, 2)	x	x	
TotalExcluding Tax	Total amount including tax	decimal(18, 2)	x	x	
TotalDryItems	Total number of dry items	int	x	x	
TotalChillerItems	Total number of chiller items	int	x	x	

Fact Purchase

Table 11: Transactional and master data for Fact Purchase

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
PurchaseKey	DW key for a row in the Purchase fact	int			
Datekey	Purchase order date	date	x		
SupplierKey	Supplier for this purchase order	int	x		
StockItemKey	Stock item for this purchase order	int	x		
LastReceiptDate	The date of the most recent receipt of stock from the supplier for this product.	date	x		x
SupplierReference	A unique identifier for the supplier of this product.	nvarchar(20)	x	x	
OrderedOuters	Quantity of outers (ordering packages)	int	x	x	
ReceivedOuter	Received outers (so far)	int	x	x	
Is Order Finalized	Is this purchase order now finalized?	bit	x	x	
ExpectedUnitPrice PerOuter	Expected unit price for each outer (package or container) of the	decimal(18, 2)	x	x	

	product as specified in the order.				

Fact Warehouse

Table 12: Transactional and master data for Fact Warehouse

Field Name	Descriptive	Type	Null	Master Data	Transaction Data
StockHolding Key	DW key for a row in the Stock Holding fact	int			
DateKey	Date for this invoice	date	x		
StockItemKey	Stock item being held	int	x		
Transaction TypeKey	Foreign key referencing the Transaction Type dimension, indicating the type of transaction for the stock item.	int	x		
CustomerKey	Foreign key referencing the Customer dimension, identifying the customer involved in the transaction.	int	x		
SupplierKey	Foreign key referencing the Supplier dimension, identifying the supplier involved in the transaction.	int	x		
StockItem TransactionID	Unique identifier for the stock item transaction.	int	x	x	

InvoiceID	Foreign key referencing the Invoice dimension, identifying the invoice associated with the transaction.	int	x	x	
Purchase OrderID	Foreign key referencing the Purchase Order dimension, identifying the purchase order associated with the transaction.	int	x	x	
Quantity	The quantity of the stock item involved in the transaction. Positive values indicate an increase in stock, negative values indicate a decrease in stock.	decimal(18, 3)	x	x	

4.1.3 Data Warehouse model (Snowflake or Star, or galaxy)

Figure 1: Relational model

When it comes to the relational model, our project decided to implement a star schema. A star schema can be thought of as a star-like shape, where a central fact table is connected to one or more dimension tables.

In a star schema, each dimension table is connected to the fact table through a primary key-foreign key relationship, which allows for efficient querying and analysis of large datasets. The star schema is often used in decision support systems and business intelligence applications, as it simplifies the process of querying and reporting on complex data.

4.2 ETL processes

In our process in order to load data to the warehouse we use Azure Data Factories to support us. Specifically, we will load data from Database Wide World Importers to the Data Warehouse as follows:

Table 13: Source & Destination of project.

Source	Destination
Wide World Importer (Cloud)	DW_tllag (Cloud)

For the loading dimension table, we will create each pipeline corresponding to each dimension table. In it, we will add "Copy data" to move data from source to our dim table, select columns to be migrated by writing a SQL query.

With the fact table, in order to load data we use a lookup Activity for each dimension:

- Create a pipeline in Azure Data Factory and add a "Data Flow" activity to the pipeline.

- In "Data Flow" , add a staging table (Query SQL) and dimension table follow Data Modelling.
- Configure the "Lookup" activity by specifying the columns for the dimension table and staging table that you want to use as a lookup table.
- Map the columns in the Lookup output to the columns in the fact table using the mapping feature of the "Sink" activity.
- Run the pipeline to load data from the dimension table to the fact table.

