UNIVERSITY OF ECONOMICS AND LAW
FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

BUSINESS INTELLIGENCE
&
DECISION SUPPORT SYSTEM

PROJECT:
BUILDING A DATA WAREHOUSE ON AZURE PLATFORM

Lecturer: Nguyen Văn Ho
Le Ba Thien

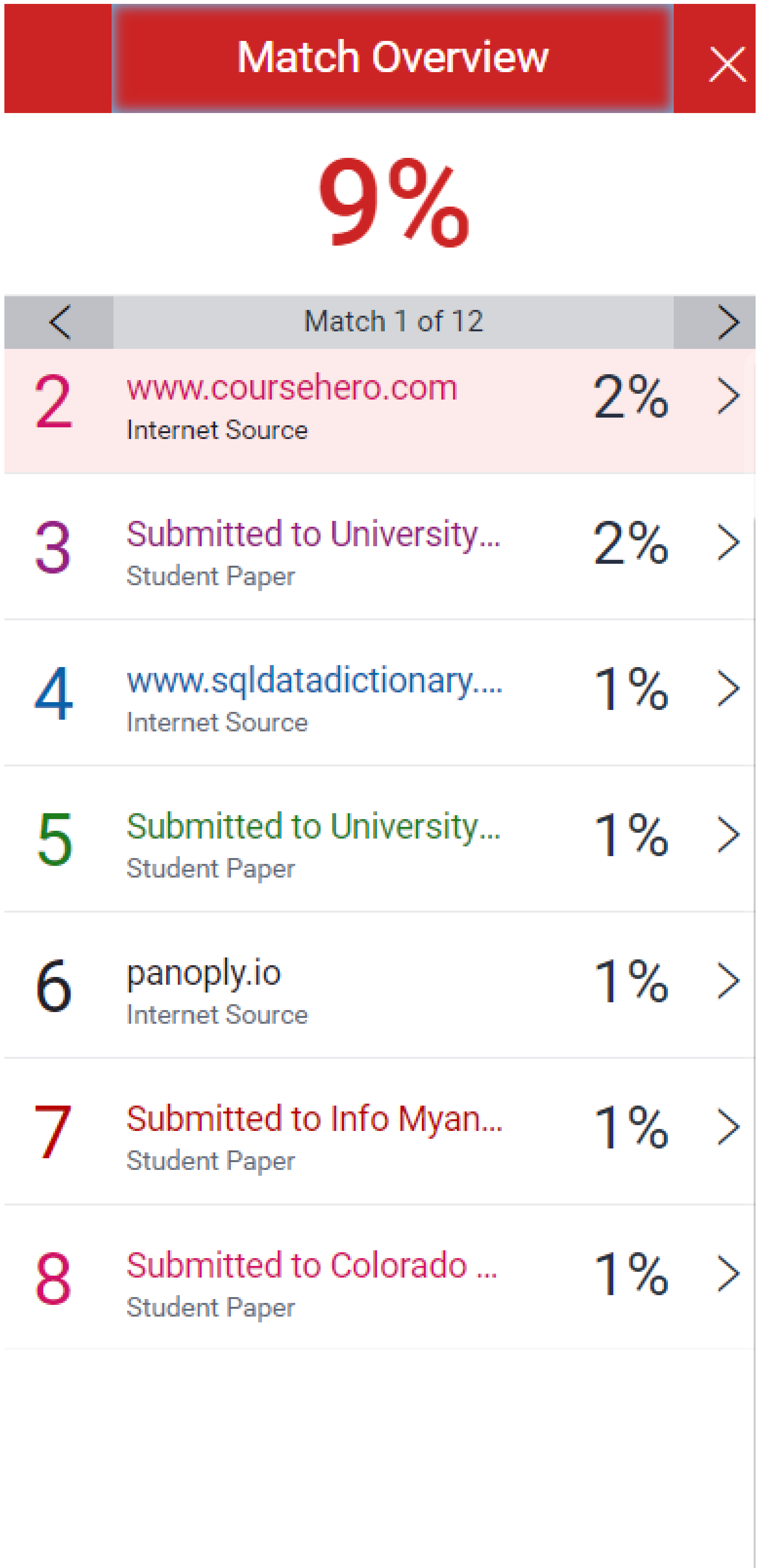
Group: Da Vu Tru

Ho Chi Minh City, May 3rd, 2023

ACKNOWLEDGEMENT

First of all, Mr. Nguyen Van Ho and Mr. Le Ba Thien deserve our team's profound gratitude. The team learned and researched new facts while doing the research, applying theoretical knowledge to real-world situations. Utilize all available information from the research effort to create the best project report. The team tried to avoid errors during the implementation process, but owing to the short time constraint and lack of in-depth knowledge, they were unable to. As a consequence, the group is eager to receive feedback from the professors on how to best build on the knowledge it has already learned and aid in the group's preparation for upcoming assignments.

