UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

COURSE: BUSINESS INTELLIGENCE AND DECISION SUPPORT SYSTEM

Topic:

BUILD A DATA WAREHOUSE IN AZURE CLOUD BASE ON THE ADVENTUREWORKS 2019 DATABASE

MODULES: SALES AND PURCHASING

Ho Chi Minh City, May 8, 2023

UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

COURSE: BUSINESS INTELLIGENCE AND DECISION SUPPORT SYSTEM

Topic:

BUILD A DATA WAREHOUSE IN AZURE CLOUD BASE ON THE ADVENTUREWORKS 2019 DATABASE

MODULES: SALES AND PURCHASING

Team UP:

NO	Name	ID	Faculty	Role	Phone	Email
1	Vu Van Son	K20410669	Khoa HTTT	Member	0944244857	sonvv20416c @st.uel.edu.vn
2	Chau Nguyen Huong Thao	K204162000	Khoa HTTT	Member	0919772802	thaocnh20416c @st.uel.edu.vn
3	Huynh Hong Uyen	K204162006	Khoa HTTT	Leader	0934190565	uyenhh20416c @st.uel.edu.vn
4	Nguyen Tuong Vy	K204162009	Khoa HTTT	Member	0764187037	vynt20416c @st.uel.edu.vn

Lecturer: **MS. Nguyen Van Ho**

Ho Chi Minh City, May 8, 2023

ACKNOWLEDGEMENTS

Throughout the course of the project, we received a significant amount of dedicated assistance and guidance from our lecturer, and also consulted numerous authors. Thanks to their contributions, we were able to successfully complete our final project.

Despite our best efforts, errors may have been made, and we welcome any feedback or constructive criticism you may have to help improve the quality of our research. Your input will serve as a tremendous source of motivation as we continue to develop our project in the future.

COMMITMENT

We would like to assure you that our final project is original and based on the research conducted by the entire team. We have also included a list of documents that we referenced in our report and have cited them accordingly.

In the event that any of the information provided is found to be incorrect or misleading, we take full responsibility for any consequences that may arise, and we will work closely with our professors to rectify the situation.

Ho Chi Minh City, May, 5, 2023

Committed person

TEAM UPgrade

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

NO.	ACRONYMS	MEANINGS
1	BI	Business Intelligence
2	ETL	Extract - Transform - Load
3	KPI	Key Performance Indicator
4	RFM	Recency, Frequency, Monetary
5	AWC	Adventure Works Cycles

CHAPTER 1: INTRODUCTION

1.1 Business case for the project

Adventure Works is a sample database developed by Microsoft to illustrate the implementation of their software solutions, and is not an actual company operating in reality. Adventure Works is a fictional company [1], specializing in selling and manufacturing metal and composite bicycles for commercial markets in North America, Europe, and Asia. Its headquarters is located in Bothell, Washington with 290 employees, and some regional sales teams are placed across their market bases.

After a successful financial year, Adventure Works Cycles is looking to expand its market share by targeting their best customers, expanding their ability to provide products through external websites, and reducing sales costs through lower production costs. The current business system (Adventure Works 2019 database) relies heavily on reporting, but the current system does not provide maximum support for that need. With efforts to change, Adventure Works needs to implement a Business Intelligence (BI) solution to optimize data analysis and reporting. Building a data warehouse enhances the ability to analyze data and make business decisions based on accurate and timely information. It also helps to minimize the time and cost of data preparation, providing fast and easy access, and making data management easier. From there, it is possible to improve business efficiency and help the Adventure Works company achieve its goals of expanding market share and upgrading products, while reducing sales costs.

1.2 Objectives of the project

1.2.1 General Objective

The main objective of this project is to implement a Business Intelligence (BI) solution for Adventure Works, a fictional company specializing in selling and manufacturing metal and composite bicycles, through the development of a data warehouse on the Azure cloud platform.

1.2.2 Specific Objectives

- To identify and understand the current business needs and requirements of Adventure Works.
- To design and develop a data warehouse on Azure.
- To implement Extract, Transform, and Load (ETL) processes to ensure the data is accurate, reliable, and consistent.
- To develop and deploy a set of reports and dashboards that can provide useful insights and analysis to support better decision-making for Adventure Works.

- To ensure the BI solution is scalable, secure, and maintainable to support future growth and changes in business requirements.

1.3 Research Objects

The research objective of this project is to gain insights from the sales and purchasing departments of the fictional company Adventure Works, thereby developing a BI solution for analytics to support better decision-making. Research components include:

- Adventure Works is the fictional company that the BI solution is being developed for.
- **Azure cloud platform:** the cloud computing platform that the data warehouse will be developed on.
- **Data sources:** Adventure Works 2019 database.
- **ETL processes:** the processes that will be used to extract, transform, and load the data from the various sources into the data warehouse.
- **BI reports and dashboards:** The reports and dashboards that will be developed to provide useful insights and analysis to support better decision-making for Adventure Works.

1.4 Scope of the Project

The project's goal is to create a data warehouse based on the AdventureWorks sample database to cater to the primary target audience, including professors, learners, and individuals interested in this field of study.

- Analysis of the current business needs and requirements of Adventure Works.
- Design and development of a data warehouse on the Azure cloud platform.
- Implementation of Extract, Transform, and Load (ETL) processes to ensure data accuracy, reliability, and consistency.
- Development and deployment of reports and dashboards that provide useful insights and analysis to support better decision-making for Adventure Works
- Ensuring the BI solution is scalable, secure, and maintainable.

In addition, since this project is conducted in the BI&DSS subject, the project is built entirely on the breadth of knowledge from this subject. As a result, the project cannot yet deploy anything out of scope, such as training end users on how to use the BI solution or developing more special features.

1.5 Value and desired outcome of the project

AdventureWorks' project to build a data warehouse for a sample database is expected to bring significant value and desired results to the company. The data warehouse will improve AdventureWorks' reporting and analytics capabilities by making it possible to produce analysis and reports that are more precise and timely and that can be utilized to make better decisions. This will give the business a competitive advantage and enable it to stay one step ahead of its rivals. Additionally, data warehousing lowers the possibility of mistakes and discrepancies during reporting and analysis, improving the accuracy and dependability of the information. As a result, AdventureWorks will be able to make data-driven decisions that will improve business outcomes and boost competitiveness.

The project will also provide benefits in terms of cost and time reduction. With the ability to streamline data processing and analysis, the data warehouse will help to reduce the time and costs associated with preparing data for reporting and analysis, leading to improved efficiency and productivity.

In conclusion, the value and desired outcome of the project are to provide AdventureWorks with a more robust and efficient data warehouse that can support the company's reporting and analysis needs, enabling them to make better-informed decisions and achieve their business objectives. With enhanced reporting and analysis capabilities, improved data integration and accessibility, reduced costs and time, increased productivity and efficiency, and improved decision-making, the project is expected to deliver significant value to the company.

1.6 Structure of project

The project structure consists of 6 parts:

- **Chapter 1: Introduction**

This chapter will introduce your project, describe the AdventureWorks sample database, and state the goals of the project. It also includes project scope and information.

- **Chapter 2: Theoretical Basis**

This chapter will cover the fundamental knowledge and theory related to the Business Intelligence (BI) solution and concepts related to data warehouse, data integration, and data analytics, and so on.

- **Chapter 3: Requirements Analytics and Introduction to BI Solution**

This chapter will analyze the requirements and needs of your project and introduce the Business Intelligence solution.

- **Chapter 4: Building Data Warehouse and Integrating Data**

This chapter will discuss the process of building a data warehouse and integrating data sources into the data warehouse. We will detail the ETL process (extract, transform, load) and the tools and software used to build the data warehouse.

- **Chapter 5: Results - Data Analytics and Visualization**

This chapter will present the project results, including data analysis and visualization using tools and software such as Power BI. We will also discuss the results achieved and how they can support Adventure Works business decisions.

- **Chapter 6: Conclusion and Future Works**

This chapter will summarize the results of your project and provide comments and evaluations about the project implementation process. This chapter also includes recommendations for future developments and upgrades to the project.

CHAPTER 2: THEORETICAL BASIS

2.1 Overview about BI

2.1.1 What is BI?

There are many definitions of BI. According to Gartner [2], BI is defined as “Business intelligence (BI) is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance." In meanwhile, Microsoft [3] defines BI more concisely and simply as “a set of technologies, applications, and practices for collecting, integrating, analyzing, and presenting business information." According to Techtarget [4], a clearer and more detailed definition says, "Business intelligence (BI) is a technology-driven process for analyzing data and delivering actionable information that helps executives, managers and workers make informed business decisions. As part of the BI process, organizations collect data from internal IT systems and external sources, prepare it for analysis, run queries against the data and create data visualizations, BI dashboards and reports to make the analytics results available to business users for operational decision-making and strategic planning.”

To recap, BI (Business Intelligence) is the process of gathering, analyzing, and presenting data to aid in business decision-making. BI is the process of transforming raw data into actionable insights using software tools and technologies, allowing firms to make data-driven decisions and achieve a competitive advantage in their sector. Data warehouses, data mining, data analytics, and data visualization are commonly used in BI to generate reports, dashboards, and other visualizations that help users better understand their company's data. Overall, BI assists firms in increasing operating efficiency, increasing revenue, decreasing costs, and improving overall performance.

2.1.2 BI Architecture

In the study of Liya Wu et al. (2007) [5], there is mention of a simple BI architecture as shown in Figure 1. . This architecture can be considered the most elementary architecture and includes the important steps in implementing a BI solution. It shows an overview of the BI architecture, including Data Source - the various data sources used to extract information, to ETL, Data Warehouse and Data Mart and the final output.