DimDate

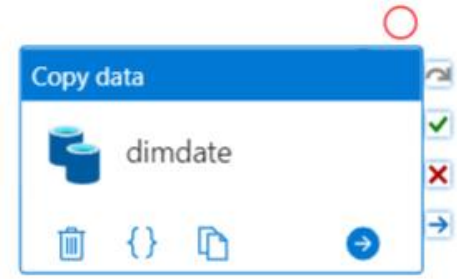


Figure 2: Load dimension for Date

Dim Customer

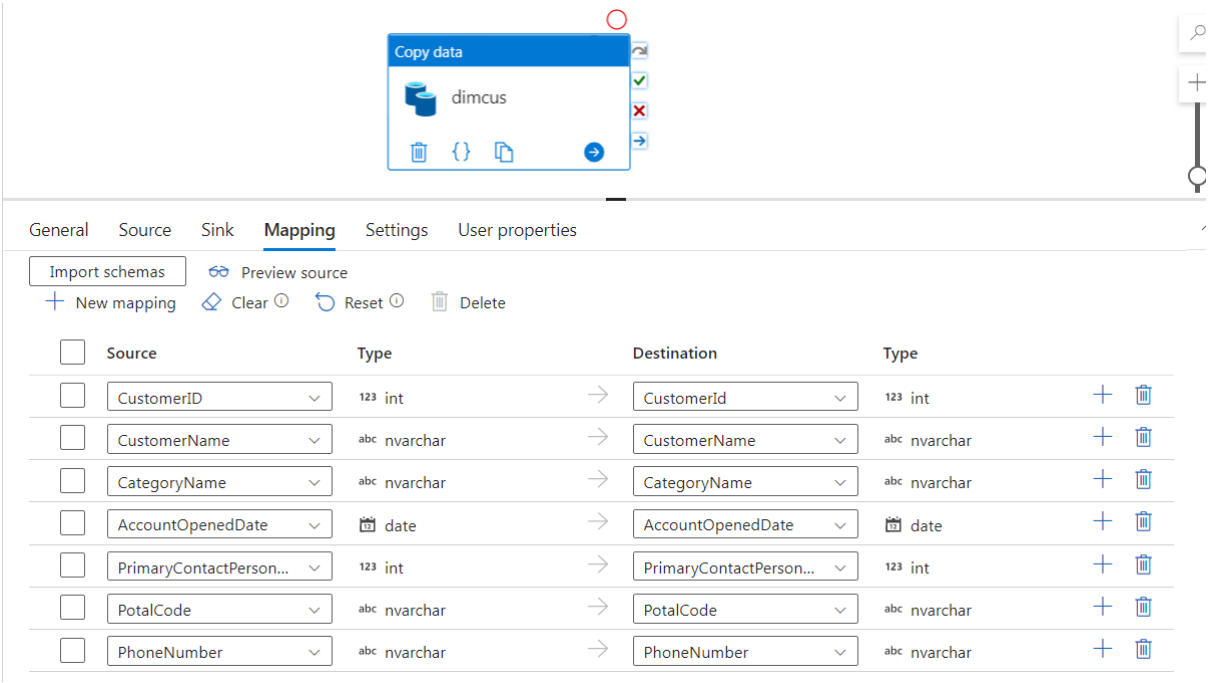


Figure 3: Load dimension for Customer

Dim Employee

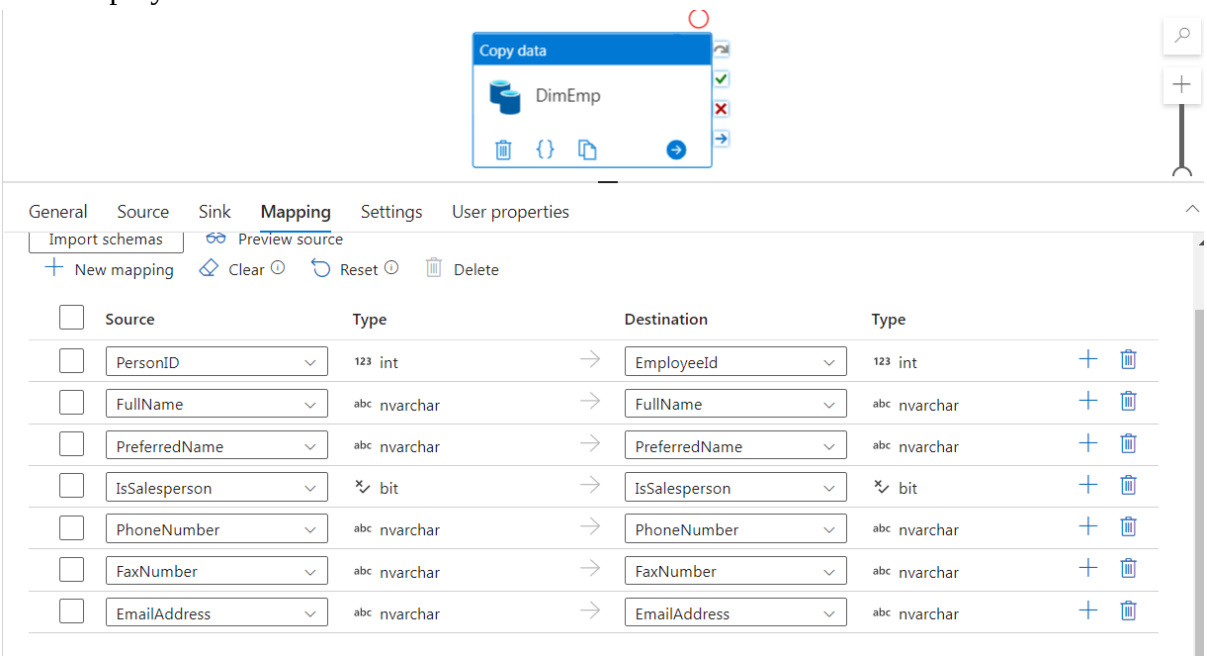


Figure 4: Load dimension for Employee

Dim Location

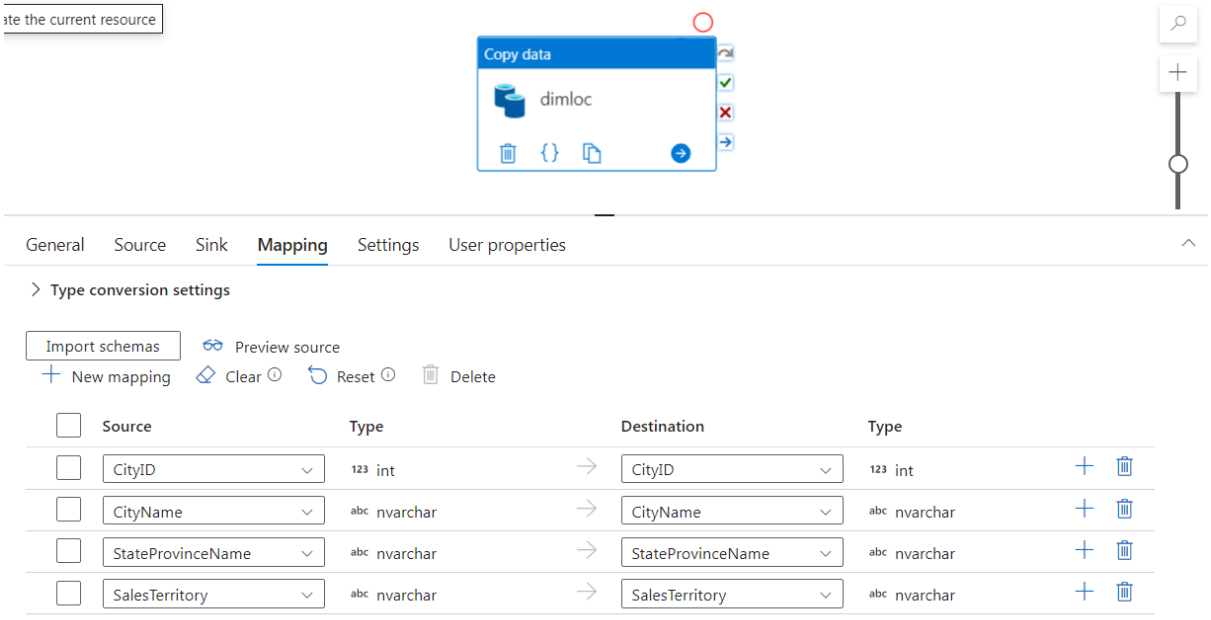


Figure 5: Load dimension for Location

Dim StockItem

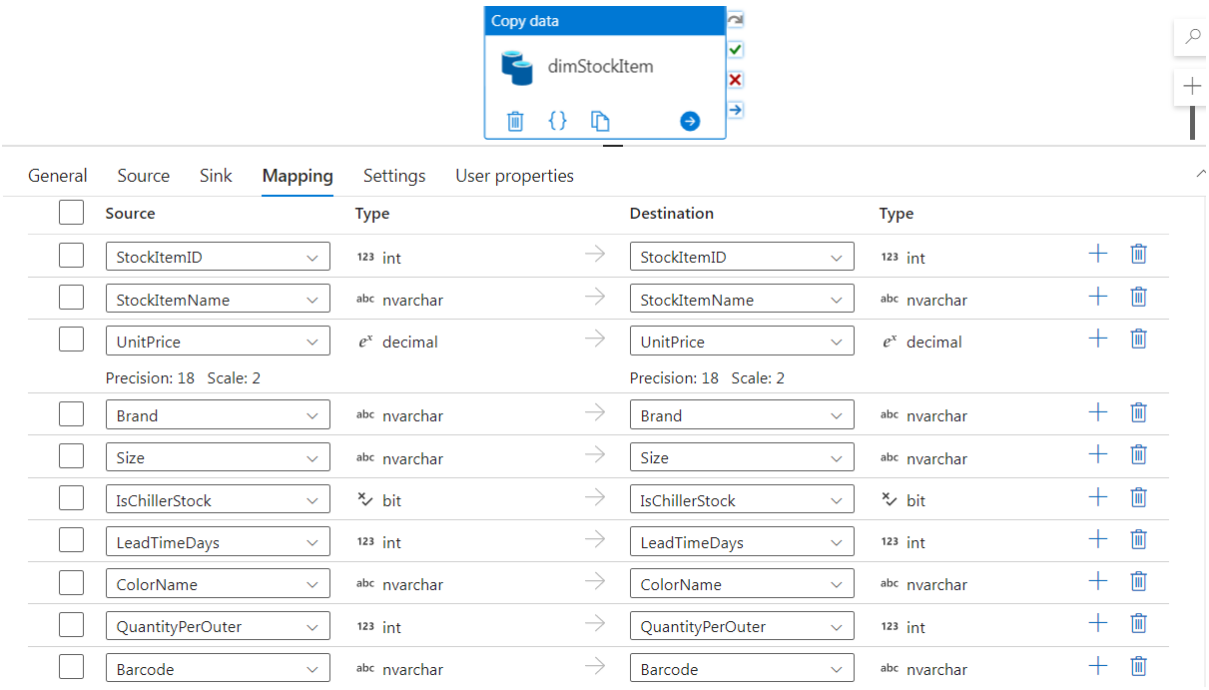


Figure 6: Load dimension for StockItem

Dim Supplier

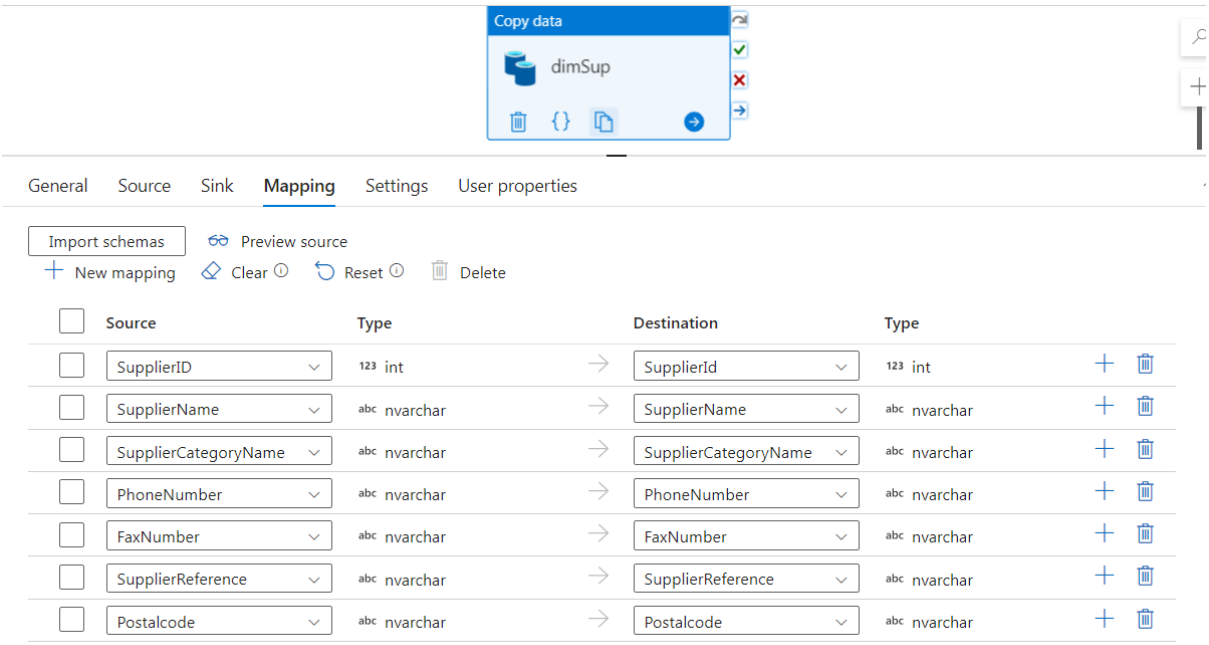


Figure 7: Load dimension for Supplier

Dim Transaction Type

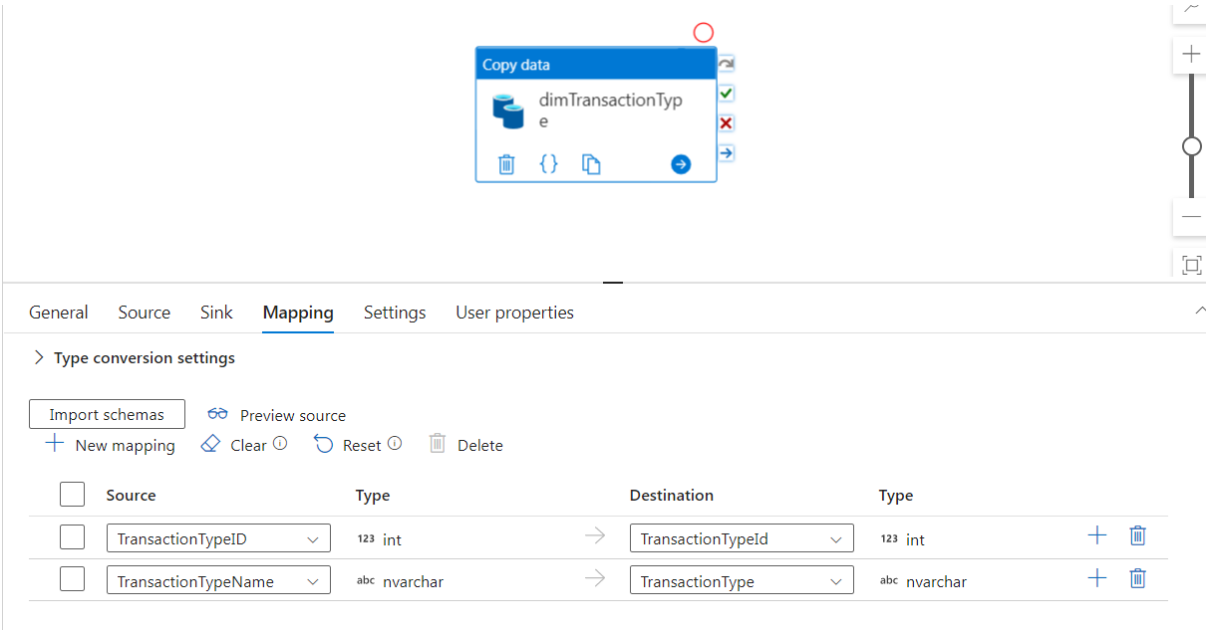


Figure 8: Load dimension for Transaction Type

Fact Sales

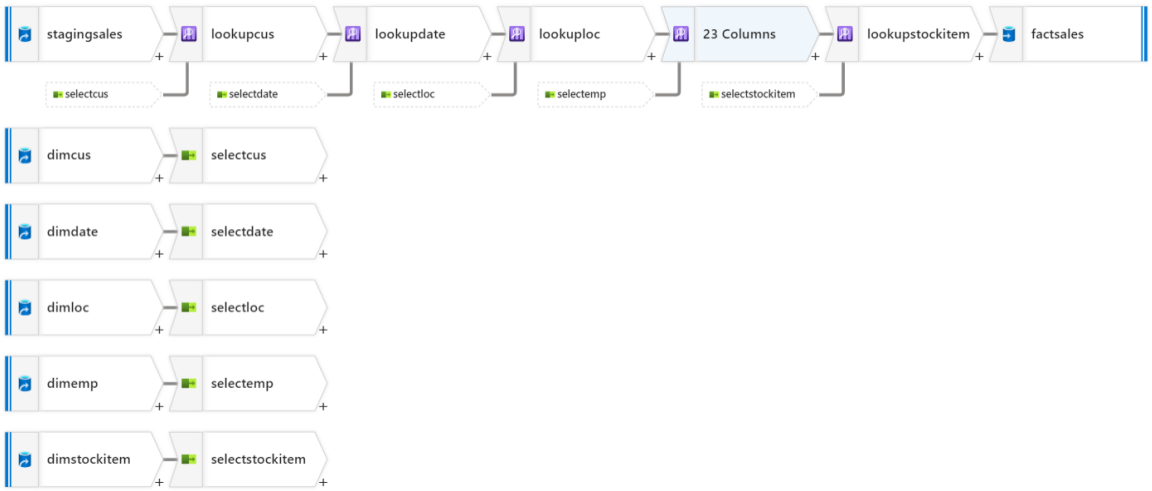


Figure 9: Load Fact Sales

Fact Purchase

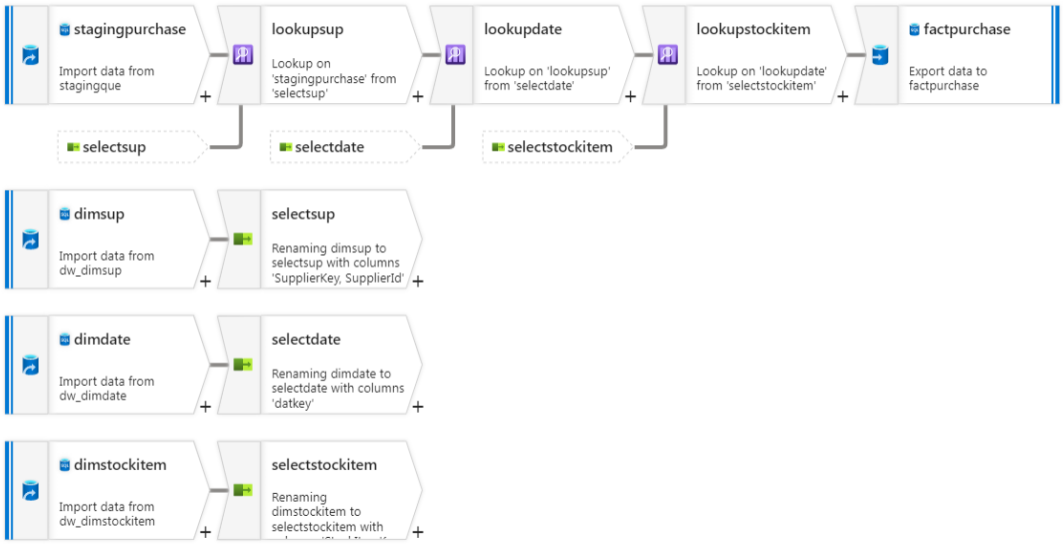


Figure 10: Load Fact Purchase

Fact Warehouse

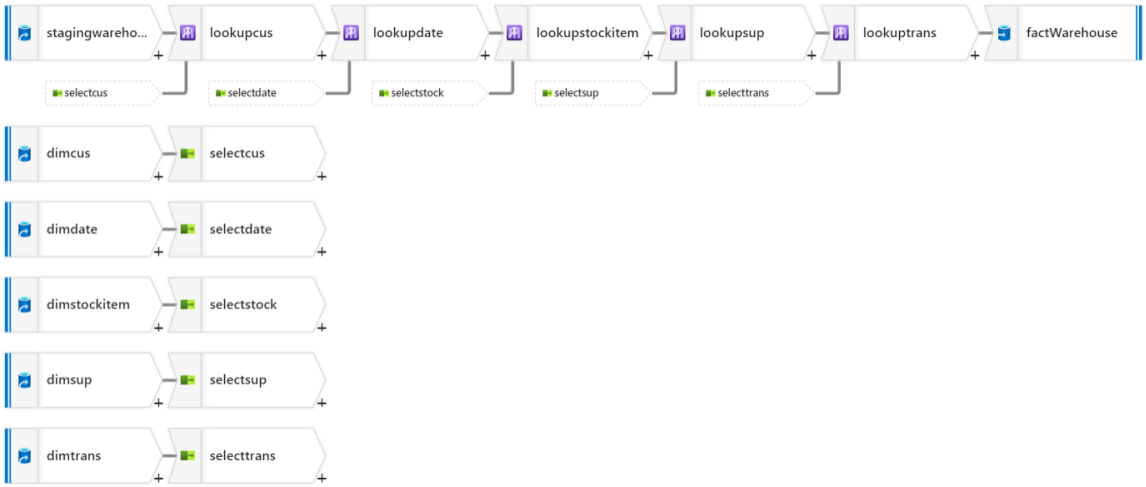
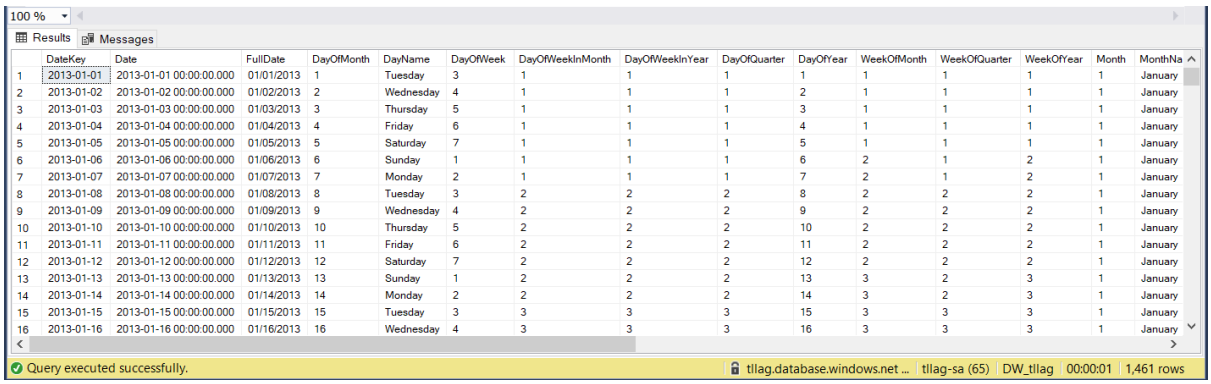


Figure 11: Load Fact Warehouse

4.2.1 Dimension Table's ETL Process

DimDate

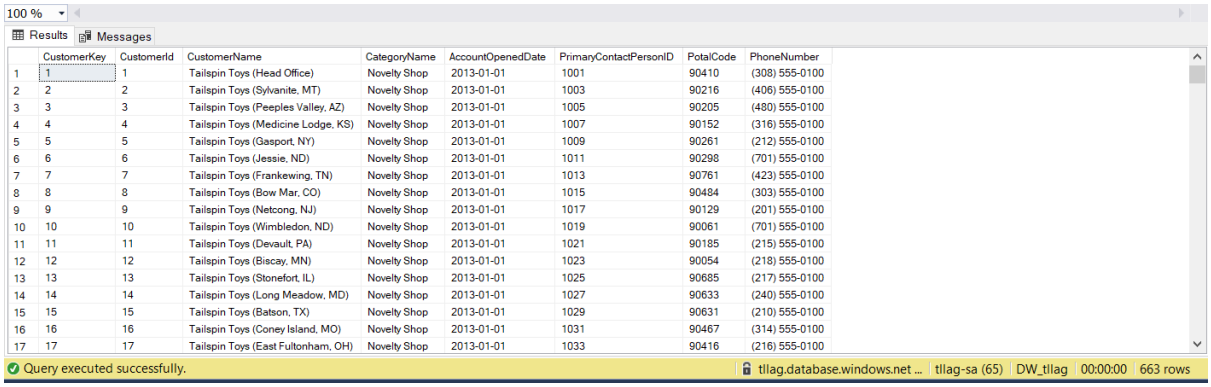


	DateKey	Date	FullDate	DayOfMonth	DayName	DayOfWeek	DayOfWeekInMonth	DayOfWeekInYear	DayOfQuarter	DayOfYear	WeekOfMonth	WeekOfQuarter	WeekOfYear	Month	MonthNa
1	2013-01-01	2013-01-01 00:00:00.000	01/01/2013	1	Tuesday	3	1	1	1	1	1	1	1	1	January