COPYRIGHTS


Match Overview
9%
Match 1 of 12
2 www.coursehero.com Internet Source 2%
3 Submitted to University... Student Paper 2%
4 www.sqldatadictionary.... Internet Source 1%
5 Submitted to University... Student Paper 1%
6 panoply.io Internet Source 1%
7 Submitted to Info Myan... Student Paper 1%
8 Submitted to Colorado ... Student Paper 1%

MEMBERS OF THE PROJECT

Number	Name	ID	Role	Contribution Rate
1	Le Chi Hao	K204161988	Leader	100%
2	Bui Thu Van	K204162007	Member	100%
3	Tran Trieu Vy	K204162011	Member	100%
4	Truong Cong Vinh	K204162008	Member	100%

Table of Contents

LiST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Business case for the project

Bicycles made of metal and composite are produced and sold by a corporation in North America, Europe, and Asia. The corporate office is located in Bothell, Washington. Managing inventory and manufacturing becomes an urgent problem when there are many items, many different types of products, and a huge number of products. Additionally, the sizable industry necessitates improved inventory management. The amount of commodities in each warehouse must thus be precisely managed. In order to manage the manufacturing process in terms of cost and quantity, the firm also has to understand the sales performance of the items based on the warehouse input and output process.

In reality, figuring out how much inventory a business should retain optimally is necessary for studying inventory management and production management. Understanding how numerous elements affect production, sales, distribution requirements, and amounts gained is necessary. To anticipate how many buyers would demand a given product by integrating all of these aspects together, decision-makers will require a clearer overall view of the activities associated with manufacturing and inventories.

However, companies frequently need to know what kind of product will be shipped to which warehouse, which items should be prioritized, or when they are scheduled to be sent in order to track these features. How many things are precisely available, how many goods have been requested, and which items have been delivered. Inventory is categorized differently at different points in the supply chain and turns into data points that need to be tracked to guarantee a seamless transformation process.

Furthermore, It is essential for the manufacturing process to anticipate customer demand in order to fulfill production, manage costs, reserve adequate inventory to satisfy consumer demand, and calculate the necessary amount of raw materials.

In summary, to maximize the company's product management, particularly during the production stage, effective procedures are required. It is crucial to provide the business with complete and accurate product data so that managers and decision-makers may create the best plans and strategies.

1.2 Objectives of the project

1.2.1 General Objective

As a result, the primary goal of our research in this project is to use the AdventureWork 2019 dataset in conjunction with technology tools like SQL Server, Azure, and Power BI to fulfill the need for reports on current inventory management and managing production process quantity and cost. Our goal is to organize the unstructured data, make it analyzable, and visualize it in a dashboard style for better and clearer reporting, enabling the identification of employee weaknesses and strengths, and assisting corporate decision-making.

1.2.2 Specific Objectives

We want to create a data warehouse where you can manage and save all the information you need, and where you can search for all the data associated with your production operations to make information retrieval and data extraction easier. Each order's manufacturing cost, performance during the production process, and product quality will all be disclosed to you. It will be a way to make performance evaluation reports for the organization more streamlined and assist managers in making the best judgments possible.

1.3 Research Objects

The AdventureWorks sample database is built using information from the Adventure Works Cycles fictitious firm. Bicycles composed of metal and composite materials are reportedly produced and sold by a sizable global manufacturing business. While the company's headquarters are in Bothell, Washington, its market spans North America, Europe, and Asia.

In this project, we use the Production module to retrieve data on product information, inventory processes, and production processes. From this module, we design a data warehouse with 3 fact tables.

First, the FactWorkOrder table includes 3 dimensions: ScrapReason, Product, and Date, used to manage the production process and check how many defective products, and the reasons for the defects, were produced during production.

Second, the FactSequenceCost table includes 3 dimensions: Location, Product, and Date, used to manage related costs in the inventory or production process.

Third, the FactInventory table includes 3 dimensions: Product, Date, and Location, used to manage stored products and track inventory levels, shipments, etc.

1.4 Scope of the project

From April 1st, 2023 to May 8th, 2023.

Study population: Data from AdventureWorks2019 database.

Theory: Business Intelligence and Decision Support System.

Purpose: Building a data warehouse to manage warehousing and production operations by putting information from the theoretical course into practice with real-world datasets.

Limitation: This study is limited to the selected tables and columns in the database to address specific issues mentioned in the previous sections.

1.5 Value and desired outcome of the project

In this study, we put "business intelligence and decision support system" knowledge into practice to offer business solutions, giving us a better understanding of how to address issues in real-world scenarios. We understand how to implement data solutions and extract knowledge from this study's findings. Applying a data warehouse and data integration solution to a business issue can lead to better business outcomes. The data may simply be arranged into spreadsheets or tables for examination and analysis.