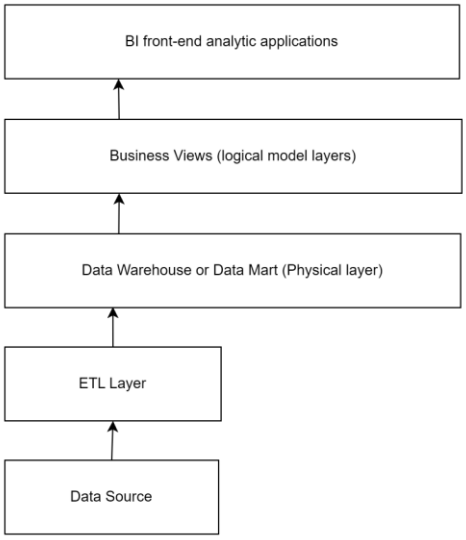

Figure 1: Conceptual architecture for business intelligence (Liya Wu, 2007)

In the study of Ong et al. (2011) [6], a BI architecture is more detailed, full of components and more steps as shown in Figure 2 below:

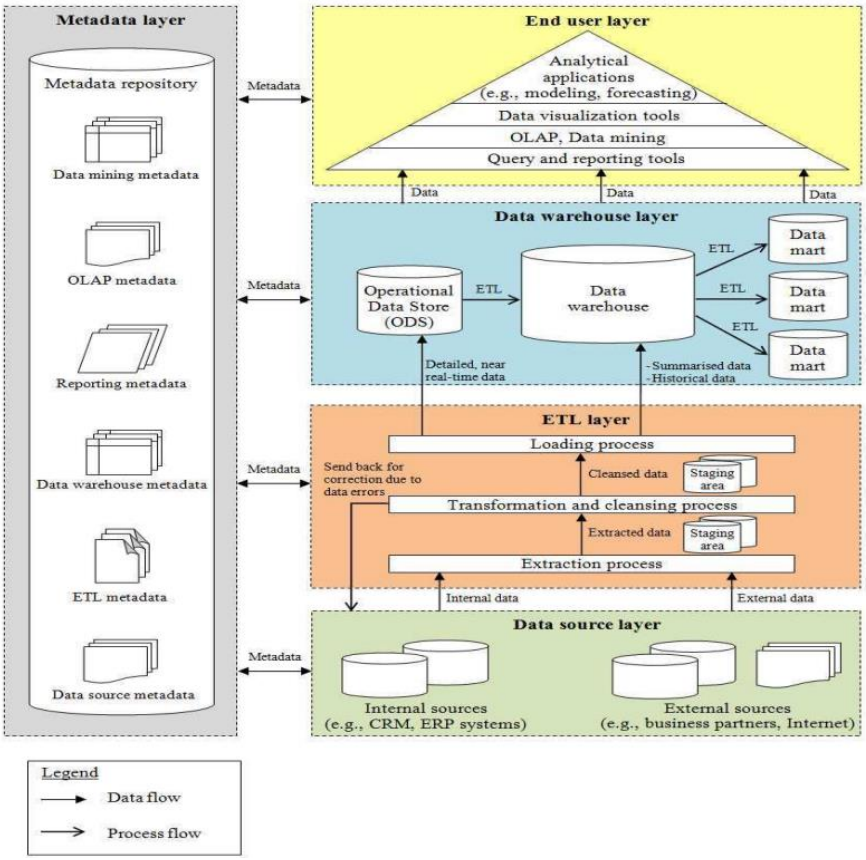

Figure 2: BI Architecture (Ong, 2011)

From Figure 2, we can see more clearly about the BI process architecture.

The first is the data source layer, now the needs of data use become much more diverse, from the requirement to use structured data to unstructured and semi-structured

data to support the use of data. Make effective and timely decisions. And all of this data can be collected from inside and outside the business.

The next layer is the ETL layer, which focuses on three main processes: data extraction, transformation, and loading. ETL processes help ensure data is synchronized and standardized for business decision and analysis purposes.

The data warehouse layer consists of three main components, namely the operational data store, the data warehouse, and the data marts. Data is first extracted from the operational data store and then moved to the data warehouse, after which it is transferred to the data marts.

The metadata layer serves as a source of information about data, providing details on its usage and storage location, its origin, and its relationships with other information. Technical and business information, along with data definitions and rules, are stored in the metadata repository. Proper management and utilization of metadata can streamline development and maintenance processes, while also providing users with useful information about data sources. For instance, stored metadata can be used for data modeling without the need to redesign data structures like table names and data types. Regular maintenance and updates to metadata in repositories are essential to ensure their accuracy and usefulness.

The end user layer includes various tools that present data in different formats to different users. These tools are arranged hierarchically in a pyramid shape, with increasing levels of data comprehensiveness and complexity as one moves up the pyramid. This hierarchical structure is designed to cater to the varying decision-making needs of different organizational levels. For instance, the highest level of the pyramid includes analytical applications primarily used by top management, while the lowest level includes query and reporting tools used mainly by operational management.

2.1.3 Advantage of BI in enterprises

Businesses benefit from Business Intelligence (BI) in a variety of ways. Here are several significant advantages:

- Data-driven decision making: Business intelligence (BI) delivers timely, accurate, and relevant data to assist firms in making educated decisions. It enables businesses to evaluate data from several sources, detect trends and patterns, and get insight into consumer behavior, market trends, and other significant aspects influencing customer behavior. Business effectiveness.
- Improve operational efficiency: By identifying bottlenecks, optimizing processes, and automating regular tasks, BI may assist firms in improving

operational efficiency. This can result in cost savings and greater resource utilization.

- Improve customer happiness: Business intelligence can help companies better understand their customers' requirements and preferences, leading to increased customer engagement and satisfaction. It enables businesses to adjust their products and services to match the needs of their customers.
- Competitive advantage: Business intelligence can provide a competitive advantage by enabling companies to make data-driven decisions faster than their competitors. It enables organizations to respond swiftly to market changes, recognize emerging trends, and stay ahead of the competition.
- Business intelligence (BI) can give a variety of benefits to various divisions within an organization.
- BI may provide insights about customers, products, and overall business performance to the sales department. This can aid in the analysis of sales performance as well as the prediction and management of sales operations. Sales employees may swiftly find new and potential market aspects, improve customer relationship management, and raise sales revenue by employing data analysis tools and software.
- BI can assist the purchasing department in analyzing data on supplier pricing, quality, and efficiency. Data analysis tools and software can assist purchasing employees in making educated decisions, saving money, and improving supplier relationship management. Furthermore, BI can provide predicted demand and market trends, allowing the purchasing department to plan wisely and make timely acquisitions.

2.1.4 BI Strategy for Business

A corporate BI strategy often entails defining specific objectives and goals for adopting BI, identifying the data sources to be used, and selecting the proper technology and tools to fulfill those objectives. It also entails building and implementing a data architecture capable of supporting the organization's data requirements, as well as ensuring that the essential data governance and security rules are in place.

The BI strategy should be consistent with the overall business strategy and concentrate on addressing specific business demands and issues. Identifying key performance indicators (KPIs) and metrics that will be used to monitor the success of the BI installation is part of this process.

To remain aligned with the increasing demands of the business and the changing technical landscape, a successful BI strategy necessitates regular examination and

continuous development. This could include doing periodic assessments of the data architecture, technological stack, and governance policies, as well as constant monitoring of KPIs and metrics to assess the efficiency of the BI implementation.

2.2 ETL Process

2.2.1 What is ETL?

ETL is an acronym for Extract, Transform, and Load, according to IBM [7]. It is a data integration procedure that combines data from multiple sources to generate a unified and consistent data store that can be fed into a data warehouse or another target system. As a result, we may conclude that ETL is a key process for transforming data from various systems for storage in a data warehouse or data analytics system. The ETL process consists of obtaining data from numerous sources, transforming the data to match the needs of the data warehouse or data analytics system, and then loading the transformed data into the warehouse, data analytics system for analysis and reporting.

2.2.2 Why do we need ETL?

ETL is required because data is frequently stored in multiple sources with varying formats, structures, and quality. ETL is required to extract pertinent data from these sources, transform and normalize it, and load it into a centralized data warehouse or destination system. We can verify data accuracy, increase data quality, decrease redundancy, and make data analysis and reporting easier. ETL also assists organizations in integrating and consolidating data from several sources, resulting in a single source of truth for decision making.

2.2.3 ETL Process

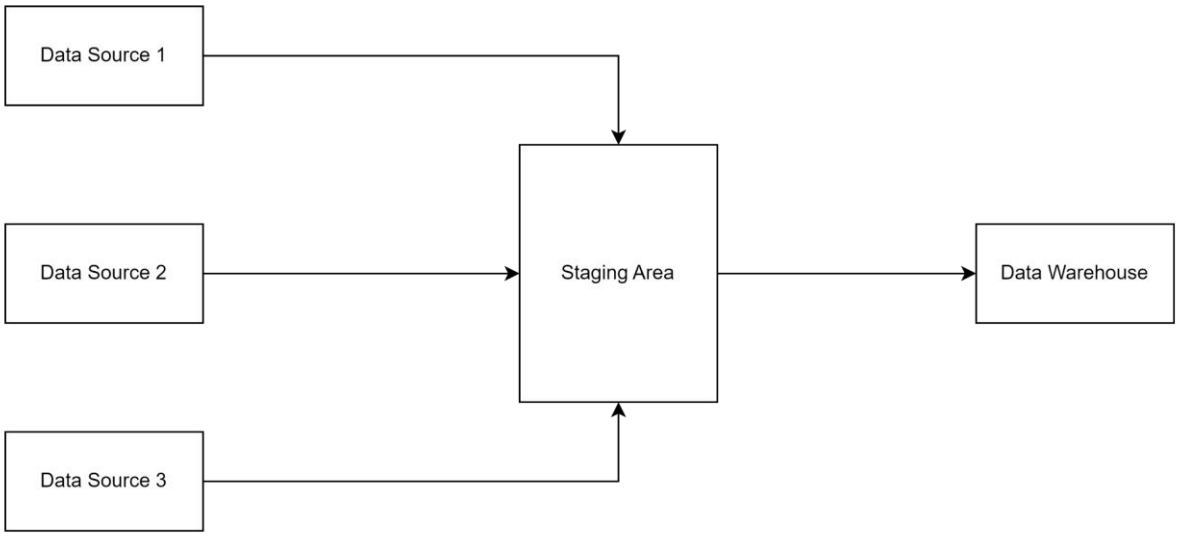

Figure 3: ETL Process (source: self-synthesized author group)

Figure 3 depicts the ETL process in a simple manner. The ETL layer is primarily concerned with three processes: data extraction, transformation, and loading. Because data is frequently received from internal and external sources that are not integrated, appropriate, incomplete, or vulnerable to duplication, the extraction process is responsible for collecting data from a range of sources and filtering out important data to support organizational choices. To ensure consistency and eliminate the need to re-extract the data, extracted data is transmitted to a temporary storage space. Data transformation and cleansing entail transforming data formats using business rules and standardizing data definitions to achieve organization-wide consistency. Data cleansing is detecting and correcting data flaws based on predefined standards to assure the data's accuracy and completeness. Once the data has been converted and cleaned, it is saved in the data area to avoid having to re-convert it if the data download fails or terminates. The data will then be loaded.

2.3 Data warehouse and Data mart

2.3.1 What are Data warehouse and Data mart?

A data warehouse is a data storage system meant to help an organization's analysis and decision-making. It is created by combining data from various sources and putting it in a single database that is structured by subject.