2	2013-01-02	2013-01-02 00:00:00.000	01/02/2013	2	Wednesday	4	1	1	1	2	1	1	1	1	January
3	2013-01-03	2013-01-03 00:00:00.000	01/03/2013	3	Thursday	5	1	1	1	3	1	1	1	1	January
4	2013-01-04	2013-01-04 00:00:00.000	01/04/2013	4	Friday	6	1	1	1	4	1	1	1	1	January
5	2013-01-05	2013-01-05 00:00:00.000	01/05/2013	5	Saturday	7	1	1	1	5	1	1	1	1	January
6	2013-01-06	2013-01-06 00:00:00.000	01/06/2013	6	Sunday	1	1	1	1	6	2	1	2	1	January
7	2013-01-07	2013-01-07 00:00:00.000	01/07/2013	7	Monday	2	1	1	1	7	2	1	2	1	January
8	2013-01-08	2013-01-08 00:00:00.000	01/08/2013	8	Tuesday	3	2	2	2	8	2	2	2	1	January
9	2013-01-09	2013-01-09 00:00:00.000	01/09/2013	9	Wednesday	4	2	2	2	9	2	2	2	1	January
10	2013-01-10	2013-01-10 00:00:00.000	01/10/2013	10	Thursday	5	2	2	2	10	2	2	2	1	January
11	2013-01-11	2013-01-11 00:00:00.000	01/11/2013	11	Friday	6	2	2	2	11	2	2	2	1	January
12	2013-01-12	2013-01-12 00:00:00.000	01/12/2013	12	Saturday	7	2	2	2	12	2	2	2	1	January
13	2013-01-13	2013-01-13 00:00:00.000	01/13/2013	13	Sunday	1	2	2	2	13	3	2	3	1	January
14	2013-01-14	2013-01-14 00:00:00.000	01/14/2013	14	Monday	2	2	2	2	14	3	2	3	1	January
15	2013-01-15	2013-01-15 00:00:00.000	01/15/2013	15	Tuesday	3	3	3	3	15	3	3	3	1	January
16	2013-01-16	2013-01-16 00:00:00.000	01/16/2013	16	Wednesday	4	3	3	3	16	3	3	3	1	January