1.6 Structure of project

Our group developed a project with 6 chapters to outline the activities and research direction of the group throughout the implementation phase in order to establish a data warehouse for the research project's objectives.

Chapter 1: Introduction - In order to define the project's research path, we will give an overview of the project, beginning with establishing the project's business case and moving on to outlining the research objectives and target audience.

Chapter 2: Theoretical basis - To assist the research endeavor, we will provide theories from fundamental to advanced linked to BI, ETL, and Data Warehouse to complement the critical foundational knowledge for the data warehouse development process. Choosing KPIs is crucial for developing a thorough understanding of the data warehouse.

Chapter 3: Requirements Analytics and Introduction to BI Solution - To construct the data warehouse, we will assess and choose the most appropriate modules. Approaching the analysis of departments in practice from there will allow you to observe how it impacts the data warehouse. Once the proper modules and characteristics have been determined, we will proceed to examine the data source's data types, attributes, and descriptions before examining the business needs to determine the best course of action.

Chapter 4: Building Data Warehouse and Integrating Data - After learning the theory and examining the pertinent business processes, we will put it to use by designing the ETL process based on SQL Server 2019 and Microsoft Azure for the design and development of the Data Warehouse.

Chapter 5: Results - Data Analytics and Visualization - After completing all Data Warehouse development procedures, we will demonstrate the project's success through a number of tasks such report writing, Power BI visualization, time series implementation, and forecasting or predictive model implementation.

Chapter 6: Conclusion and Future Works - In the final part, we will present the project conclusion, along with the strengths and weaknesses that can be drawn from the project. From there, we will analyze and provide future prospects in our work to further improve the completed project.

CHAPTER 2: THEORETICAL BASIS

2.1 Overview about BI

2.1.1 What is BI?

Theo Thor Olavsrud and Josh Fruhlinger (2023), Business intelligence (BI) is a set of strategies and technologies enterprises use to analyze business information and transform it into actionable insights that inform strategic and tactical business decisions. BI tools access and analyze data sets and present analytical findings in reports, summaries, dashboards, graphs, charts, and maps to provide users with detailed intelligence about the state of the business.

2.1.2 BI Architecture

Theo Tatyana Korobeyko (2022), Business intelligence (BI) architecture refers to the technical framework that an organization uses to support all phases of the BI process, including data collecting, cleansing, organizing, storing, and analysis, as well as the delivery of reports and dashboards and the operationalization of insights.

The business architecture, which specifies the IT environment (the structure) and business processes (the operation of the organization), includes BI architecture. The enterprise architecture helps the corporation accomplish its strategic and tactical goals.

From the article above, we can understand simply about the 5 components to build a business intelligence architecture as follows:

Data sources: A data source is anything that generates the digital data that BI systems use. Data sources can be internal or external, depending on where the information is created and stored. Internal data sources are where an organization captures and maintains its own data.

Data integration and data quality management layer: The goal of the second phase in the BI process is to combine datasets from several sources into a single perspective so that the information may be used for analytical and operational needs. The kind, structure, and volume of the information as well as the intended application—operational reporting, business analysis, ML use cases, etc.—determine the best data integration approach.

Data repositories: This component includes multiple repositories for organizing and storing data so that it may be processed later. There are two main categories of data repositories:

- Analytical data stores
- Data marts

In fact, some companies also have complementary repositories. Having analytics repositories alone may not be adequate for a corporation these days due to the wide variations in data volumes and content. It takes a lot of resources, time, and money to store all data in these repositories.

BI and analytics layer: Solutions for accessing and interacting with data that are geared for data analysts, data scientists, or business users are included in this layer. This layer naturally reflects the organization's BI maturity and its data analytics goals: while some businesses only need descriptive and diagnostic analytics, others need running thorough analyses backed by ML and AI through a self-service user interface.

Data governance layer: This component's primary function is to oversee and manage the whole BI process, which makes it closely related to the other four. When a corporation has data governance rules and policies in place, it can regulate who has access to the information, how it is accessed, if the data is of good quality and is properly protected, etc. An automated data management program may be created using the data governance tools by combining all of these policies and standards.

2.1.3 Advantage of BI in enterprises

Maximized data value: A business may create a high-performing information management environment with all of its components connected and cooperating by establishing a BI architecture. By putting such a system in place, businesses may maximize the value of their data with little manual work and reduce the amount of "dark data."

Offloaded IT department: Even though dedicated IT and data analytics teams cannot be completely offloaded by BI architecture, it does significantly relieve the IT and DA departments from onerous data management tasks like gathering data from corporate systems, modeling, creating routine reports, etc.

Increased efficiency of business users: Business users don't have to wait for choices or make them instinctively since the installation of the BI architecture results in simplified information management and analytics procedures.