A data warehouse's key elements include:

- **Long-term data storage:** Data warehouses are built for long-term data storage, which includes both historical and current data. It allows an organization to access historical data and compare it to current data to get a sense of patterns and changes.
- **Organize data by topic:** Data in a data warehouse is arranged by topic or in some other way to suit the objective of the analysis. This enables people to efficiently access and utilize data.
- **Supports complex processes:** Data warehouses enable complex queries and data analyses, including combining data from multiple sources and calculating complex metrics.
- **Provide correct and consistent data:** Although data warehouses can be updated from a variety of sources, all data must be standardized and consistent before being stored in the system. This guarantees that the information collected from the data warehouse is correct and consistent.
- **Easy access and integration:** Data warehouse allows different applications and tools to easily access and integrate data. These programs can be utilized within the organization for analysis and decision-making.

A Data Mart is a simplified version of a Data Warehouse that provides users data on one specific aspect or an enterprise edge. A Data Mart's principal function is often to give essential information required to make critical decisions in a specific department or sector of the firm.

Building Data Mart makes it easier to manage and retrieve data. Instead of having to search for data in a large and complex Data Warehouse, Data Mart provides users with only the data needed to meet the needs of a particular department or department. The creation of Data Marts makes the organization's analytical and decision-making processes more efficient, because departments will be provided with accurate and consistent data to make decisions in the field. mine. By giving only the information that is required to a limited range, Data Mart additionally contributes to lowering errors and data inconsistencies.

Bill Inmon proposed the top-down approach for data warehousing, which involves creating a comprehensive data warehouse that includes all the information of a business before building data marts to serve different business needs. The Data Warehouse construction process is shown as Figure 4.

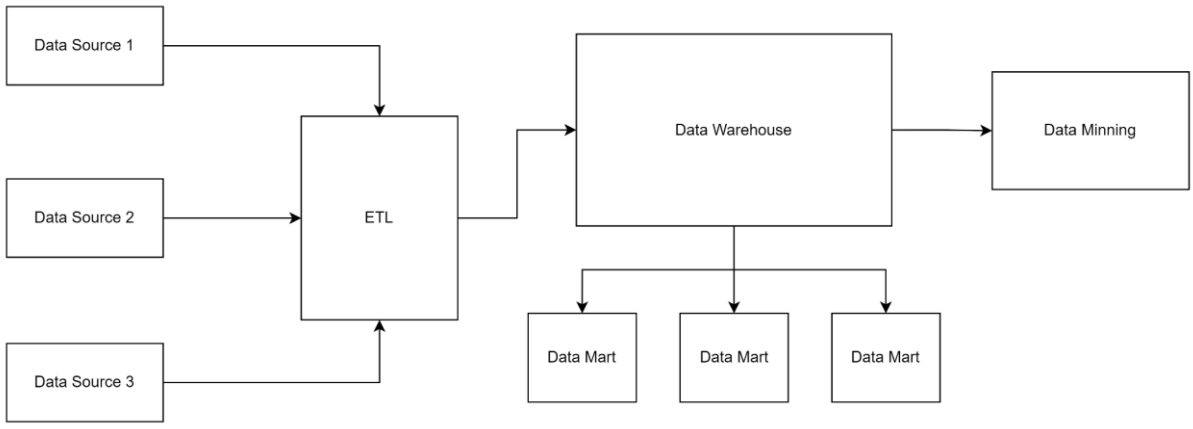

Figure 4: Bill Inmon - building Data Warehouse

(source: self-synthesized author group)

While the Bottom Up strategy (Ralph Kimball) proposes that Data Marts are established initially to satisfy the business's reporting needs. A Data Mart will be associated with a single business area, such as sales or finance. Data Mart is then combined to create a whole data warehouse. Therefore, the process of building Data Warehouse also changes as shown in Figure 5 below.

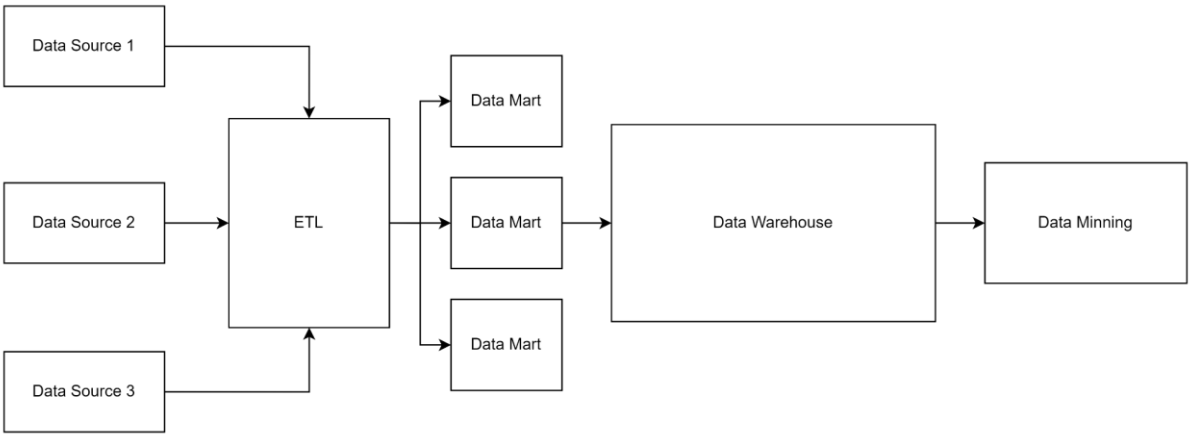

Figure 5: Ralph Kimbal - building Data Warehouse

(source: self-synthesized author group)

2.3.2 Snowflake and Star schemas?

❖ **Star Schema**

The fact table and numerous related dimension tables can be found in the center of the star in a data warehouse's "Star" structure. It is known as a star schema due to its star-like structure. The Star Schema data model is the most fundamental kind of Data Warehouse schema. It is also known as Star Join Schema and is designed specifically for searching through large data collections. Because there are fewer joins, SQL queries are efficient. High data redundancy necessitates additional storage space.

❖ **Snowflake Schema**

By logically arranging the tables in a multidimensional database known as a snowflake schema, a snowflake ER diagram is produced. The snowflake schema adds extra dimensions and expands the star schema. The normalization of the dimension tables causes the data to be split into extra tables. Because there are more joins, SQL queries perform slightly worse than star schema. Data redundancy is modest and takes up less disk space as compared to star schemes.

2.3.3 Who needs Data warehouse and Data mart?

A data mart is limited to a single focus for a certain line of business or a certain part of the business. While a data warehouse, with more data, lasts longer, will be used more granularly across the enterprise and spans many different regions. This means, the data mart is personalized for each department and the users will often be the ones who need to use that department's information for work, while the data warehouse will represent the whole company. and will usually be the new managers or data warehouse managers with access.

2.3.4 Advantages and disadvantages of Data warehouse

Table 1: Advantages and disadvantages of Data warehouse

No.	Advantages of Data Warehouse	Disadvantages of Data Warehouse
1	Provides a consolidated view of data from various sources, enabling better decision-making	High implementation cost
2	Improves data quality and consistency	Requires a lot of time and effort to design and implement
3	Allows for historical data analysis and trend identification	Requires specialized skills and knowledge to maintain
4	Enables faster and more efficient data retrieval and analysis	May require significant changes to existing business processes
5	Facilitates integration of data from disparate systems	May not be suitable for small organizations or those with limited data
6	Supports data mining and advanced analytics	Can result in security and privacy concerns if not properly secured

2.4 KPIs

2.4.1 KPIs Definition

KPIs are measurements that evaluate how effectively an organization is executing its strategic vision. They are quantifiable metrics that reflect the performance of an organization in achieving its goals and objectives. KPIs align all levels of an organization, from business units to departments and individuals, with clearly defined and cascaded targets and benchmarks to create accountability and track progress. KPIs

are strategic value drivers that go beyond measuring non-critical business activities and processes.

A KPI must be a practical tool rather than a discursive one. It should be a fundamental characteristic that forms a type of language to measure the effectiveness of a project and state its chances of success. A KPI can rarely exist alone, and a group of KPIs is usually required. Without agreement on the meaning of KPIs, they would become a source of chaos rather than a source of focused action.

2.4.2 Categories of KPIs

❖ **Sales module**

- RFM score

RFM score is a customer segmentation technique that helps businesses understand the behavior of their customers. The score is based on three factors: Recency (when was the last time the customer made a purchase), Frequency (how often the customer makes purchases), and Monetary Value (how much money the customer spends). The formula for calculating the RFM score is simply the sum of the three individual scores, with higher scores indicating more valuable customers.

$$RFM\ Score = Recency\ Score\ \_\ Frequency\ Score\ \_\ Monetary\ Score \quad (1)$$

Currently there are many methods to determine the RFM score, but in this case, we use the quartile ratio from the separate calculation of the index of each factor (R, F, M) to conduct the scoring. RFM aggregates and groups customers later.

Figure 6: A matrix of the RFM score.

We need to group our customers into different categories based on our company's goals. We think that the best customers for us are the ones who have bought recently, frequently and with enough spending, according to these criteria. We use the quartiles of Recency, Frequency, and Monetary values to create 64 possible combinations of RFM scores and assign each customer segment to one of them based on their similarity and alignment with our technical and business objectives. We give a score from 1 to 4 for each RFM value, where 1 is the best and 4 is the worst. However, we have to reverse the scoring for Recency, because a higher Recency means a lower score. For example, a customer with RFM-score = 111 is the best customer we have, while a customer with RFM-score = 444 is the customer we have lost.

- Customer Churn

Customer Churn is a measure of how many customers stop doing business with a company over a given period of time. High churn rates can indicate that a company is failing to meet customer needs or that there is increased competition in the market. The formula for calculating customer churn rate is

$$\frac{umber\ of\ Customers\ Lost}{Total\ Number\ of\ Customers}\ x\ 100 \qquad (2)$$

- Sales Quota

Total employee sales quota is determined by the total sales of sales staff during a given period. Thanks to this calculation, the company will be able to measure the performance of each salesperson.

❖ **Purchasing model**

- Vendor rejection rates and costs are measures of the quality of goods and services provided by a vendor. It can be calculated by the following formula.

$$Supplier\ rejection\ rate\ and\ cost = \frac{total\ of\ the\ rejected\ product}{total\ of\ received\ product}\ x\ 100 \qquad (3)$$

Here, "total cost of rejected products" refers to the total cost of rejected products due to quality or other problems among the products received from the vendor, and "total cost of received products" refers to the total cost of the product received from the vendor.

- *Supplier Lead Time:* This measures the time it takes for a supplier to deliver goods or services after an order has been placed. The formula to calculate supplier lead time is

$$Supplier\ Lead\ Time = \frac{Delivery\ Date - Order\ Date}{Total\ Number\ of\ Orders} \qquad (4)$$

2.4.3 The advantages and disadvantages of KPIs

❖ **Advantages**

- *Aligns Business Objectives:* KPIs align business objectives with the overall organizational strategy, creating a clear understanding of what needs to be achieved and how it will be measured. It ensures that everyone is operating from the same playbook, reducing confusion and ambiguity.
- *Identifies Opportunities for Improvement:* KPIs help identify areas for improvement by measuring performance against specific targets and benchmarks. They provide valuable insight into areas that need improvement and highlight successes, which can be replicated in other areas of the organization.
- *Improves Decision Making:* KPIs provide accurate and relevant data that can be used to make informed decisions. They help identify trends and

patterns, which can be used to adjust strategies and make data-driven decisions.