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:01 | 1,461 rows

Figure 12: Data in Date Dimension

Dim Customer

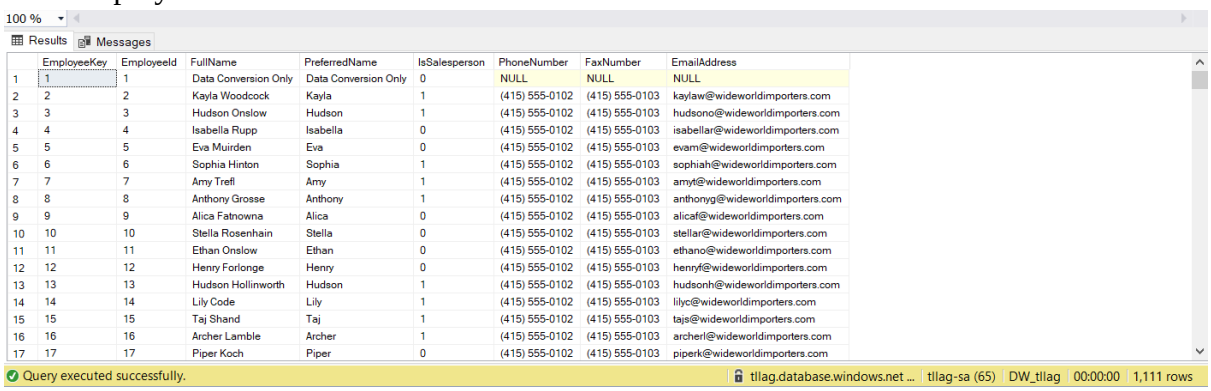


	CustomerKey	CustomerId	CustomerName	CategoryName	AccountOpenedDate	PrimaryContactPersonID	PotalCode	PhoneNumber
1	1	1	Tailspin Toys (Head Office)	Novelty Shop	2013-01-01	1001	90410	(308) 555-0100
2	2	2	Tailspin Toys (Sylvanite, MT)	Novelty Shop	2013-01-01	1003	90216	(406) 555-0100
3	3	3	Tailspin Toys (Peeples Valley, AZ)	Novelty Shop	2013-01-01	1005	90205	(480) 555-0100
4	4	4	Tailspin Toys (Medicine Lodge, KS)	Novelty Shop	2013-01-01	1007	90152	(316) 555-0100
5	5	5	Tailspin Toys (Gasport, NY)	Novelty Shop	2013-01-01	1009	90261	(212) 555-0100
6	6	6	Tailspin Toys (Jessie, ND)	Novelty Shop	2013-01-01	1011	90298	(701) 555-0100
7	7	7	Tailspin Toys (Frankewing, TN)	Novelty Shop	2013-01-01	1013	90761	(423) 555-0100
8	8	8	Tailspin Toys (Bow Mar, CO)	Novelty Shop	2013-01-01	1015	90484	(303) 555-0100
9	9	9	Tailspin Toys (Netcong, NJ)	Novelty Shop	2013-01-01	1017	90129	(201) 555-0100
10	10	10	Tailspin Toys (Wimbledon, ND)	Novelty Shop	2013-01-01	1019	90061	(701) 555-0100
11	11	11	Tailspin Toys (Devault, PA)	Novelty Shop	2013-01-01	1021	90185	(215) 555-0100
12	12	12	Tailspin Toys (Biscay, MN)	Novelty Shop	2013-01-01	1023	90054	(218) 555-0100
13	13	13	Tailspin Toys (Stonefort, IL)	Novelty Shop	2013-01-01	1025	90685	(217) 555-0100
14	14	14	Tailspin Toys (Long Meadow, MD)	Novelty Shop	2013-01-01	1027	90633	(240) 555-0100
15	15	15	Tailspin Toys (Batson, TX)	Novelty Shop	2013-01-01	1029	90631	(210) 555-0100
16	16	16	Tailspin Toys (Coney Island, MO)	Novelty Shop	2013-01-01	1031	90467	(314) 555-0100
17	17	17	Tailspin Toys (East Fultonham, OH)	Novelty Shop	2013-01-01	1033	90416	(216) 555-0100

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:00 | 663 rows

Figure 13: Data in Customer Dimension

Dim Employee

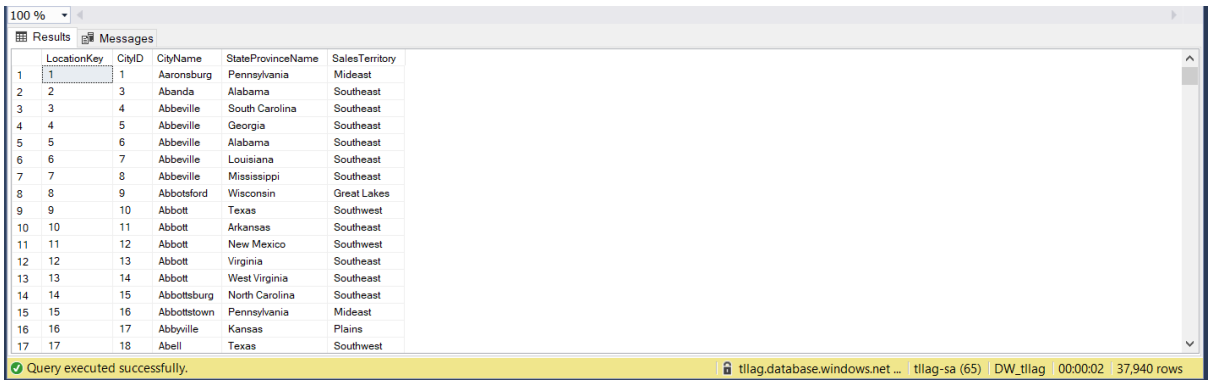


	EmployeeKey	EmployeeId	FullName	PreferredName	IsSalesperson	PhoneNumber	FaxNumber	EmailAddress
1	1	1	Data Conversion Only	Data Conversion Only	0	NULL	NULL	NULL
2	2	2	Kayla Woodcock	Kayla	1	(415) 555-0102	(415) 555-0103	kaylaw@wideworldimporters.com
3	3	3	Hudson Onslow	Hudson	1	(415) 555-0102	(415) 555-0103	hudsono@wideworldimporters.com
4	4	4	Isabella Rupp	Isabella	0	(415) 555-0102	(415) 555-0103	isabellar@wideworldimporters.com
5	5	5	Eva Muirden	Eva	0	(415) 555-0102	(415) 555-0103	evam@wideworldimporters.com
6	6	6	Sophia Hinton	Sophia	1	(415) 555-0102	(415) 555-0103	sophiah@wideworldimporters.com
7	7	7	Amy Trefl	Amy	1	(415) 555-0102	(415) 555-0103	amyt@wideworldimporters.com
8	8	8	Anthony Grosse	Anthony	1	(415) 555-0102	(415) 555-0103	anthonyg@wideworldimporters.com
9	9	9	Alica Fatnowna	Alica	0	(415) 555-0102	(415) 555-0103	alicaf@wideworldimporters.com
10	10	10	Stella Rosenhain	Stella	0	(415) 555-0102	(415) 555-0103	stellar@wideworldimporters.com
11	11	11	Ethan Onslow	Ethan	0	(415) 555-0102	(415) 555-0103	ethano@wideworldimporters.com
12	12	12	Henry Forlonge	Henry	0	(415) 555-0102	(415) 555-0103	henryf@wideworldimporters.com
13	13	13	Hudson Hollinworth	Hudson	1	(415) 555-0102	(415) 555-0103	hudsonh@wideworldimporters.com
14	14	14	Lily Code	Lily	1	(415) 555-0102	(415) 555-0103	lilyc@wideworldimporters.com
15	15	15	Taj Shand	Taj	1	(415) 555-0102	(415) 555-0103	tajs@wideworldimporters.com
16	16	16	Archer Lamble	Archer	1	(415) 555-0102	(415) 555-0103	archerl@wideworldimporters.com
17	17	17	Piper Koch	Piper	0	(415) 555-0102	(415) 555-0103	piperk@wideworldimporters.com

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:00 | 1,111 rows

Figure 14: Data in Emloyee Dimension

Dim Location

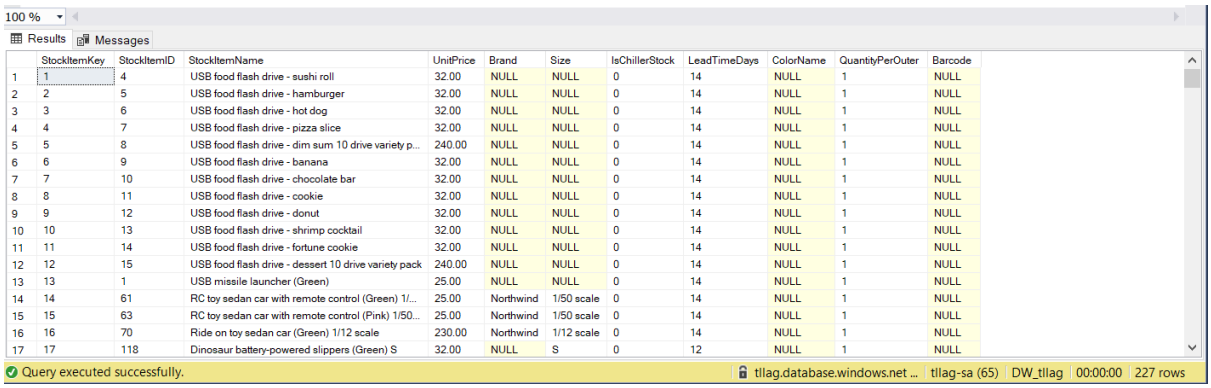


	LocationKey	CityID	CityName	StateProvinceName	SalesTerritory
1	1	1	Aaronsburg	Pennsylvania	Mideast
2	2	3	Abanda	Alabama	Southeast
3	3	4	Abbeville	South Carolina	Southeast
4	4	5	Abbeville	Georgia	Southeast
5	5	6	Abbeville	Alabama	Southeast
6	6	7	Abbeville	Louisiana	Southeast
7	7	8	Abbeville	Mississippi	Southeast
8	8	9	Abbotsford	Wisconsin	Great Lakes
9	9	10	Abbott	Texas	Southwest
10	10	11	Abbott	Arkansas	Southeast
11	11	12	Abbott	New Mexico	Southwest
12	12	13	Abbott	Virginia	Southeast
13	13	14	Abbott	West Virginia	Southeast
14	14	15	Abbottsburg	North Carolina	Southeast
15	15	16	Abbottstown	Pennsylvania	Mideast
16	16	17	Abbyville	Kansas	Plains
17	17	18	Abell	Texas	Southwest