Cost savings: Despite being an expensive endeavor, establishing a full-scale BI system offers cost benefits since a company: (1) Doesn't need the deployment and operation of many systems to meet the unique analytics requirements of various departments; (2) Automates and integrates all aspects of data administration; (3) Reduces errors caused by inadequate data quality, ineffective security, etc.; (4) Stops shadow IT.

2.1.4 BI Strategy for Business

According to the Tableau (2022), Your road map to success is a BI strategy. In the early stages, it will be necessary to choose how data will be utilized, identify important roles, and specify duties. Starting with company objectives is your key to success, even if it may appear straightforward at first glance.

Here is how to build a BI strategy from scratch:

- Know your business's objectives and strategy.
- Identify the important parties.
- Pick a sponsor from among your important stakeholders.
- Pick your BI tools and platform.
- Form a BI team.
- Establish your domain.
- Your data infrastructure should be ready.
- Establish your plans and objectives.

2.2 ETL Process

2.2.1 What is ETL?

According to the IBM Cloud (2022), ETL is an abbreviation for Extract - Transform - Load, which refers to the procedure of extracting, transforming, and loading data. It is a data integration procedure that transfers raw data from a system source to a separate data system (data lake or data warehouse) that is housed on a particular server. The data is then turned into information that may be used in accordance with the objectives of the company.

ETL was created as a technique for integrating and loading data for computation and analysis as databases gained popularity in the 1970s. Eventually, it became the primary method for data processing in data warehousing operations.

ETL serves as the foundation for workstreams in data analytics and machine learning. In order to satisfy specific business intelligence requirements, such as monthly reporting, ETL cleans and organizes data using a set of business rules. However, it can also handle more complicated analytics to enhance back-end operations or end-user experiences. Frequently, a business will use ETL to:

Extract information from old systems.
Cleanse the data to improve its consistency and quality.
Load the chosen database with data.

2.2.2 Why do we need ETL?

Complex business queries cannot be answered by legacy systems using traditional databases; only ETL can achieve this.

It makes it possible for businesses to retrieve old information that can be utilized to offer context and a thorough picture of the company throughout time.

It enhances and incorporates business intelligence decision-making solutions.

It transforms fragmented data into a unified framework and produces pertinent patterns and insights.

It integrates information gathered from partners, suppliers, and recent mergers and acquisitions of businesses.

It enables sample data comparison between the source and destination systems as well as the establishment of a centralized hub or data repository for easy access.

By codifying and reusing without requiring additional technical knowledge, it improves efficiency. Developers and analysts may now invest time on business research and growth rather than creating specialized tools for study.

2.2.3 ETL Process

In the world of data, we are aware that ETL is the cornerstone of the data warehouse and that the data warehouse is the foundation of business intelligence. Data is extracted from numerous source systems using ETL tools, transformed inside the organizational environment, and then loaded into the data warehouse system.

Extraction: Diverse types of databases, Excel files, or raw files are just a few of the many forms in which source data from diverse source systems can be retrieved. Instead of storing the data straight into the data warehouse, it is crucial to first extract the data from multiple source systems and stage it in a staging area since the retrieved data may be damaged or have different formats.
Transformation: The raw data is now processed using data at this step. For further analysis, the data is combined and converted. This stage will need the usage of transformation procedures like:
choosing the essential and acceptable data columns.
changing data. For instance, changing the number 1 to "Nam" or the opposite. New calculated columns being created. An average score column, for instance.
Data sorting and filtering.

Executing actions for aggregation. Counting the number of rows, or adding the total of the columns.

Creating fresh values. Creating an auto-incremented key is one example.

Data research or comparison.

Loading: Pushing the data into the data warehouse following the final transformation is what this procedure entails. Developers may create apps and end users can base decisions on the data and the applications they have access to thanks to the data being loaded into a centralized storage place.

2.3 Data warehouse and Data mart

2.3.1 What are Data warehouse and Data mart?

Data warehouse

Theo Mary K. Pratt and Jacqueline Biscobing (2023), a data warehouse is a sizable central data repository inside a company that holds data from several sources. The gathered data is used to guide corporate decisions through the use of analysis, reporting, and data mining tools.

Inmon vs. KimBall

Two data warehouse pioneers, Bill Inmon and Ralph Kimball differ in their views on how data warehouses should be designed from the organization's perspective.

Bill Inmon

The key component of an organization's data systems, according to Bill Inmon's top-down approach, which advocates a centralized data repository, is the data warehouse.

The comprehensive corporate data model developed using the Inmon technique is thought of as having a physical representation in the data warehouse. The data warehouse may be used to create dimensional data marts specific to particular business lines as needed.

According to the Inmon model, the data warehouse serves as the source for the data that is eventually used in the various data marts. This guarantees data consistency and integrity throughout the company.

Ralph Kimball

The most important business processes occur first in Ralph Kimball's data warehouse design. Using this strategy, a company produced data marts that gather crucial data on particular subject areas. The data warehouse combines the many data marts used by the company.