- *Facilitates Performance Management:* KPIs create a culture of accountability and track progress towards goals and objectives. They provide a basis for evaluating employee performance, identifying training needs, and coaching opportunities.
- *Accelerates Collaborative Planning:* KPIs facilitate seamless and collaborative planning across the organization by creating a shared understanding of goals and objectives. It ensures that everyone is working towards the same goals, reducing duplication of effort and improving efficiency.

❖ **Disadvantages**

- *Over-reliance on KPIs:* Organizations can become overly reliant on KPIs and use them as the sole measure of success. This can result in a narrow focus on short-term goals, which can be detrimental to the long-term success of the organization.
- *Incorrect KPI selection:* Selecting the wrong KPIs can result in counterproductive behavior and sub-optimized results. Organizations must carefully consider which KPIs are most relevant to their goals and objectives, and ensure that they are measuring what is truly important.
- *Disregarding qualitative data:* KPIs are typically based on quantitative data, which can result in a lack of consideration for qualitative data. Qualitative data can provide valuable insights into the performance of an organization, and should not be disregarded in favor of quantitative data.
- *Unrealistic targets:* Setting unrealistic targets can demotivate employees and result in a lack of buy-in for KPIs. Organizations must ensure that KPIs are achievable and aligned with the capabilities of their employees and resources.

2.5 Cloud Service

2.5.1. Cloud Services Definition

Cloud services, also known as cloud computing services, are a collection of resources and applications that are delivered to users over the Internet. According to the National Institute of Standards and Technology (NIST), cloud services are "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and

services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell & Grance, 2011).

In simpler terms, cloud services refer to computing resources that can be accessed and used over the internet, instead of being stored or processed on local computers or servers. These resources and applications can include data storage, computing power, and software applications, among others. Cloud services are fully managed by cloud computing vendors and service providers, and they are made available to customers from the providers' servers, eliminating the need for a company to host applications on its own on-premises servers.

Cloud services are typically provided by cloud service providers (CSPs), who own and operate the infrastructure and resources necessary to deliver cloud services to users. Some of the most well-known CSPs include Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

2.5.2. Types of Cloud Services

There are three main types of cloud services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

IaaS provides users with access to computing resources such as virtual machines, storage, and networking infrastructure. With IaaS, users have the flexibility to create and manage their own virtual machines, configure their own networking, and install their own operating systems and applications. IaaS is a great option for companies that need to quickly scale their infrastructure up or down based on demand.

PaaS provides users with a platform for developing, testing, and deploying applications, without having to manage the underlying infrastructure. With PaaS, users can focus on developing and deploying applications, while the platform takes care of the underlying infrastructure. PaaS is a great option for developers who need to quickly develop, test, and deploy applications.

SaaS provides users with access to software applications that are hosted and managed by a third-party provider. With SaaS, users do not need to worry about installing, maintaining, or upgrading software applications on their own systems. SaaS is a great option for companies that need access to software applications but do not want to manage the underlying infrastructure.

2.5.3. Benefits of Cloud Services

Cloud services have become increasingly popular in recent years due to their flexibility, scalability, and cost-effectiveness. One of the biggest benefits of cloud

services is that they allow users to access computing resources and applications from anywhere in the world, as long as they have an internet connection. This makes it easy for companies to offer remote work options and allows employees to work from anywhere in the world.

Cloud services also offer scalability, which means that users can quickly scale their infrastructure up or down based on demand. This makes it easy for companies to handle spikes in traffic or demand without having to invest in additional infrastructure.

Cloud services are cost-effective because users only pay for what they use. This eliminates the need for companies to invest in expensive infrastructure that they may not use to its full capacity. Additionally, because cloud services are fully managed by cloud computing vendors and service providers, users do not need to worry about the costs associated with maintaining their own infrastructure.

CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION

3.1 Business processes

3.1.1 Sales department

❖ **The purpose of Sales**

A company's sales division is in charge of selling its products and services. A salesperson works with others to close agreements, boost earnings, and build and maintain relationships with clients to encourage repeat business and brand loyalty. To market products and services to both present and potential customers, look for business opportunities. Make the customer an engaging presentation about the products or services they are seeking. Keep in touch with clients frequently to understand their wants and needs. Set annual budget objectives and monitor quarterly budget implementation rates. The business opportunity for each client should be taken into account when creating the target budget or forecast list. For the purpose of learning about their future objectives, develop relationships with all present and potential customers.

❖ **Sales process**

Sales activities of AventureWorks include Two sales channels, including: online through distribution channels and direct sales via the Internet. Bicycles, clothing and accessories & parts are the three main product categories of AWC's online sales. Sells bikes under three major brands, including mountain bikes, road bikes, and touring bikes.

Multiple sales territories were given to sales representatives from regional sales offices. A team manager and multiple sales agents work in each regional office.

For sales channels through distribution channels, the sales process goes through four steps: Order processing → Packing, delivery of goods → Invoicing, debit recognition → Payment .

For the Internet sales channel, through the company's website, there are 2 main steps: Order, payment → Delivery, invoice.

3.1.2 Purchasing Department

❖ **The purpose of Purchasing**

Planning high-level procurement activities is part of the purchasing department's business purpose. The goal of the strategic purchasing department is to make it easier to source products strategically (through e-sourcing) at competitive prices and high

standards. The decision between buying from external vendors and internal production is decided during the strategic purchasing phase.

❖ **Purchasing process**

The purchasing division at Adventure Works Cycles purchases the raw materials and components needed to make the company's bicycles.

In addition, Adventure Works Cycles purchases goods for resale such as cycling accessories like water bottles and pumps as well as clothing for cyclists. The AdventureWorks sample database contains details about these goods and their providers.

3.2 Data source and challenges

3.2.1 Data Source

Transactional Databases: The AdventureWorks 2019 database itself is a good source of data for sales and purchasing departments, as it contains information on orders, customers, products, and suppliers.

The schemas used in the AdventureWorks database are shown in the following Table 2:

Table 2: AdventureWorks 2019 database

Schema	**Description**	**Number of tables**
Human Resources	Employees of the company Adventure Works Cycles.	6
Person	Names and addresses of customers: individual customers, suppliers and employees.	13
Production	Products manufactured and sold by Adventure Works Cycles.	25
Purchasing	Suppliers of products that the company buys.	5
Sales	Customers and data related to the purchase.	18

3.2.2 Challenges

3.3. Business Requirements Analysis (Purchasing, Production, Sales, or HR)

3.3.1. Sales Requirements Analysis

Involve identifying the data needs of the Sales department to support our operations and decision-making. This includes identifying the sources of data, such as sales transactions, customer interactions, and marketing campaigns. It also involves determining the types of data that are needed, such as customer demographics, sales performance metrics, and inventory levels. Finally, Sales Requirements Analysis involves identifying the key performance indicators (KPIs) that will be used to measure sales performance and track progress toward sales targets.

3.3.2. Purchasing Requirements Analysis

Involve identifying the data needs of the Purchasing department to support our operations and decision-making. This includes identifying the sources of data, such as purchase orders, vendor contracts, and inventory levels. It also involves determining the types of data that are needed, such as vendor performance metrics, purchase history, and inventory turnover. Finally, Purchasing Requirements Analysis involves identifying the key performance indicators (KPIs) that will be used to measure purchasing performance and track progress toward purchasing targets.

3.4. IT Requirements Analysis (IT & Infrastructure)

- **Define the objectives and scope of the BI solution:** Identify the business goals that the BI solution should help achieve, and define the scope of the solution in terms of data sources, data models, and analytics requirements.
- **Identify technical requirements:** Define the technical infrastructure requirements necessary to support the BI solution, including the hardware and software needed to store and process data. We considered the Azure cloud infrastructure requirements necessary for the solution.
- **Analyze data sources:** Identify the data sources needed to support the BI solution and determine how the data will be integrated into the cloud-based data warehouse. This may involve performing data profiling and data quality assessments to ensure the data is accurate and complete.
- **Design data models:** Develop a logical data model that can be used to structure the data in the cloud-based data warehouse. This model should be based on the data sources identified in step 3.

- **Determine reporting and analytics requirements:** Define the reporting and analytics requirements for the BI solution. This can include identifying the types of reports and visualizations needed, as well as the tools required to create and distribute these reports.

CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

4.1 Designing Data Warehouse

4.1.1 Bus Matrix

Table 3: Bus Matrix

	COMMON DIMENSIONS							
BUSINESS PROCESSES	Date	Products	Customers	Employees	Vendors	Shipping	Location	Promotion
Issues Purchase Orders	x	x		x	x			
Supply management	x	x			x			
Sales Quota management	x			x				
Vendors Evaluation	x	x			x	x	x	
Promotion Tracking	x	x	x					x
Sales Forecasting	x	x	x	x	x	x	x	x

4.1.2 Master Data

For Sales module, we got a list of necessary master data as follows:

- *Product data:* contains information about the products that are sold, such as the product name, subcategory, category, stock quantity and other details for the price of each product.

- *Customer data:* contains information about the customers who purchase the products, such as their name, address, contact information, etc. In some cases, customers may not have done any transactions with the firm.
- *Time data:* contains information about the time dimension, such as the date, month, year, etc. This is a calendar recording of any datetime entities.
- *Promotion data:* represents the specific information of promotion applying to any orders. This kind of data may be useful for marketing evaluation later.
- *Employee data:* persons are responsible for making customer orders. Generally, the company will evaluate the work performance of these salespeople and treat them with appropriate policies.

Similarly, Fact Purchasing also includes Time, Employee, and Product product dimensions. On the other hand, this event table will contain vendors' dimensions to store vendor personal information relating to evaluating and beneficial for making orders decision.

4.1.3 Transaction Data

Sales transaction data: contains information about the individual sales transactions, such as the product sold, the customer who purchased the product, the quantity sold, the price per unit, the date of sale, etc.

Purchase transaction data: contains information about the individual purchase transactions, such as the product purchased, the vendor from whom the product was purchased, the quantity purchased, the price per unit, the date of purchase, etc

4.1.4 Fact and dimension tables

- DimCustomers

Table 4: DimCustomers

Column name	Datatype	Description	Key
CustomerKey	int	Unique key	PK
CustomerID	int	Natural Key	
FirstName	nvarchar(50)	First name of the person.	
LastName	nvarchar(50)	Last name of the person.	
EmailAdress	nvarchar(50)	Contact email	
EmailPromotion	int	0 = Contact does not wish to receive e-mail	

		promotions, 1 = Contact does wish to receive e-mail promotions from AdventureWorks, 2 = Contact does wish to receive e-mail promotions from AdventureWorks and selected partners.	
PhoneNumber	int	Telephone number identification number.	
TerritoryName	nvarchar(50)	Customer territory located	
CountryRegion Code	nvarchar(3)	Country Code	
CountryRegion Name	nvarchar(50)	Country name	
TransactionFlag	bit	0 = had not recorded transaction, 1 is reversed.	