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:02 | 37,940 rows

Figure 15: Data in Location Dimension

Dim StockItem

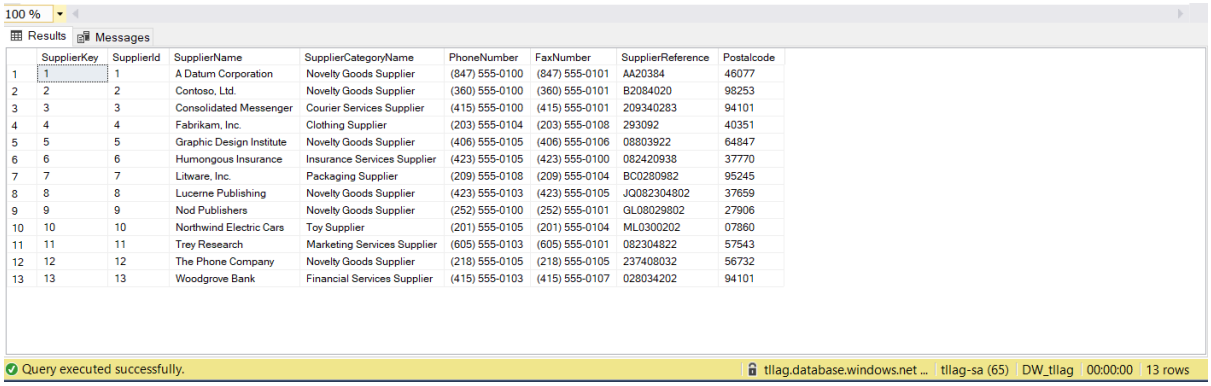


	StockItemKey	StockItemID	StockItemName	UnitPrice	Brand	Size	IsChillerStock	LeadTimeDays	ColorName	QuantityPerOuter	Barcode
1	1	4	USB food flash drive - sushi roll	32.00	NULL	NULL	0	14	NULL	1	NULL
2	2	5	USB food flash drive - hamburger	32.00	NULL	NULL	0	14	NULL	1	NULL
3	3	6	USB food flash drive - hot dog	32.00	NULL	NULL	0	14	NULL	1	NULL
4	4	7	USB food flash drive - pizza slice	32.00	NULL	NULL	0	14	NULL	1	NULL
5	5	8	USB food flash drive - dim sum 10 drive variety p...	240.00	NULL	NULL	0	14	NULL	1	NULL
6	6	9	USB food flash drive - banana	32.00	NULL	NULL	0	14	NULL	1	NULL
7	7	10	USB food flash drive - chocolate bar	32.00	NULL	NULL	0	14	NULL	1	NULL
8	8	11	USB food flash drive - cookie	32.00	NULL	NULL	0	14	NULL	1	NULL
9	9	12	USB food flash drive - donut	32.00	NULL	NULL	0	14	NULL	1	NULL
10	10	13	USB food flash drive - shrimp cocktail	32.00	NULL	NULL	0	14	NULL	1	NULL
11	11	14	USB food flash drive - fortune cookie	32.00	NULL	NULL	0	14	NULL	1	NULL
12	12	15	USB food flash drive - dessert 10 drive variety pack	240.00	NULL	NULL	0	14	NULL	1	NULL
13	13	1	USB missile launcher (Green)	25.00	NULL	NULL	0	14	NULL	1	NULL
14	14	61	RC toy sedan car with remote control (Green) 1/...	25.00	Northwind	1/50 scale	0	14	NULL	1	NULL
15	15	63	RC toy sedan car with remote control (Pink) 1/50...	25.00	Northwind	1/50 scale	0	14	NULL	1	NULL
16	16	70	Ride on toy sedan car (Green) 1/12 scale	230.00	Northwind	1/12 scale	0	14	NULL	1	NULL
17	17	118	Dinosaur battery-powered slippers (Green) S	32.00	NULL	S	0	12	NULL	1	NULL

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:00 | 227 rows

Figure 16: Data in StockItem Dimension

Dim Supplier

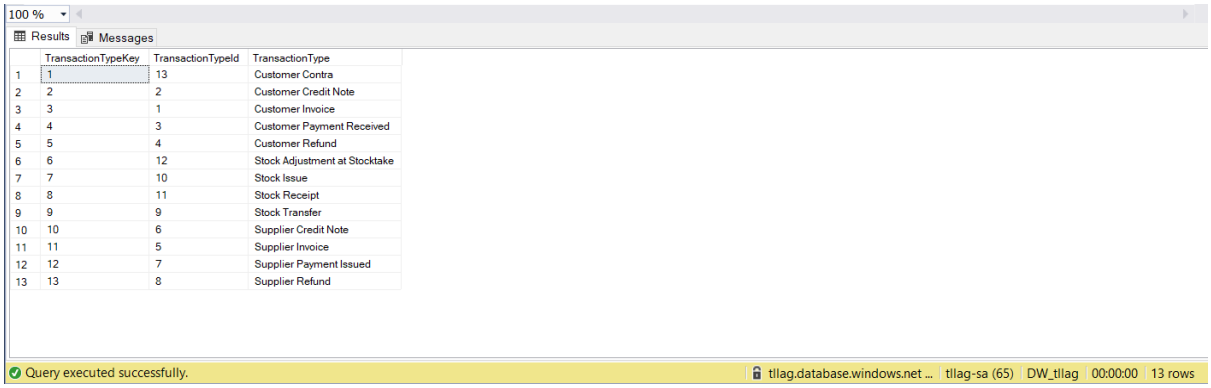


	SupplierKey	SupplierId	SupplierName	SupplierCategoryName	PhoneNumber	FaxNumber	SupplierReference	Postalcode
1	1	1	A Datum Corporation	Novelty Goods Supplier	(847) 555-0100	(847) 555-0101	AA20384	46077
2	2	2	Contoso, Ltd.	Novelty Goods Supplier	(360) 555-0100	(360) 555-0101	B2084020	98253
3	3	3	Consolidated Messenger	Courier Services Supplier	(415) 555-0100	(415) 555-0101	209340283	94101
4	4	4	Fabrikam, Inc.	Clothing Supplier	(203) 555-0104	(203) 555-0108	293092	40351
5	5	5	Graphic Design Institute	Novelty Goods Supplier	(406) 555-0105	(406) 555-0106	08803922	64847
6	6	6	Humongous Insurance	Insurance Services Supplier	(423) 555-0105	(423) 555-0100	082420938	37770
7	7	7	Litware, Inc.	Packaging Supplier	(209) 555-0108	(209) 555-0104	BC0280982	95245
8	8	8	Lucerne Publishing	Novelty Goods Supplier	(423) 555-0103	(423) 555-0105	JQ082304802	37659
9	9	9	Nod Publishers	Novelty Goods Supplier	(252) 555-0100	(252) 555-0101	GL08029802	27906
10	10	10	Northwind Electric Cars	Toy Supplier	(201) 555-0105	(201) 555-0104	ML0300202	07860
11	11	11	Trey Research	Marketing Services Supplier	(605) 555-0103	(605) 555-0101	082304822	57543
12	12	12	The Phone Company	Novelty Goods Supplier	(218) 555-0105	(218) 555-0105	237408032	56732
13	13	13	Woodgrove Bank	Financial Services Supplier	(415) 555-0103	(415) 555-0107	028034202	94101

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:00 | 13 rows

Figure 17: Data in Supplier Dimension

Dim Transaction Type



100 %

Results | Messages

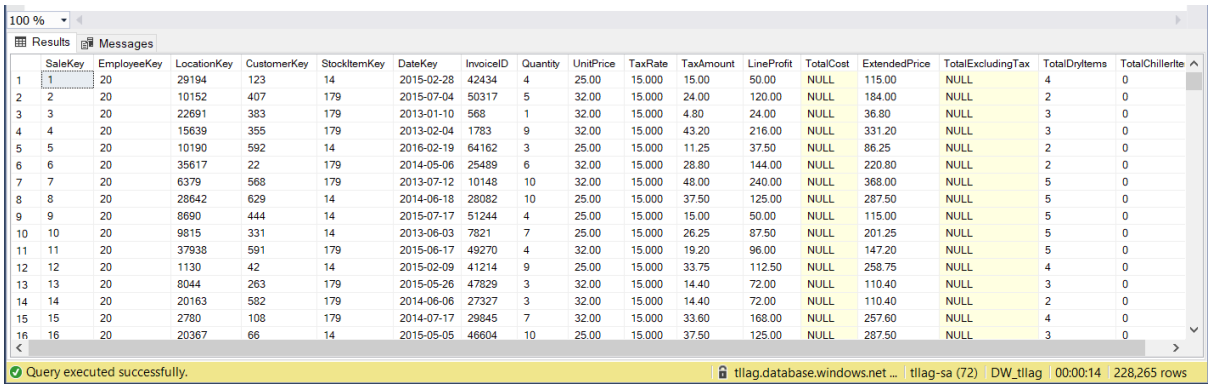
	TransactionTypeKey	TransactionTypeId	TransactionType
1	1	13	Customer Contra
2	2	2	Customer Credit Note
3	3	1	Customer Invoice
4	4	3	Customer Payment Received
5	5	4	Customer Refund
6	6	12	Stock Adjustment at Stocktake
7	7	10	Stock Issue
8	8	11	Stock Receipt
9	9	9	Stock Transfer
10	10	6	Supplier Credit Note
11	11	5	Supplier Invoice
12	12	7	Supplier Payment Issued
13	13	8	Supplier Refund

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (65) | DW_tllag | 00:00:00 | 13 rows