The Kimball approach creates a specific data warehouse out of several data marts. compared to Inmon's approach, which relies on data from the warehouse to construct data marts. Kimball remarked in 1997 that "the data warehouse is nothing more than the union of all data marts."

Data mart

A data mart is a section of a data warehouse that is targeted toward a certain industry. Data marts are storage locations for compiled data that has been acquired for analysis on a certain division or unit inside a business, such the sales division.

2.3.2 Who needs Data warehouse and Data mart?

DWH is needed for all types of users like:

- Those who base their judgments on a lot of information.
- Users who use intricate, specialized procedures to get data from several data sources.
- It is also used by those who favor simple technology for information gathering.
- Additionally, it is essential for those who want to make logical assessments.
- If the customer wants speedy performance on a big amount of data that is needed for reports, grids, or charts, data warehouses might be useful.

A data mart is simple to access and utilize for business analysts, other end users, and the BI and data analysts assigned to the business unit. Data marts assist departments in making critical business decisions. Data marts can be used, for instance, by a marketing team to evaluate customer behavior and by a sales team to provide quarterly sales reports. The teams don't require access to all corporate data because these responsibilities are carried out inside their individual departments.

2.3.3 Advantages and disadvantages of Data warehouse

Advantages of DWH:

Data warehouses allow business users to quickly and easily access important data from several sources.

Data warehouse provides consistent data on several cross-functional tasks. Additionally supported are ad hoc reporting and querying.

Data warehouses help with the integration of many data sources to minimize the burden on the production system.

Making advantage of a data warehouse may generally speed up analysis and reporting.

Thanks to reorganization and integration, the user may make better use of it for reporting and analysis.

Data warehouses enable users to access essential data from several sources in a single spot. Users may therefore access data from diverse sources more quickly as a consequence.

In data warehouses, a substantial amount of historical data is preserved. This may be used by users to analyze past events and trends and predict the future.

Disadvantages of DWH:

A poor option for unstructured data.

It goes without saying that creating and implementing a data warehouse takes time.

Data Warehouse is prone to being outdated.

It can be difficult to make changes to data types, ranges, data source structure, indexes, and searches.

Although the data warehouse can seem straightforward, most users would consider it to be overly complex.

Even the best project management efforts cannot stop the scope of a data warehousing project from growing.

Warehouse users periodically develop own business rules.

Organizations must devote a large portion of their budget to implementation and training.

2.3.4 Snowflake and Star schemas

According to David Taylor (2023), Data structure method called Dimensional Modeling (DM) is designed specifically for data warehouse storage. Dimensional modeling is used to enhance databases for quicker data retrieval. The "fact" and "dimension" tables that make up the Dimensional Modelling idea were created by Ralph Kimball.

Star Schemas (SS):

The star schema is the fundamental building component used in dimensional modeling. A star schema is made up of several smaller tables called dimension tables that radiate out from a central, larger table called the fact table.

Each dimension table has a one-to-many link with the fact table. The fact table often shows business transactions, events, or a quick snapshot of the transactions, events, or transactions.

Snowflake (SF):

The star schema is expanded upon in the snowflake schema. Each dimension in a snowflake schema is normalized and linked to further dimension tables.

Galaxy Schema (GS):

The data warehouse schema's successor is the Galaxy Data Warehouse Schema, sometimes referred to as a Fact Constellation Schema. Unlike the Star and Snowflake Schemas, the Galaxy Schema makes use of a number of fact tables connected by shared normalized dimension tables. Galaxy Schema prevents data duplication and discrepancies by connecting and appropriately normalizing a number of star schema.

2.4 KPIs

2.4.1 KPIs Definition

According to Alexandra Twin (2023), Key performance indicators (KPIs) are a group of quantitative metrics used to assess the overall long-term performance of an organization. KPIs in particular aid in determining a company's strategic, financial, and operational accomplishments, particularly when compared to those of rival companies in the same industry.

2.4.2 The advantages and disadvantages of KPIs

Advantages:

KPIs swiftly demonstrate current performance, target attainment, or other strategic objectives.

Better decision-making: Based on exact measurements and accompanying numerical data, decisions are likely to be made more quickly and with more knowledge.

KPIs offer quantifiable goals, making performance reviews more precise and simple to carry out. This helps to reduce disagreements and discrepancies during the job execution process.

Disadvantages:

KPIs must adapt readily to the business's objectives.

The performance evaluation system and job management system of companies and enterprises would suffer if the KPIs developed do not match the SMART requirements.

The indications will become unattainable and unachievable if the construction KPI is erroneous and confusing.

2.4.3 Categories of KPIs

2.4.3.1 Manufacturing

Manufacturing Efficiency: This KPI measures how efficiently the manufacturing process is operating. It is usually calculated as the ratio of actual production output to the planned production output. This KPI is important for identifying any inefficiencies in the production process, which can help the company reduce costs and improve profitability. By monitoring this KPI, the company can make informed decisions about how to optimize the production process and improve overall efficiency.