- DimEmployees

Table 5: DimEmployees

Column name	**Datatype**	**Description**	**Key**
EmployeeKey	int	Unique key	PK
EmployeeID	int	Natural Key	
NationalIDNumber	nvarchar(15)	National ID number of the employee	
LoginID	nvarchar(256)	Login ID number of the employee	
JobTitle	nvarchar(50)	Employee’s job title	
BirthDate	date	Employee’s birth day	
Gender	nvarchar(1)	Employee's gender Rule: M = Male, F = Female	
CurrentFlag	bit	Rule: 0 = Inactive, 1 = Active	
HireDate	date	Employee’s hire date	

SalariedFlag	bit	Employee's salaried flag Rule: 0 = Yes, 1 = No	
VacationHours	smallint	Employee's hours spending vacation	
SickLeaveHours	smallint	Employee's hours sick leave	
CountryRegionCode	nvarchar(3)	Country Code	
CountryRegionName	nvarchar(50)	Country Name	
SalesPersonFlag	bit	Define whether an employee is a salesperson. Rule: 0 = False, 1 = True	

- DimProducts

Table 6: DimProducts

Column name	**Datatype**	**Description**	**Key**
ProductKey	int	Unique key	PK
ProductID	int	Natural Key	
Name	nvarchar(50)	English name of product	
FinishedGoodsFlag	bit	0 = Finished goods is False, 1 = Finished goods is True	
Color	nvarchar(15)	Product' Color	
SafetyStockLevel	smallint	Minimum inventory quantity.	
ReorderPoint	smallint	Inventory level that triggers a purchase order or work order.	
StandardCost	money	Standard cost of the product.	
ListPrice	money	Selling price.	
Size	nvarchar(5)	Product size	

Weight	decimal(8,2)	Product weight.	
DaysToManufacture	int	Number of days required to manufacture the product.	
ProductLine	nvarchar(2)	R = Road, M = Mountain, T = Touring, S = Standard	
Class	nvarchar(2)	H = High, M = Medium, L = Low	
Style	nvarchar(2)	W = Womens, M = Mens, U = Universal	
ProductSubcategoryName	nvarchar(50)	Name of product's subcategory	
ProductCategoryName	nvarchar(50)	Name of product's category	

- DimShipMethods

Table 7: DimShipMethods

Column name	**Datatype**	**Description**	**Key**
ShipMethodsKey	int	Unique key	PK
ShipMethodsID	int	Natural Key	
Name	nvarchar(50)	English name of product	
ShipBase	money	Minimum shipping charge. Default: 0.00	
ShipRate	money	Shipping charge per pound. Default: 0.00	

- DimVendors

Table 8: DimVendors

Column name	**Datatype**	**Description**	**Key**

VendorsKey	int	Unique key	PK
VendorsID	int	Natural Key	
Name	nvarchar(50)	English name of company	
CreditRating	tinyint	1 = Superior, 2 = Excellent, 3 = Above average, 4 = Average, 5 = Below average	
PreferredVendorStatus	bit	0 = Do not use if another vendor is available. 1 = Preferred over other vendors supplying the same product. Default: 1	
CountryRegionCode	nvarchar(3)	Country Code	
CountryRegionName	nvarchar(50)	Country Name	
City	nvarchar(30)		

- DimPromotion

Table 9: DimPromotion

Column name	**Datatype**	**Description**	**Key**
PromotionKey	int	Unique key	PK
PromotionID	int	Natural Key = SpecialOfferID	
Description	nvarchar(255)	English name of product	
DiscountPct	smallmoney	Discount precentage. Default: 0.00	
Type	nvarchar(50)	Discount type category	

Category	nvarchar(50)	Group the discount applies to such as Reseller or Customer	

- DimSalesTerritory

Table 10: DimSalesTerritory

Column name	Datatype	Description	Key
TerritoryKey	int	Unique key	PK
TerritoryID	int	Natural Key	
Name	nvarchar(50)	English name of product	
CountryRegion Code	nvarchar(3)	Country Code	
Group	nvarchar(50)	Geographic area to which the sales territory belong	
SalesYTD	money	Sales in the territory year to date. Default: 0.00	
SalesLastYear	money	Sales in the territory the previous year. Default: 0.00	
CostYTD	money	Business costs in the territory year to date. Default: 0.00	
CostLastYear	money	Business costs in the territory the previous year. Default: 0.00	

- DimDate

Table 11: DimDate

Column name	Datatype	Description	Key
DateSK	int	Unique key	PK

FullDateAlternateKey	date		
DayNumberOfWeek	tinyint		
DayNameOfWeek	nvarchar(10)		
DayNumberOfMonth	tinyint		
MonthName	nvarchar(10)		
MonthNumberOfYear	tinyint		
CalendarQuarter	tinyint		
CalendarYear	smallint		

- FactSales

Table 12: FactSales

Column name	**Datatype**	**Description**	**Key**
SalesKey	int	Unique key	PK
OrderKey	int	Natural Key	
ProductKey	int	ProductKey	
CustomerKey	int		
SalesPersonKey	int		
TerritoryKey	int		
ShipMethodKey	int		
PromotionKey	int		
DateTimeKey	int		

OrderDate	datetime		
DueDate	datetime		
ShipDate	datetime		
Status	bit	Order current status. 1 = In process; 2 = Approved; 3 = Backordered; 4 = Rejected; 5 = Shipped; 6 = Cancelled Default: 1	
OnlineOrderFlag	bit	0 = Order placed by sales person. 1 = Order placed online by customer. Default: 1	
Freight	money	Shipping cost. Default: 0.00	
TaxAmt	money	Tax amount. Default: 0.00	
OrderQty	smallint	Order quantity of product in one order	
UnitPriceDiscount	money	unit price discount of product in one order	
TotalProductCost	money	Sales subtotal. Computed as SUM(SalesOrderDetail.LineTotal)for the appropriate SalesOrderID. Default: 0.00	
TotalPaid	money	Total due from customer. Computed as Subtotal + TaxAmt + Freight. Computed: isnull(([SubTotal]+[TaxAmt])+[Freight],(0))	

- FactPurchasing

Table 13: FactPurchasing

Column name	**Datatype**	**Description**	**Key**

PurchaseKey	int	Unique key	PK
DateTimeKey	int	Natural Key	
PurchaseOrderKey	nvarchar(50)	English name of product	
EmployeeKey	int	Employee ID	
ShipMethodKey	int	Ship Method ID	
VendorKey	int	Vendor ID	
ProductKey	int	Product ID	
UnitPrice	money	Vendor's selling price of a single product	
OrderQty	smallint		
ReceivedQty	decimal(8, 2)	Quantity actually received from the vendor	
RejectedQty	decimal(8, 2)	Quantity rejected during inspection	
AverageLeadTime	int		
LastReceiptDate	datetime	Date the product was last received by the vendor.	
LastReceiptCost	datetime	The selling price when last purchased.	
StandardPrice	money	The vendor's usual selling price	
MinOrderQty	int	The minimum quantity that should be ordered.	
MaxOrderQty	int	The maximum quantity that should be ordered.	

OrderDate	datetime	Dates the sales order was created. Default: getdate()	
ShipDate	datetime	Date the order was shipped to the customer	
Status	tinyint	Order current status. 1 = Pending; 2 = Approved; 3 = Rejected; 4 = Complete Default: 1	
TaxAmt	money	Tax amount. Default: 0.00	
Freight	money	Shipping cost. Default: 0.00	
TotalDue	money	Total due to vendor. Computed as Subtotal + TaxAmt + Freight. Computed: isnull(([SubTotal]+[TaxAmt])+[Freight],(0))	

4.1.5 Data Warehouse Model

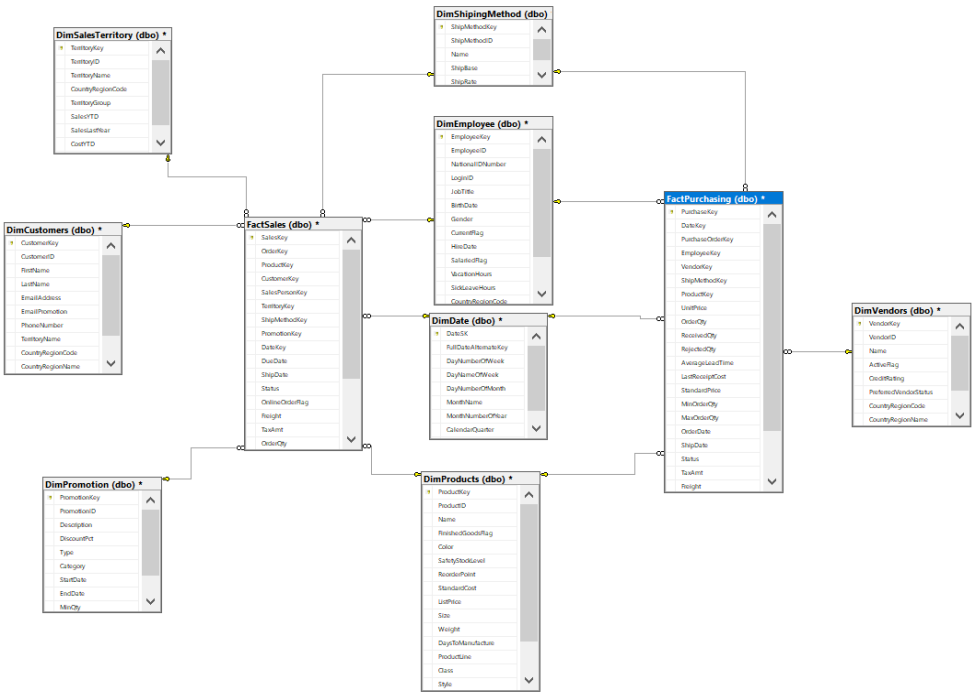

Figure 7: Data warehouse Model

4.2 ETL processes

Before doing the data ETL process, we create a database of empty tables called destination database as Figure 8.