Figure 18: Data in Transaction Type Dimension

4.2.2 Fact Table's ETL Process

Fact Sales



100 %

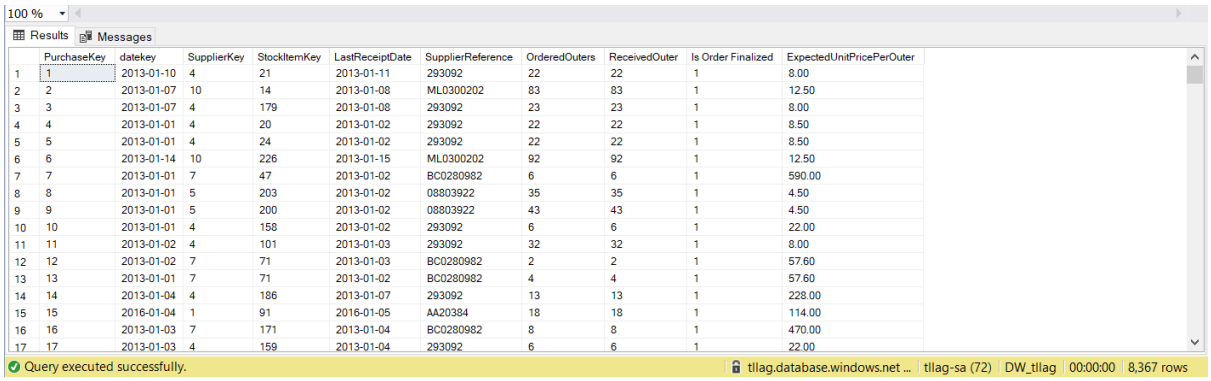
Results | Messages

	SaleKey	EmployeeKey	LocationKey	CustomerKey	StockItemKey	DateKey	InvoiceID	Quantity	UnitPrice	TaxRate	TaxAmount	LineProfit	TotalCost	ExtendedPrice	TotalExcludingTax	TotalDryItems	TotalChillerIte
1	1	20	29194	123	14	2015-02-28	42434	4	25.00	15.000	15.00	50.00	NULL	115.00	NULL	4	0
2	2	20	10152	407	179	2015-07-04	50317	5	32.00	15.000	24.00	120.00	NULL	184.00	NULL	2	0
3	3	20	22691	383	179	2013-01-10	568	1	32.00	15.000	4.80	24.00	NULL	36.80	NULL	3	0
4	4	20	15639	355	179	2013-02-04	1783	9	32.00	15.000	43.20	216.00	NULL	331.20	NULL	3	0
5	5	20	10190	592	14	2016-02-19	64162	3	25.00	15.000	11.25	37.50	NULL	86.25	NULL	2	0
6	6	20	35617	22	179	2014-05-06	25489	6	32.00	15.000	28.80	144.00	NULL	220.80	NULL	2	0
7	7	20	6379	568	179	2013-07-12	10148	10	32.00	15.000	48.00	240.00	NULL	368.00	NULL	5	0
8	8	20	28642	629	14	2014-06-18	28082	10	25.00	15.000	37.50	125.00	NULL	287.50	NULL	5	0
9	9	20	8690	444	14	2015-07-17	51244	4	25.00	15.000	15.00	50.00	NULL	115.00	NULL	5	0
10	10	20	9815	331	14	2013-06-03	7821	7	25.00	15.000	26.25	87.50	NULL	201.25	NULL	5	0
11	11	20	37938	591	179	2015-06-17	49270	4	32.00	15.000	19.20	96.00	NULL	147.20	NULL	5	0
12	12	20	1130	42	14	2015-02-09	41214	9	25.00	15.000	33.75	112.50	NULL	258.75	NULL	4	0
13	13	20	8044	263	179	2015-05-26	47829	3	32.00	15.000	14.40	72.00	NULL	110.40	NULL	3	0
14	14	20	20163	582	179	2014-06-06	27327	3	32.00	15.000	14.40	72.00	NULL	110.40	NULL	2	0
15	15	20	2780	108	179	2014-07-17	29845	7	32.00	15.000	33.60	168.00	NULL	257.60	NULL	4	0
16	16	20	20367	66	14	2015-05-05	46604	10	25.00	15.000	37.50	125.00	NULL	287.50	NULL	3	0

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (72) | DW_tllag | 00:00:14 | 228,265 rows

Figure 19: Data in Fact Sales

Fact Purchase



100 %

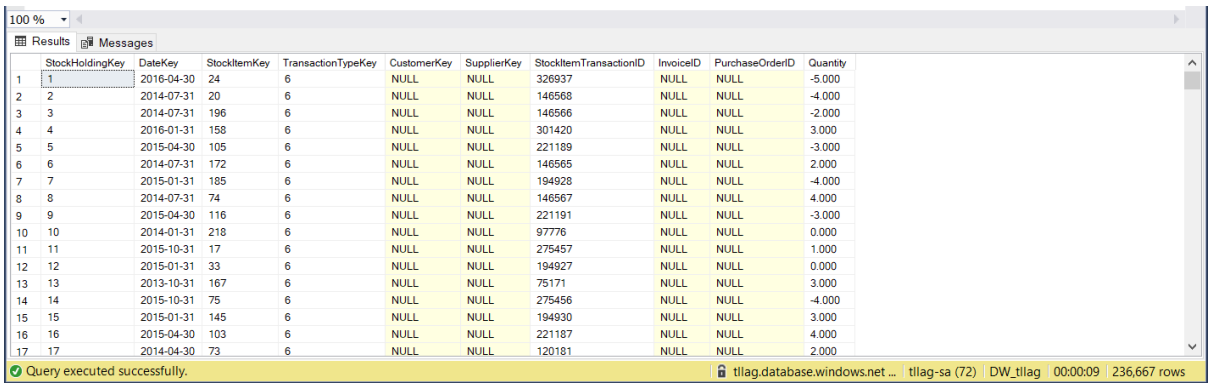
Results | Messages

	PurchaseKey	datekey	SupplierKey	StockItemKey	LastReceiptDate	SupplierReference	OrderedOuters	ReceivedOuter	Is Order Finalized	ExpectedUnitPricePerOuter
1	1	2013-01-10	4	21	2013-01-11	293092	22	22	1	8.00
2	2	2013-01-07	10	14	2013-01-08	ML0300202	83	83	1	12.50
3	3	2013-01-07	4	179	2013-01-08	293092	23	23	1	8.00
4	4	2013-01-01	4	20	2013-01-02	293092	22	22	1	8.50
5	5	2013-01-01	4	24	2013-01-02	293092	22	22	1	8.50
6	6	2013-01-14	10	226	2013-01-15	ML0300202	92	92	1	12.50
7	7	2013-01-01	7	47	2013-01-02	BC0280982	6	6	1	590.00
8	8	2013-01-01	5	203	2013-01-02	08803922	35	35	1	4.50
9	9	2013-01-01	5	200	2013-01-02	08803922	43	43	1	4.50
10	10	2013-01-01	4	158	2013-01-02	293092	6	6	1	22.00
11	11	2013-01-02	4	101	2013-01-03	293092	32	32	1	8.00
12	12	2013-01-02	7	71	2013-01-03	BC0280982	2	2	1	57.60
13	13	2013-01-01	7	71	2013-01-02	BC0280982	4	4	1	57.60
14	14	2013-01-04	4	186	2013-01-07	293092	13	13	1	228.00
15	15	2016-01-04	1	91	2016-01-05	AA20384	18	18	1	114.00
16	16	2013-01-03	7	171	2013-01-04	BC0280982	8	8	1	470.00
17	17	2013-01-03	4	159	2013-01-04	293092	6	6	1	22.00

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (72) | DW_tllag | 00:00:00 | 8,367 rows

Figure 20: Data in Fact Purchase

Fact Warehouse



100 %

Results | Messages

	StockHoldingKey	DateKey	StockItemKey	TransactionTypeKey	CustomerKey	SupplierKey	StockItemTransactionID	InvoiceID	PurchaseOrderID	Quantity
1	1	2016-04-30	24	6	NULL	NULL	326937	NULL	NULL	-5.000
2	2	2014-07-31	20	6	NULL	NULL	146568	NULL	NULL	-4.000
3	3	2014-07-31	196	6	NULL	NULL	146566	NULL	NULL	-2.000
4	4	2016-01-31	158	6	NULL	NULL	301420	NULL	NULL	3.000
5	5	2015-04-30	105	6	NULL	NULL	221189	NULL	NULL	-3.000
6	6	2014-07-31	172	6	NULL	NULL	146565	NULL	NULL	2.000
7	7	2015-01-31	185	6	NULL	NULL	194928	NULL	NULL	-4.000
8	8	2014-07-31	74	6	NULL	NULL	146567	NULL	NULL	4.000
9	9	2015-04-30	116	6	NULL	NULL	221191	NULL	NULL	-3.000
10	10	2014-01-31	218	6	NULL	NULL	97776	NULL	NULL	0.000
11	11	2015-10-31	17	6	NULL	NULL	275457	NULL	NULL	1.000
12	12	2015-01-31	33	6	NULL	NULL	194927	NULL	NULL	0.000
13	13	2013-10-31	167	6	NULL	NULL	75171	NULL	NULL	3.000
14	14	2015-10-31	75	6	NULL	NULL	275456	NULL	NULL	-4.000
15	15	2015-01-31	145	6	NULL	NULL	194930	NULL	NULL	3.000
16	16	2015-04-30	103	6	NULL	NULL	221187	NULL	NULL	4.000
17	17	2014-04-30	73	6	NULL	NULL	120181	NULL	NULL	2.000

Query executed successfully. | tllag.database.windows.net ... | tllag-sa (72) | DW_tllag | 00:00:09 | 236,667 rows

Figure 21: Data in Fact Warehouse

CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION

5.1 Report and dashboard systems (structure)

The system's structure includes the following components:

- Define purpose: define the purpose of the report or dashboard. What question should be answered or displayed to solve a problem or make a decision.
- Data collection: After defining the purpose, it is necessary to collect data from different sources (tables) and prepare them for analysis.
- Design a data model: Before starting to build a report or dashboard, it is necessary to design a data model. This is the process of defining tables, relations, attributes, and primary keys.
- Create reports or dashboards: use Power BI Desktop to create reports or dashboards. Choose chart and table types to display data, create filters to categorize data, and create analytics to answer questions.

Dashboard structure is the main elements and layout of the dashboard that assist users to understand information easily and quickly. Dashboard structure plays an important role in creating an effective and valuable dashboard.

A dashboard structure that integrates arrangement, formatting, and sentence classification visualizes information, helping to create an attractive and easy-to-use layout for users. Here are some of the key elements commonly used in the dashboard structure:

- Title: The title should be clear, describe the topic, and image if possible.
- Tables and charts: Easy-to-understand charts and tables provide users with useful information quickly. They should be made clear, simple and easy to read.
- Tags: Tags are used to categorize the information in the dashboard. It makes it easy for users to navigate to the element they are interested in.
- Metrics: A key metric or indicator should be included in the dashboard to provide an overall view of the performance of a company or a given region. When comparing figures between time points, using symbols such as color icons makes the metrics easier.
- Actions: Dashboards can provide links/operations that users can use to make decisions or take necessary actions.

5.2 Data analysis with Power BI

5.2.1 Sales

Using the BI tool to help us clearly see the contrast and interaction between variables such as Quantity and Untprice, combining the two variables will produce revenue results from which we can compare sales. revenue of each product, each region.