Scrap Ratio: Scrap ratio measures the percentage of raw materials or finished products that are scrapped or rejected during the manufacturing process. This KPI is important for identifying any issues in the production process that may be causing excessive waste or inefficiencies. By tracking this KPI, the company can identify areas for improvement in the manufacturing process and reduce costs associated with waste.

Production Quantity over time: This KPI measures the quantity of products produced over a specific period of time. It is important for monitoring overall production capacity and identifying any fluctuations in demand. By tracking this KPI, the company can make informed decisions about production planning and inventory management, ensuring that they are able to meet customer demand while minimizing costs associated with excess inventory.

2.4.3.2 Inventory

Inventory Carrying Costs: the full amount businesses spend to stock and store items before they’re sold.

Safety Stock Level (SSL): SSL is a crucial metric for businesses, especially those involved in inventory management and supply chain operations. SSL represents the minimum level of inventory that a company should maintain to ensure that customer demand can be met even if there are disruptions in supply or unexpected spikes in demand.

Reorder Point (ROP): ROP is a critical metric for businesses involved in inventory management and supply chain operations. ROP represents the inventory level at which a company needs to place a new order with its suppliers to replenish the inventory.

2.4.3.3 Costs

Cost of Production: The whole cost of manufacturing a good, including all direct and indirect production expenses, is tracked by this KPI. Organizations can find places where they can cut expenses and increase profitability by measuring this indicator.

Resource Hours: This KPI calculates the proportion of time that labor and other resources, such equipment, are being used efficiently. A low utilization rate suggests that resources are not being used effectively and are being squandered. Increased expenses and decreased profitability may follow from this. Organizations can pinpoint areas where they can streamline resource allocation and cut expenses by monitoring this indicator.

Forecast Accuracy: This KPI measures the accuracy of the organization's forecasted costs. By tracking this metric, organizations can identify areas where they need to improve their forecasting methods.

2.5 Power BI

According to an article by Microsoft (2023), Power BI is a group of software services, applications, and connections that combine to transform your disparate data sources into coherent, engaging visuals, and interactive insights. Your data may be stored in a hybrid data warehouse that is both cloud-based and on-premises, or it may be an Excel spreadsheet. Power BI makes it simple to connect to your data sources, view the data, identify the key information, and share it with whoever you choose.

2.6 Azure Microsoft Portal

According to an article by Microsoft (2022), Command-line tools can be replaced with the Azure portal, a web-based unified console. You may utilize the graphical user interface of the Azure portal to manage your Azure membership. In the portal, you can create, administer, and keep an eye on everything from straightforward web apps to intricate cloud deployments.

The Azure interface is built with continuity and resilience in mind. Every Azure datacenter has it present. By being proximate to users, this arrangement makes the Azure portal robust to outages in particular datacenters and prevents network slowdowns. The Azure site is constantly updated and doesn't need to go down for maintenance.

CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION

3.1 Business processes

3.1.1 Production department

Adventure Works 2019 refers to the division in charge of creating and assembling items as the "Production" department. In order to guarantee that goods are produced in a timely and effective way, the production department is in charge of arranging the manufacturing process. To improve production schedules, they employ a number of methods and instruments, including production scheduling software and capacity planning. To start the production process, the production department generates a work order. Work orders provide details on the product name, quantity, and deadline for completion. In addition to maintaining the inventory of completed items and raw materials, the production department also keeps an eye on the quality of the commodities being produced.

3.1.2 Production Control Department

The manufacturing Control division is a component of the Production schema, which is in charge of overseeing the company's product manufacturing. The management of the tools and resources, such as labor, machinery, and raw materials, required for production, may fall within the purview of the production control department. It is essential to ensure the company's efficient and effective manufacture of items.

3.2 The purpose of the Product and Production Control Department

In order to fulfill client demands and increase profitability, AdventureWorks must be able to create high-quality goods quickly and effectively. This is where the Product and Production Control Department comes in. To guarantee that items are available when needed and to reduce the expenses of retaining excess inventory, the department is in charge of controlling inventory levels. The department is in charge of making sure that products adhere to the business's quality standards and that any flaws or problems are swiftly found and fixed. The department is in charge of spotting possibilities to boost the efficacy and efficiency of the business' manufacturing processes and putting such improvements into action to promote continual improvement.