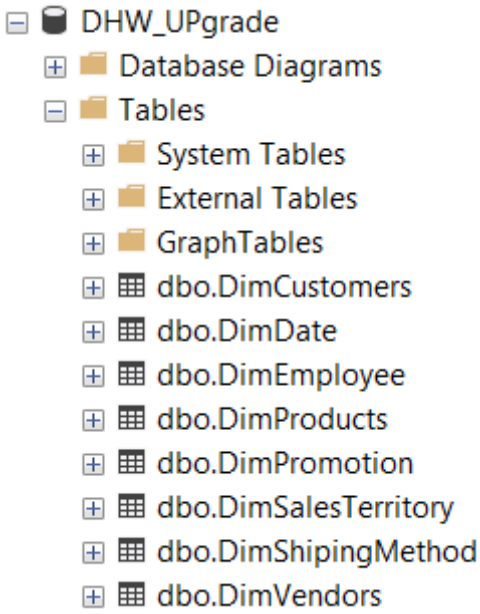

Figure 8: Destination database

4.2.1 Dimension Table's ETL Process

First, we create a new pipeline, after renaming, we select the component "**Copy data**" and define the data source from Azure SQL Database which we uploaded to the AdvantureWork2019 database from earlier. In this step, we use query statements to get data from the source database, following Figure 9 below.

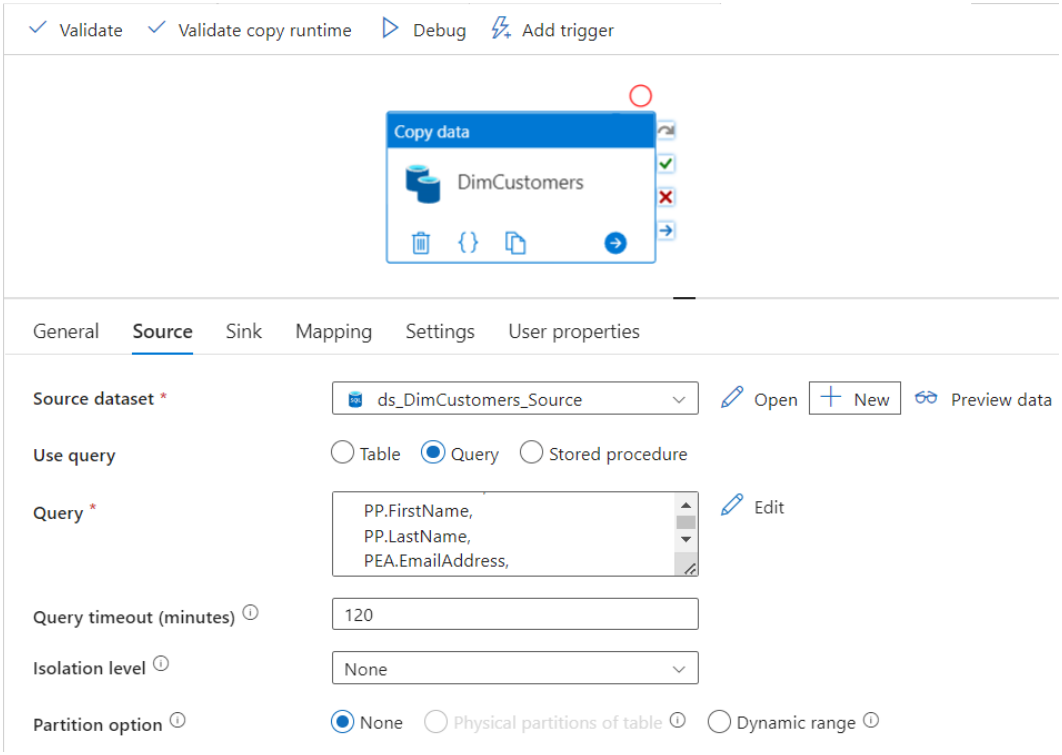

Figure 9: Define data source.

Next, we go on to define the destination as the empty dimension table that was created earlier in our data warehouse, is showed in Figure 10.

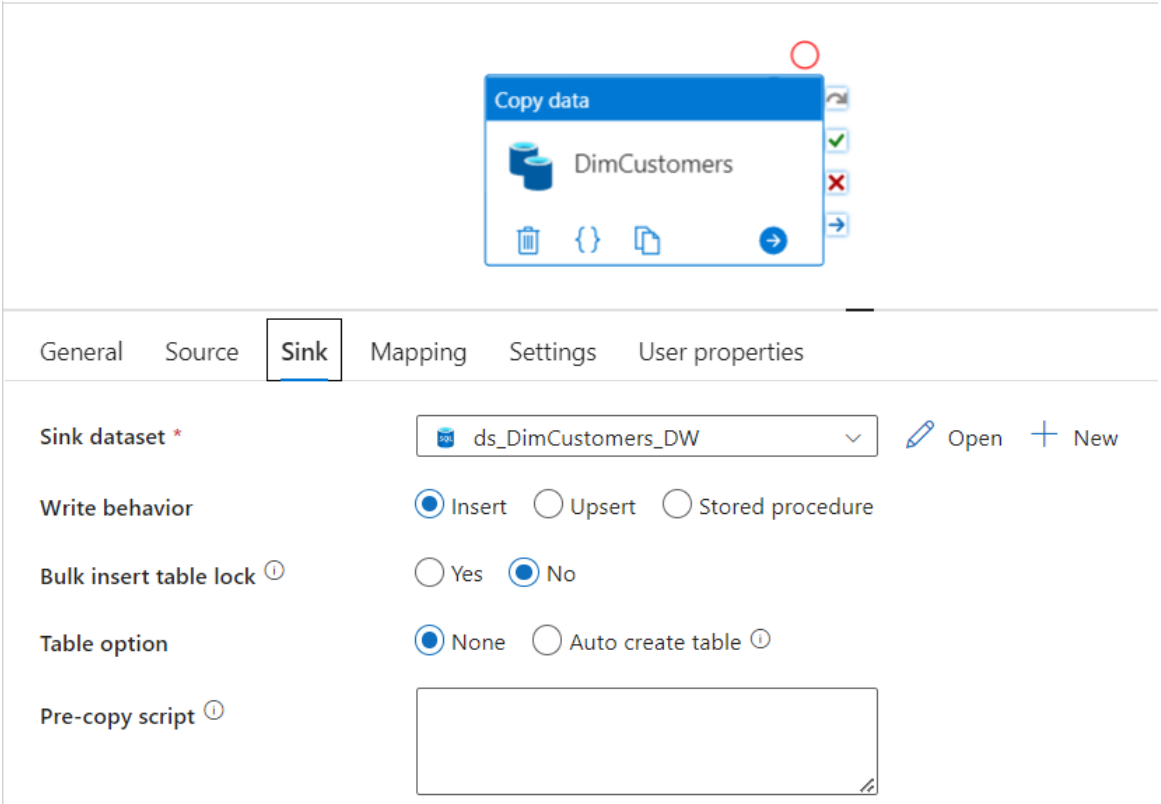

Figure 10: Define destination.

After defining the source and destination data, we proceed to map the columns between the source and destination data. Once the above steps have been completed, we press the "**Publish all**" button and finally the "**Trigger**" button and select "**Trigger now**".

The other tables do the same process as above, but the query statements for the source data are different for each table as Figure 11.

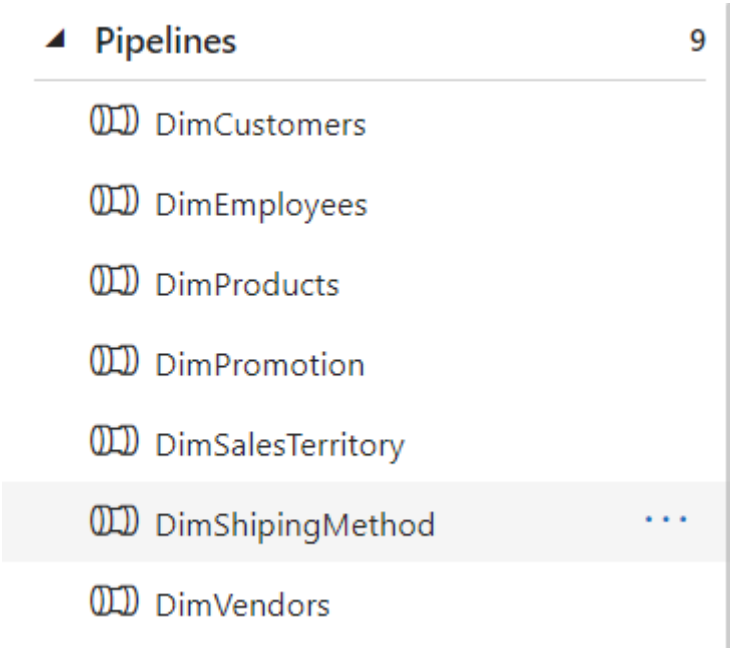

Figure 11: Pipelines of Dimension Tables.

4.2.2 Fact Table's ETL Process

Generate an ADF data pipeline for the Fact table. We initially create a pipeline, which is subsequently moved by utilizing the Data Flow activity, following Figure 12.

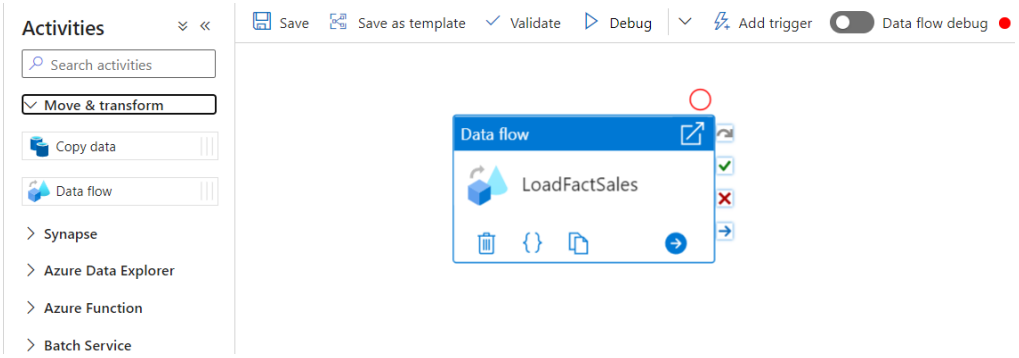

Figure 12: Load Fact table

Look at the Figure 13, each Fact table includes source data and a lookup with associated dimensions tables. Here are the parts of the ETL process a Fact table.