PI gives us a better overview of data visualization, making it easier for users to understand the parameters related to the model we are interested in.

Steps to perform visualization using PI:

- Load data from SQL through the tool on PI (select the appropriate table for calculation). Tables already have real data from sql side
- Create related charts
- Customize and filter data
- Finally, share and integrate

Thanks to PI analysis, we can easily access the data, create amazing charts and reports.

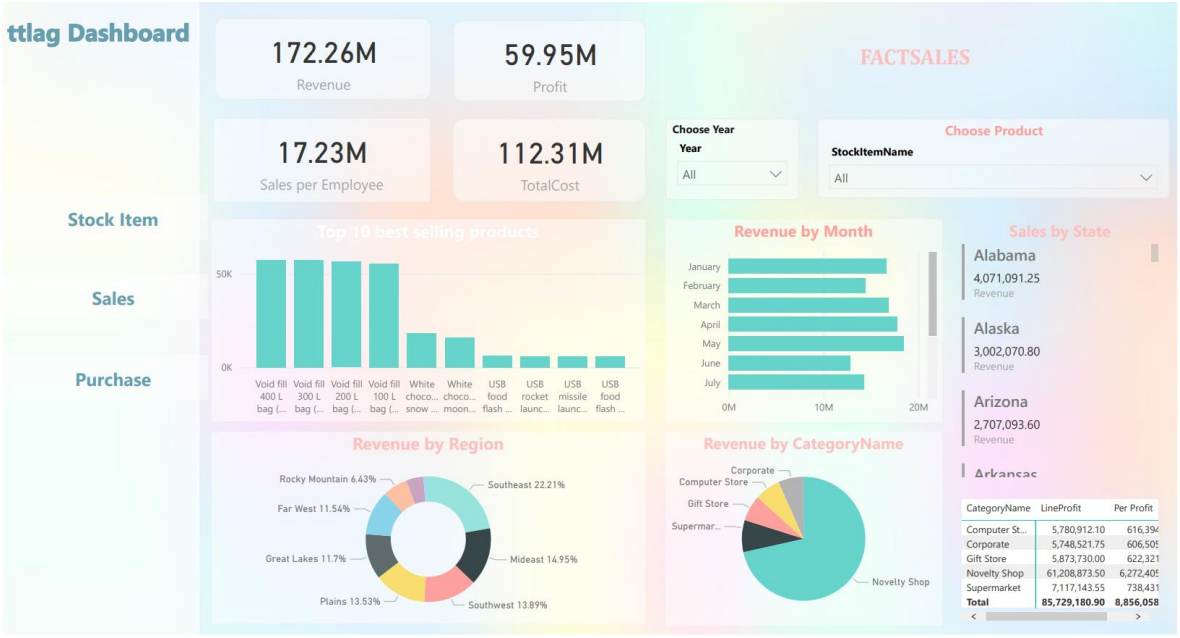


Figure 22: Dashboard factsales

5.2.2 Purchase Department

Using Microsoft Power BI to analyze and visualize data. Power BI provides a variety of tools that allow for quick insights into data. Some of the key features of Power BI that have been used for purchasing visualizations are:

- Data visualization: create interactive dashboards and reports that provide insights into data such as total product categories purchased, top most purchased products, total key product categories, etc. .
- Data Analytics: Use these tools to perform ad-hoc analytics on data and gain insights into the data such as vendor analysis to see what is the best supplier, trends purchase goods through each business stage to have a development direction for each product to help grow revenue.

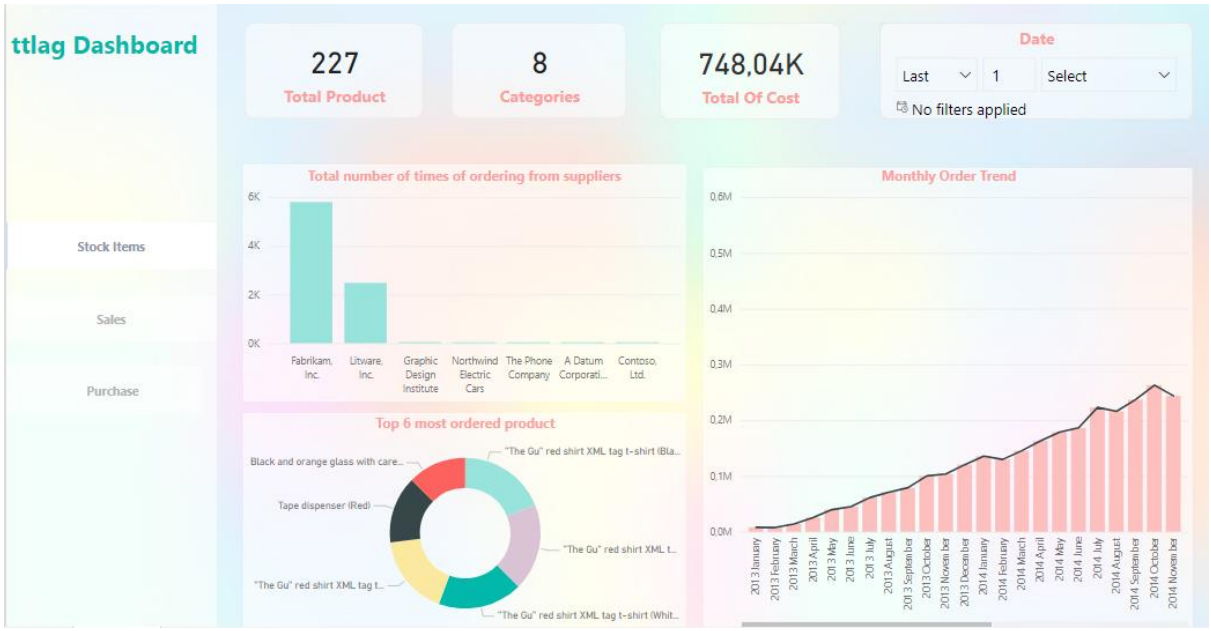


Figure 23: Dashboard Purchase

5.2.3 Warehouse Department

Using Microsoft Power BI to analyze and visualize data can provide valuable insights into warehouse performance. Power BI offers a range of features and tools that are useful for creating interactive dashboards and reports, and performing ad-hoc analytics on data. Some of the key Power BI features that are relevant to warehouse analytics and visualization include:

- Data visualization: Power BI enables the creation of interactive dashboards and reports that provide insights into warehouse data such as inventory levels, quantity of stockitem, warehouse import and export. With data visualization tools, it's possible to track key performance indicators (KPIs) and metrics and monitor trends over time.
- Data analytics: With Power BI, it's possible to perform ad-hoc analytics on warehouse data to gain insights into the data such as identifying inventory trends, forecasting demand for certain products.

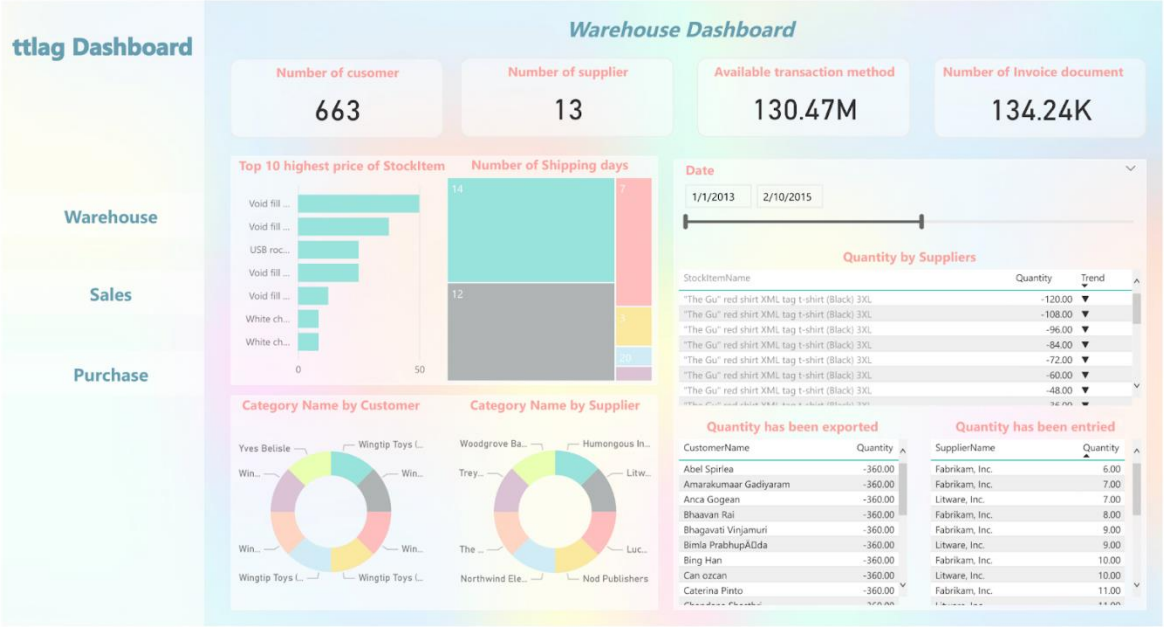


Figure 24: Dashboard Warehouse

5.3 Evaluation and Discussion

5.3.1 Sales

With the completion of the data analysis dashboard, here are the thoughts and conclusions of each variable:

- Sales Revenue: helps identify customer behavioral trends and their product preferences, allowing businesses to optimize products, sales and marketing strategies to maximize revenue.
- Sales per employee: used to measure the efficiency and performance of the business's sales team. This metric reflects whether each employee effectively contributes to the company's sales. Analyzing sales per employee can help a company better understand which employees are top performers, incentivize employees to improve sales performance, and reduce sales costs.
- Sales by Region: Provides information about the company's performance in different markets. With this information, businesses can understand what products and services are required in each region and adjust their sales and marketing strategies accordingly, such as investing more in one area or product. specific product.
- Sales by product category: provides information on which products are performing well and which are not. This can help businesses realize opportunities for product diversity, promotions, and improved customer experience.

Accurate sales statistics help identify weaknesses, opportunities and gaps in sales performance, maximize profits and help businesses compete in dynamic markets.

Below are the results of the research to answer the questions posed at the beginning of the article:

Figure 25: Calculate relevant parameters

Based on the above results, we can see the revenue and profit achieved on the total data. There is also revenue calculated by employees of 17.23 million.

Figure 26: Top 10 best selling product

The combination of the product variable and the revenue variable shows which products are selling the most and generating the most revenue in the existing product categories.