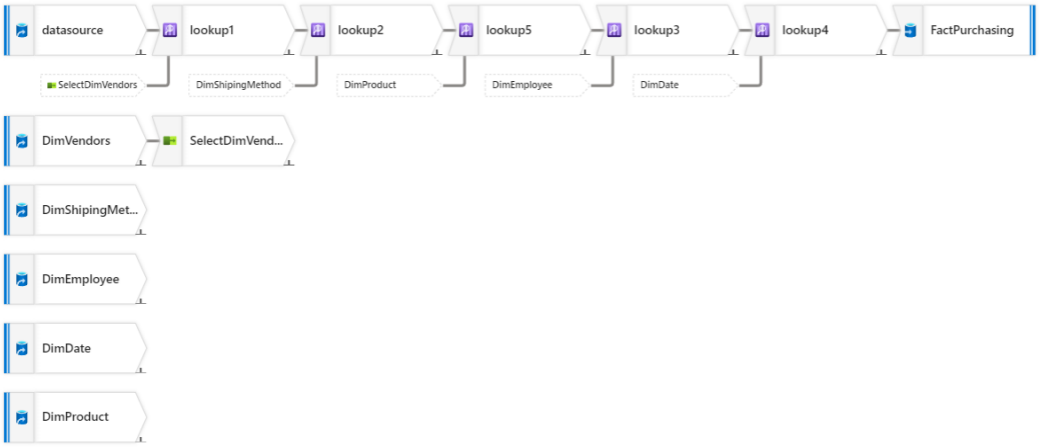

Figure 13: Dataflow of the Process Load Fact table.

First, we create an empty table similar to the destination table for the Dimension table. Next, we query data from the data source and simultaneously retrieve data from the corresponding Dimension tables combined together at the lookup component. We repeat this process for all the Dimension tables. Finally, we add a destination for the data from the sources. After these steps, we proceed to map the data columns to match the data in the destination table. Now, perform "Publish all" and "Trigger Now" for the data flow to execute the ETL process. Repeat the same steps for the remaining Fact tables.

CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION

5.1. Report and dashboard systems (structure)

5.1.1. Sales module

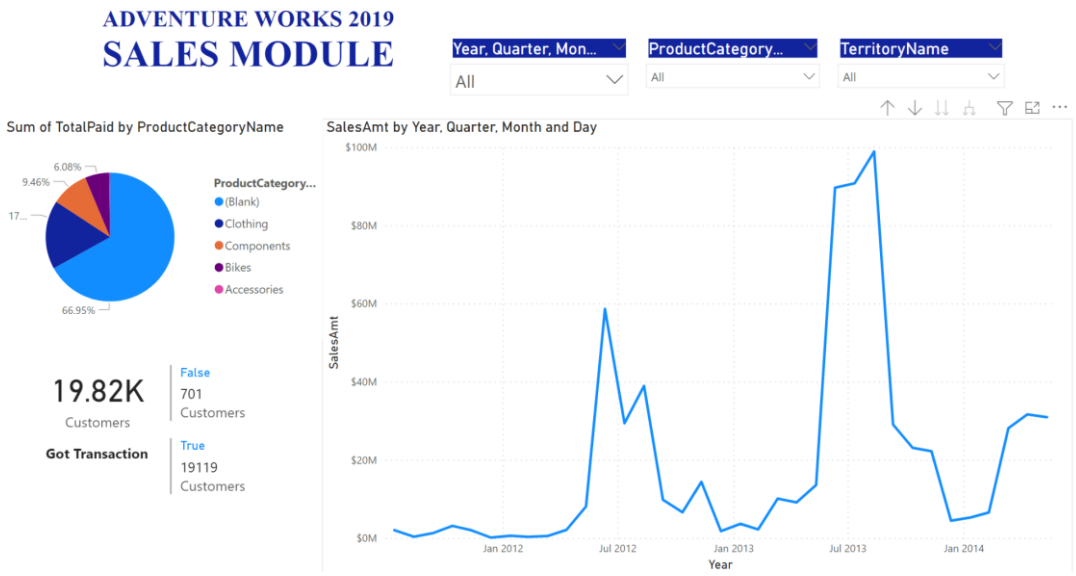

Figure 14: Overall of Sales.

The dashboard shows the company's revenue from year to year (2011-2014) and management can track the specific revenue of the selected period with factors such as Product Category and SalesTerritory.

It can be seen that there are more than 19,8 thousand customers in total but our firm got 701 customers, who did not do any transactions and recorded various null records. Therefore, our RFM model will remove these customers.

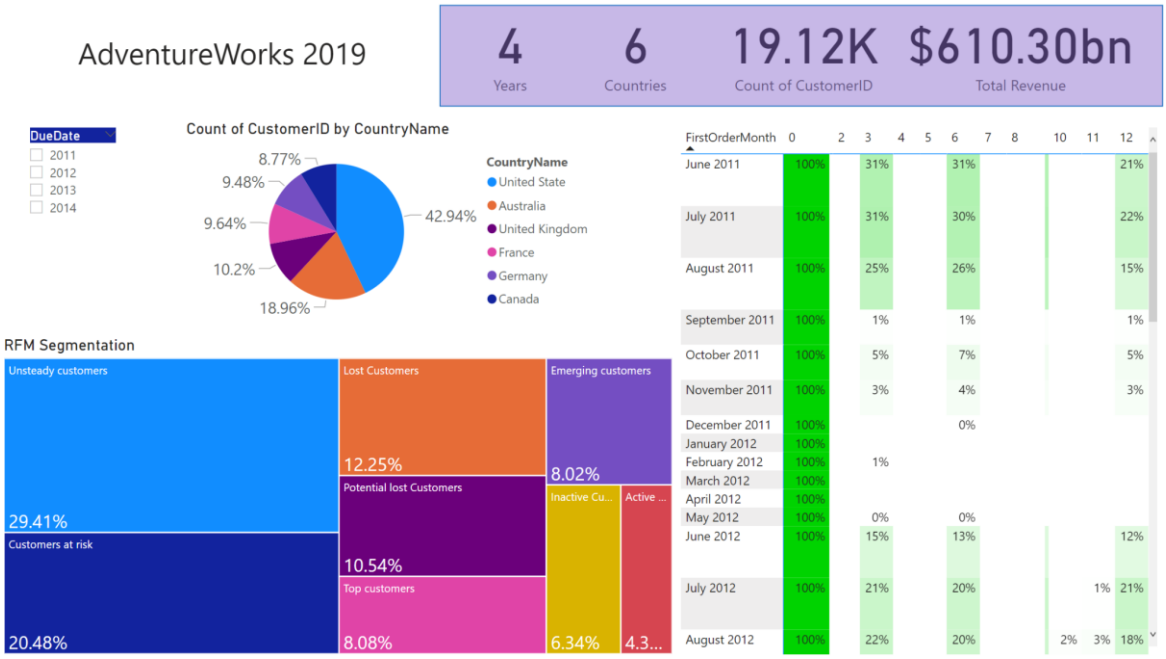

Figure 15: Customers' KPI dashboard.

The map tree chart shows 8 different groups based on the RFM score. The Best Customers group has the lowest percentage (8.80%), which means that this group has very high standards and scores very well on all 3 factors. On the other hand, the Active Customers group is the most common one in the segmentation, with the highest percentage at 29.41%.

5.1.2. Purchase module

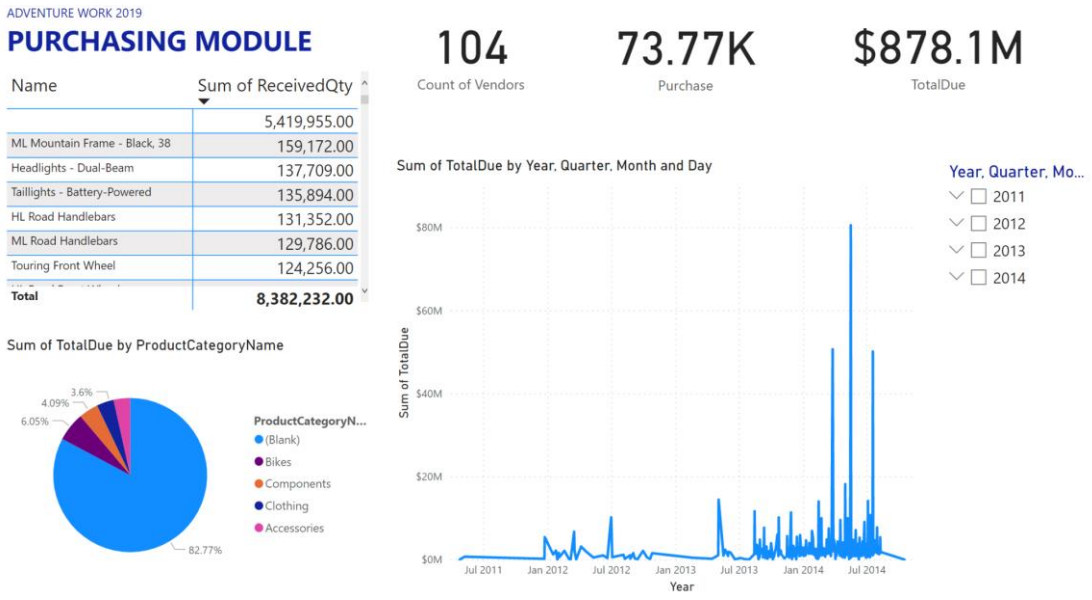

Figure 16: Overall of Purchasing.

The Dashboard of the company's Purchasing process shows that from 2011 to 2014, the company currently has 104 suppliers with nearly 74,000 orders transacted.

Which, accessory products are the category with the lowest purchase expenditure (3.49%). The next position belongs to Clothing, Components, Bike, and others, respectively.

5.2 Time series and Forecasting or Predictive model with Python and Power BI

For forecasting time series by date of revenue and number of orders, We first identify trend, seasonality, and residuals, which are the three main components of a time series. The trend is the long-term behavior of the data, seasonality is the regular pattern of fluctuations at fixed intervals, and residuals are the random fluctuations that cannot be explained by the trend or seasonality. These components can be identified visually from a time series plot, where the trend appears as a long-term direction, seasonality as repeating patterns, and residuals as the fluctuations around the trend and seasonal patterns. To identify each component from a plot, we can use a time series decomposition plot, which shows the original time series data, the trend, the seasonality, and the residuals.