Some of the common bottlenecks we encounter that have been resolved through data analysis include:

- We remove duplicate information to avoid affecting the prediction results.
- Data analytics can help determine which lead generation channels are most effective and which aren't.

Analysis helps find bottlenecks, inefficiencies, from improvements to increased productivity, efficiency and higher revenue.

5.3.2 Purchase Department

Dashboards and reporting systems have enabled the Purchase Department to better understand its data and make data-driven decisions. The system has brought the following benefits:

- Improved visibility: better visibility into its data, allowing it to identify trends and patterns that were previously hidden.
- Improved decision-making: Departments can make data-driven decisions based on insights gained from dashboards and reporting systems.
- Improved collaboration: The system has enabled the department to collaborate with other parts of the organization such as Sales, share insights, and improve the overall decision-making process.
- Reduced manual effort: The system has reduced the amount of manual effort required to analyze and report data, freeing up resources to focus on higher value-added tasks.

For Purchase Department, by using BI tools (PowerBI), we have visualized the dashboard for Purchase Department, for detail:

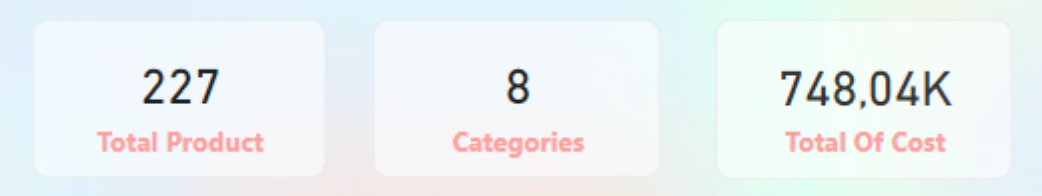


Figure 27: Total number of products, total list of goods and total cost of goods

By examining the data presented above, we can determine both the overall quantity of products that the company is selling and the total categories of goods that the company is currently providing to the market. It can be seen that the company is currently not diversifying many products, has very few product categories.

Figure 28: Top 6 most ordered products

With the most purchased products from the Procurement department. We can look there to get an overview of which products are popular and bought by customers, so that we can focus on developing those products better.

Figure 29: Monthly Order Trend

With the order trend of the Purchase Department, we can see the purchase volume over the months, thereby evaluating and giving strategies to improve the months with low purchasing volume.

5.3.3 Warehouse Department

Warehouse departments can benefit from using data to identify trends and opportunities for improvement. By analyzing key metrics with Microsoft Power BI, warehouse managers can make informed decisions to optimize warehouse operations and improve performance. Some of the key metrics that can be tracked and visualized with Power BI include:

- Inventory turnover: measures how often inventory is sold and replaced over a period of time. This metric helps warehouse managers understand how efficiently inventory is moving through the warehouse and identify potential areas for improvement such as reducing excess inventory or improving order fulfillment times.
- Shipping time: measures how long it takes to fulfill and ship orders to customers. By tracking shipping time with Power BI, warehouse managers can identify areas for improvement such as reducing order processing times, optimizing warehouse layout, and improving logistics.

For warehouse department, by using BI tools (PowerBI), we have visualized the dashboard for warehouse department, for detail:

Figure 30: Customers and suppliers frequently appear in category name list

Looking at the pie chart above, the number of customers and suppliers over the past 3 years is almost stable, moreover, they are also wholesale buyers and focal suppliers, so they can be identified as loyal customers of the warehouse department

Figure 31: Customers and suppliers frequently appear in category name list

Based on the amount of inventory in the warehouse, it is possible to determine whether the warehouse has just imported or shipped goods, if the value of goods in the warehouse is negative, this proves that the warehouse has just shipped goods to the customer and vice versa, if the value of the goods is. There is a positive sign in the warehouse, which means that the new warehouse has imported more goods from suppliers.

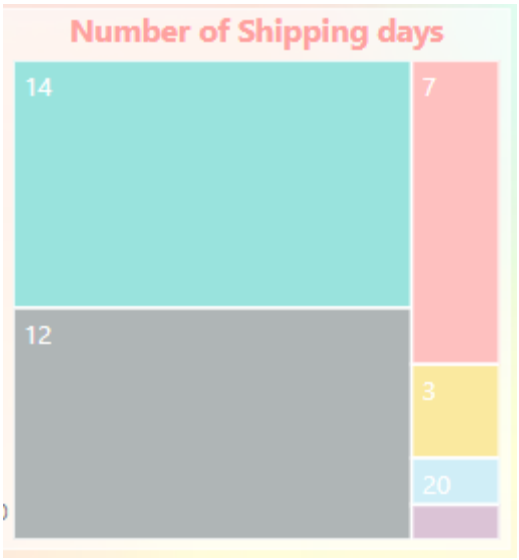


Figure 32: Number of shipping days

Shipping is not very efficient because it often takes more than 10 days for goods from the warehouse to be shipped to the customer, part of the reason for this is the volume of orders is often large. To improve this, it is also a good idea to work with a shipping company, but if you want to save costs, you can transport yourself, but you must reduce the delivery time so that it is below ten.

Due to the limitation of the dataset, we cannot determine the inventory factor of the warehouse as well as whether the invoice documents are fully exported or not.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1 Results

Building data and applying BI (Business Intelligence) to the research paper helped the company have a more comprehensive view of its business activities. By using Aruze to load data, companies can efficiently and quickly collect, store, and analyze data.

From the analysis of this data, the company can recognize the market trends, the most popular products and the most effective marketing channels. This helps the company to proactively respond to the market and come up with appropriate business strategies, thereby helping to increase sales and profits.

In addition, applying BI also helps companies improve business processes by optimizing operations for maximum efficiency. Such as reducing order processing time, improving production processes or optimizing warehouse management.

Therefore, using Aruze to load and build data, apply BI has helped the company become more developed and improve business in many aspects, thereby increasing the competitiveness in the market.

6.2 Limitations

Limitations when analyzing and predicting with WWI . datasets

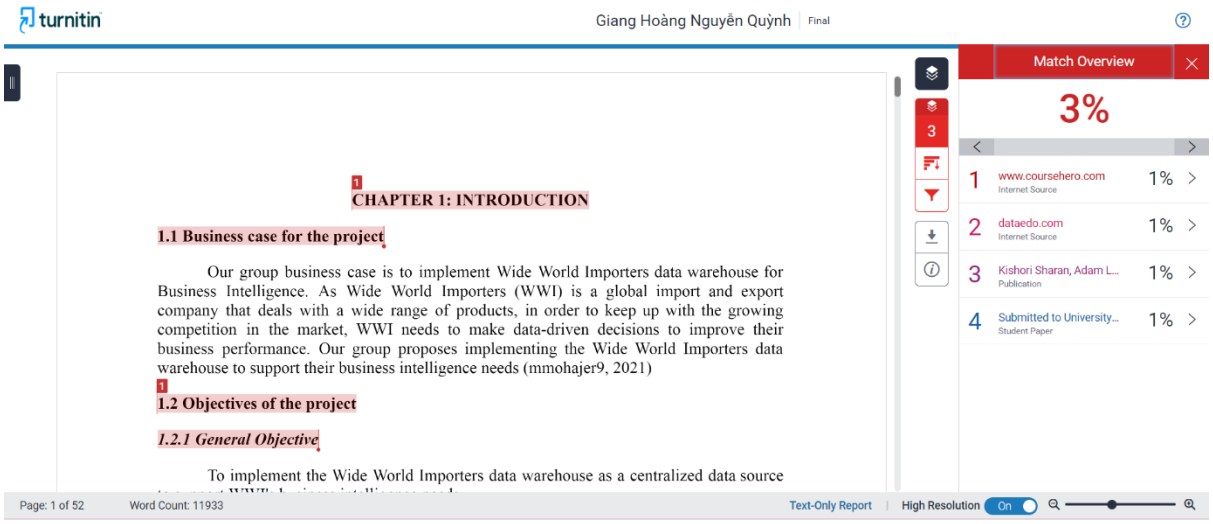
- Insufficient richness: Small-scale datasets can lead to inaccurate or misleading analysis results.
- Heterogeneous data: Data is not uniform in terms of do, format, and region and regional level. It also leads to confusion and difficulty in data analysis.
- Lack of real-time information: Data sets become obsolete. Information related to businesses often changes constantly, so if you do not update new information, the analysis results are also inaccurate.
- The WWI dataset has not been widely used in the analysis community, so it is difficult for us to find supporting topics and answer questions related to the data.

6.3. Future works

Future works of Wide World Importers data may include:

- Add new data: includes detailed product descriptions, product photos, and more detailed information about suppliers or customers.
- Combining Wide World Importers data with a CRM system will help businesses better manage customer relationships and improve customer care.
- Data optimization: Data stored on Wide World Importers needs to be optimized to speed up data retrieval and minimize storage costs.
- Integration with online sales platforms: helping businesses reach a large number of potential customers and increase sales.
- Develop data analysis features: help businesses better understand market trends, customer needs and improve business strategies

REFERENCES

[1] Costello, T., & Blackshear, L. (2020). Prepare Your Data for Tableau. USA.
[2] Data Warehouse Star Schema vs Snowflake Schema. ((n.d.)). Javatpoint.
[3] Genesys. ((n.d.)). Bus Matrix. Genesys Documentation.
[4] Migrate workloads to Azure. Tech Data Cloud. . ((n.d.)). Microsoft.
[5] Mohajer, M. M. (2021). wideworld-importers-data-warehouse. github.
[6] Ranjan, J. (2005). Business Intelligence: Concepts, Components, Techniques And Benefits. Journal of Theoretical and Applied Information Technology (pp. Vol 9. No 1. (pp 060-070)). India: http://www.jatit.org/.
[7] TechTarget. ((n.d.)). Windows Azure. TechTarget.
[8] University, V. (2022). What Is a Business Process? | Definition, Importance and Examples. Villanova University.
[9] Vigintas Sakys, Kestutis Kapocius, Rimantas Butleris, Audrius Lopata and Saulius Gudas. (2013). The Framework for Business Intelligence Driven Analysis of Study Course Teaching Efficiency. Transformations in Business & Economics, (pp. Vol. 12, No 1A (28A), pp.42-59).
[10] Wide World Importers: What is it? Microsoft Learn. ((n.d.)). Microsoft.
[11] Zeng, L., Xu, L., Shi, Z., Wang, M. and Wu, W. (2007). Techniques, process, and enterprise solutions of business intelligence. IEEE Conference on Systems, Man, and Cybernetics, (pp. Vol. 6, pp. 4722.). Taipei, Taiwan October 8-11, 2006.

Nhóm đã thực hiện check phần mềm Turnitin