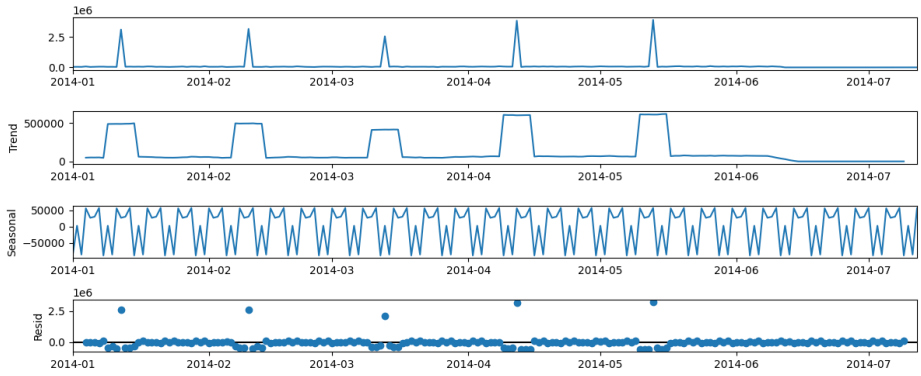

Figure 17: Decomposition of Total Revenue

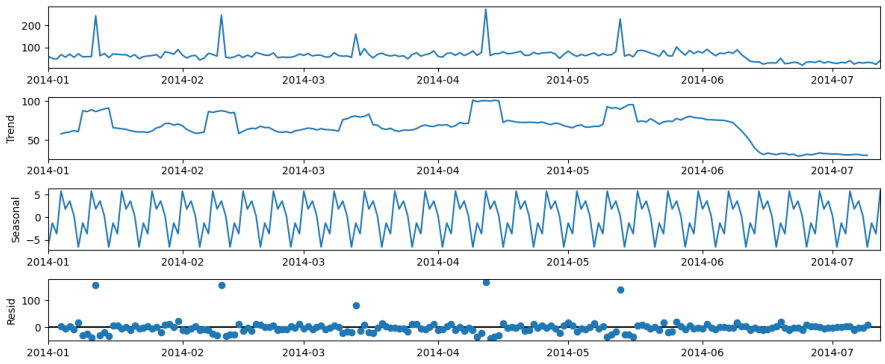

Figure 18: Decomposition of Number of orders

From 2 plots above, we can see that trend tends to be very stable, and only rises very high in the first days of the month, stabilizes for a few days and drops very low again in the middle of the month, and the series is also cyclical, increasing and decreasing with the rule of the week. So that we can use **Power BI** (which is based on

built-in predictive forecasting models using exponential smoothing), which is to assist with time series forecasting.

With the time intelligence feature in Power BI, it can generate custom time periods, such as weeks or months, and summarize data according to those periods. This can help in analyzing data over time and recognizing patterns and trends.

We set the confidence interval to 95% means that the forecasting algorithm will create a prediction range around the forecasted values with a confidence level of 95%. And set the seasonality to "Auto", so the algorithm will automatically detect the seasonality pattern in the data and adjust the model accordingly. By detecting and incorporating seasonality patterns in the model, the algorithm can produce more accurate forecasts that take into account the expected changes in the data over time.

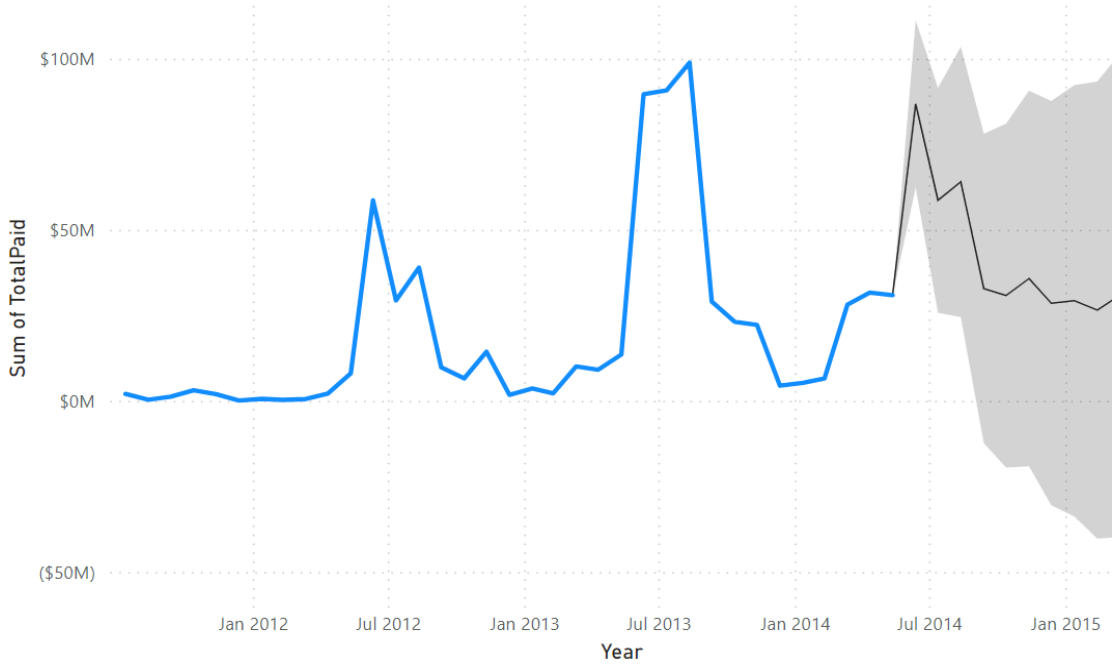

Figure 19: . A 10-point length forecasting for Sales amounts

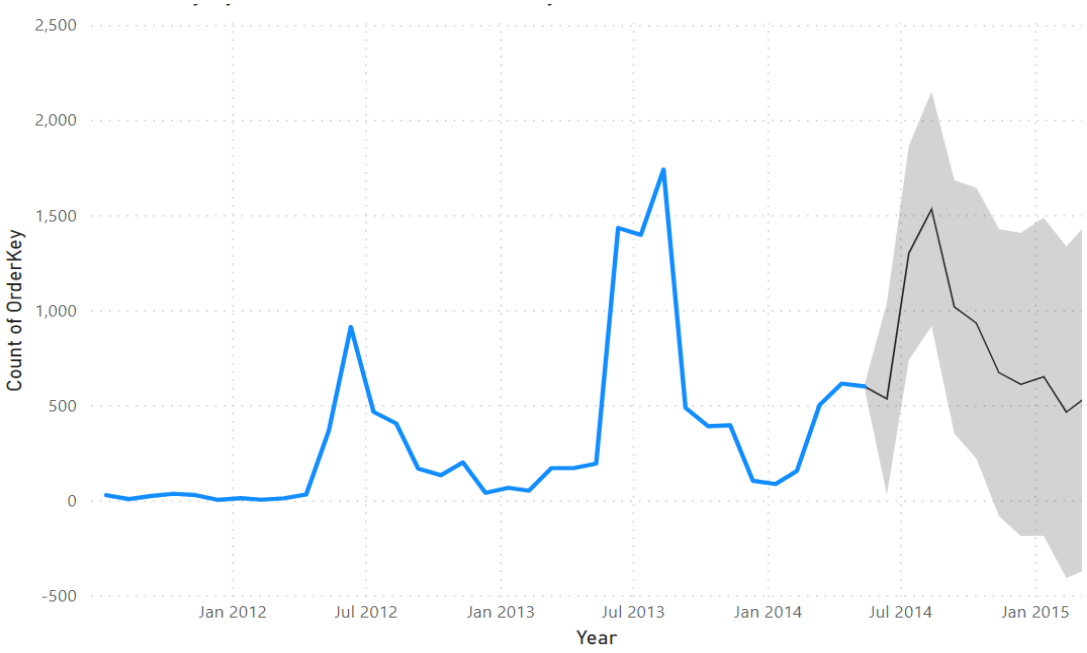

Figure 20: A 10-point length forecasting for sales orders' number.

Two above plots that show the prediction curve has a period that is quite similar to the past data. By forecasting with time series, we can predict sales and total orders in the near future (within 1 year) so that we can be well prepared to be able to grow sales.

5.3. Evaluation and Discussion

Furthermore, we aim to be able to conduct in-depth research on the proposed future value prediction sector in the near future, because, as previously said, prediction is extremely difficult. It is vital to study, assess, and precisely pick the parts inside the model in order for it to be truly applicable in practice.

The chapter describing the development of a cloud-based BI solution for the Adventure Work 2019 database provides a thorough account of the project and its outcomes. The chapter focuses on three main areas: data analytics and visualization, Power BI-based report and dashboard systems, and time series forecasting or predictive modeling.

The project's results are presented in detail, highlighting the solution's key features. The report and dashboard systems built with Power BI offer a comprehensive overview of the sales and purchase modules, with detailed metrics and charts that provide valuable business insights. Additionally, the chapter delves into time series forecasting or predictive modeling, which provides decision-makers with a glimpse into the business's future performance and aids in informed decision-making.

The project evaluation reveals that the cloud-based BI solution is effective in providing an accurate and detailed view of the business. The report and dashboard

systems are user-friendly and provide access to a wealth of information that can be easily analyzed. The time series forecasting and predictive models are valuable tools for decision-makers, enabling them to plan and prepare for the future.

Overall, the chapter demonstrates the benefits of a cloud-based BI solution and its ability to enhance decision-making by providing real-time data and powerful analytics tools. The successful implementation of a cloud-based BI solution for the Adventure Work 2019 database highlights the significance of investing in data analytics and visualization tools.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1. Results

The primary objective of the project was to construct a cloud-based data warehouse by extracting data from the AdventureWork2019 database, intended for utilization in business intelligence (BI) solutions. The data warehouse is structured from data marts that contain information sourced from the sales and purchasing modules, thereby meeting the specifications related to the sales and purchase processes. Additionally, we created a dashboard utilizing Power BI to visualize the data, providing comprehensive insights into the performance of the business.

6.2. Limitations

Due to time constraints and limited experience in building a data warehouse, the data warehouse we built is still small and has some non-optimal and inefficient points. Moreover, the AdventureWorks database is complex and has several modules that we do not fully understand. Although we have checked it thoroughly, there might still be some shortcomings in our project, and we are open to feedback and suggestions for improvement.

6.3. Future works

In the future, we plan to expand the data warehouse to include other modules such as Person and Production. We also aim to optimize and improve the data warehouse further. Additionally, we will continue to develop the dashboard to provide more comprehensive visualizations and insights into the business's performance. Overall, our project lays a strong foundation for building a robust cloud-based data warehouse and BI solution using the AdventureWork2019 database.

REFERENCES

[1] Microsoft. Adventure Works Cycles Business Scenarios. Retrieved 4/5/2023 from: https://learn.microsoft.com/en-us/previous-versions/sql/sql-server 2008/ms124825(v=sql.100)

[2] Gartner. Analytics and Business Intelligence (ABI). Retrieved 4/5/2023 from: https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi

[3] Microsoft. What is business intelligence?. Retrieved 4/5/2023 from:https://powerbi.microsoft.com/en-cy/what-is-business-intelligence/

[4] Techtarget. business intelligence (BI). Retrieved 4/5/2023 from: https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI

[5] Wu, L., Barash, G., & Bartolini, C. (2007, June). A service-oriented architecture for business intelligence. In IEEE international conference on service-oriented computing and applications (SOCA'07) (pp. 279-285). IEEE.

[6] Ong, I. L., Siew, P. H., & Wong, S. F. (2011). A five-layered business intelligence architecture. Communications of the IBIMA, 2011, 1-11.

[7] IBM. What is ETL?. Retrieved 5/5/2023 from: https://www.ibm.com/topics/etl

APPENDIX

PROJECT MEMBER EVALUATION (Signed by all members)