3.3 Production process

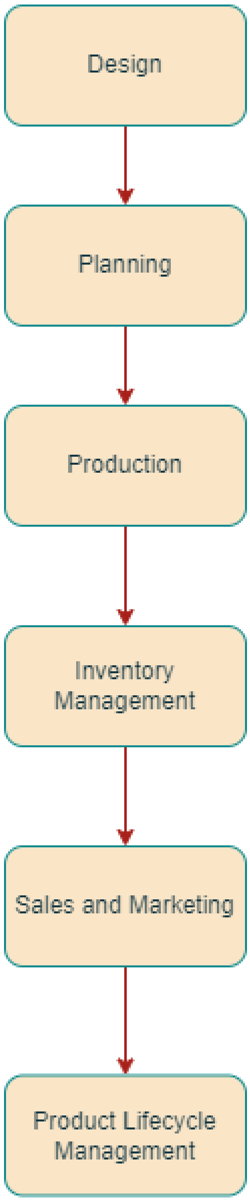


*Figure 3.**Error! No text of specified style in document.**.1 Production Process*

(The author summarizes and recommends)

In AdventureWorks 2019, managing goods often entails the following steps:

Product Design: To develop new items or improve current ones, the product department collaborates with the design team. This includes designing prototypes, formulating product specifications, and testing the finished goods to make sure they adhere to quality standards.

Product Planning: The Product Department develops a strategy for the products' manufacture when the product designs are completed. This involves figuring out the required raw material amount, the manufacturing schedule, and the anticipated expenses.

Production: Following the creation of the plan in the preceding stage, the items are next made. The Product Department is in charge of making sure that the goods are produced on schedule, adhere to quality requirements, and are delivered to clients on schedule.

Inventory Control: The Product Department is in charge of overseeing the stock levels at the business. This entails keeping an eye on stock levels, making demand projections, and placing orders for fresh goods or raw materials as required.

Sales and Marketing: To market the company's products and increase sales, the Product Department collaborates closely with the Sales and Marketing departments. This includes creating marketing campaigns, going to trade exhibitions, and educating salespeople about the products.

Product Lifecycle Management: Management of the full product lifespan, from introduction to retirement, is within the purview of the product department. Making choices here includes determining when to discontinue a product and how to get rid of any lingering inventory.

3.4 Data source and challenges

The Adventure Works database supports typical online transaction processing situations for a bicycle manufacturer (Adventure Works Cycles). Scenarios include those in human resources, production, sales, and purchasing. In AdventureWorks 2019, the production division is in charge of overseeing the manufacture of finished items, keeping track of work orders, and controlling the creation of raw materials. In our project, we focus on the Production process with different requirements and tables from the original data source of Adventure Works. After searching and analyzing the data sources, we finally have our own database. We mainly worked on the Production process with 11 tables and lots of user requirements.

In the original database, the Production process of Adventure Work has 24 different tables as shown in Schema Data Source:

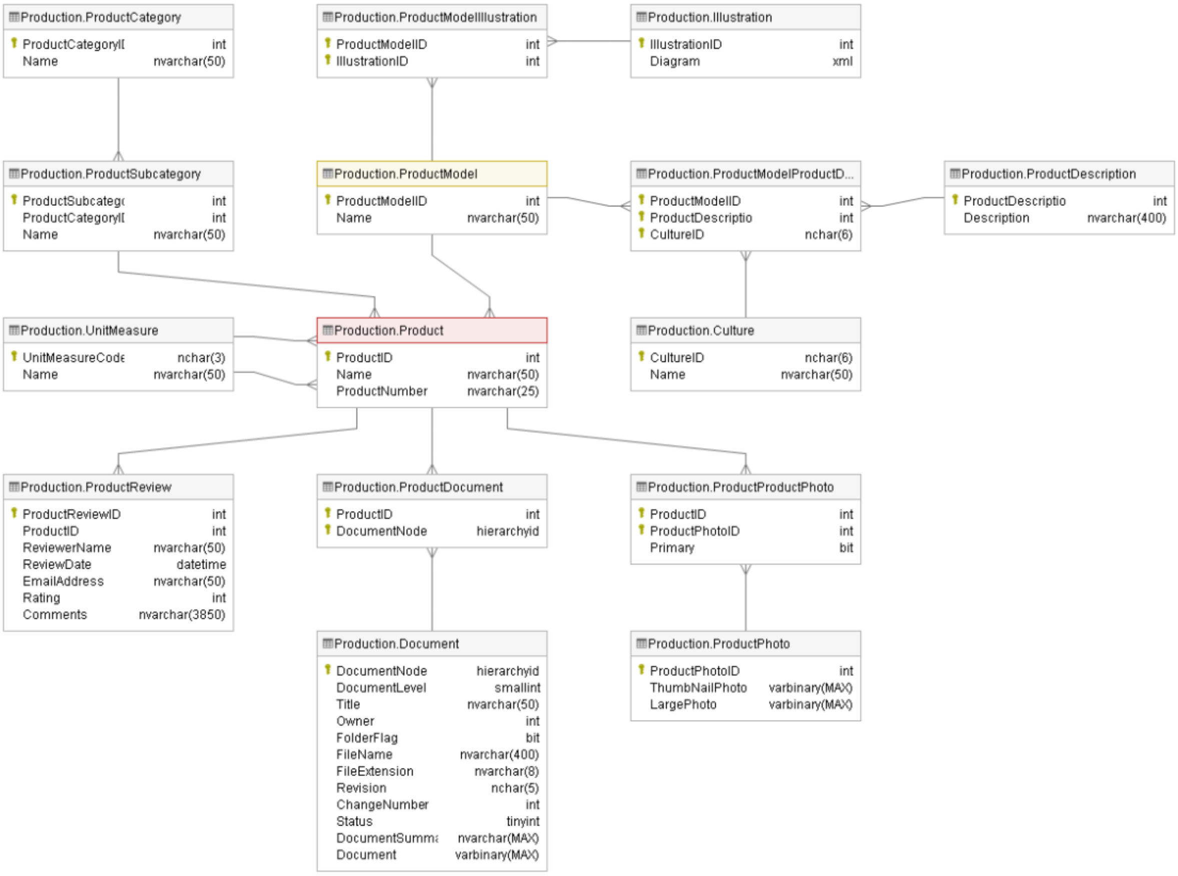


Figure 3.2 Schema Data Source

However, we only focus on 6 tables on this schema: Product, Location, ProductSubCategory, ProductCategory, ScrapReason, WorkOrder.

Table Production.Product:

Key	**Name**	**Data type**	**Null**	**Attributes**	**Description**
PK	ProductID	int		Identity	Primary key for Product records.
	Name	nvarchar (50)			Name of the product.

	Product Number	nvarchar (25)			Unique product identification number.
	MakeFlag	bit		Default: 1	0 = Product is purchased, 1 = Product is manufactured in-house.
	Finished GoodsFlag	bit		Default: 1	0 = Product is not a salable item. 1 = Product is salable.
	Color	nvarchar (15)	x		Product Color.
	Size	nvarchar (5)	x		Product size.
	SizeUnit MeasureCode	nchar(3)	x		Unit of measure for Size column.
	WeightUnit MeasureCode	nchar(3)	x		Unit of measure for Weight column.
	Weight	decimal(8, 2)	x		Product weight.

	Class	nchar(2)	x		H = High, M = Medium, L = Low
	Style	nchar(2)	x		W = Womens, M = Mens, U = Universal
FK	ProductSubcategoryID	int			Product is a member of this product subcategory. Foreign key to ProductSubCategory.ProductSubCategoryID.
	ModifiedDate	datetime		Default: getdate( )	Date and time the record was last updated.

*Table 3.**Error! No text of specified style in document.**.1 Table Production.Product*

The table "Production.Product " contains information about products sold or used in the manufacturing of sold products. It has several columns, each representing a different attribute of the product.

Table Production.ProductCategory:

Key	Name	Data type	Null	Attributes	Description
PK	ProductCategoryID	int		Identity	Primary key for ProductCategory records.
	Name	nvarchar(50)			Category description.

	ModifiedDate	datetime		Default: getdate()	Date and time the record was last updated.

*Table 3.**Error! No text of specified style in document.**.2 Table Production.ProductCategory*

In AdventureWorks2019, the Production.ProductCategory table stores information about the High-level product categorization of products that are manufactured or sold by the company.

Table Production.ProductSubcategory

Key	Name	Data type	Null	Attributes	Description
PK	ProductSubcategoryID	int		Identity	Primary key for ProductSubcategory records.
FK	ProductCategoryID	int			Product category identification number. Foreign key to ProductCategory.ProductCategoryID.
	Name	nvarchar(50)			Subcategory description.
	ModifiedDate	datetime		Default: getdate()	Date and time the record was last updated.

*Table 3.**Error! No text of specified style in document.**.3 Table Production.ProductSubcategory*

Information about the High-level product subcategorization of goods produced or sold by the firm is kept in AdventureWorks2019 "Production.ProductSubcategory" table.

Table: Production.ScrapReason

Key	Name	Data type	Null	Attributes	Description
PK, FK	ScrapReasonID	smallint			Primary key for ScrapReason records Identity / Auto increment column
	Name	nvarchar(50)			Failure description
	ModifiedDate	datetime		Default: getdate()	Date and time the record was last updated.

*Table 3.**Error! No text of specified style in document.**.4 Table Production.ScrapReason*

The Production.ScrapReason table in AdventureWorks 2019 is a lookup table that stores the reasons for why products are being scrapped during production.

Table: Production.WorkOrder

Key	Name	Data type	Null	Attributes	Description
PK	WorkOrderID	int			Primary key. Foreign key to WorkOrder.WorkOrder ID.

PK	ProductID	int			Primary key. Foreign key to Product.ProductID.
PK	Operation Sequence	smallint			Primary key. Indicates the manufacturing process sequence.
	LocationID	smallint			Manufacturing location where the part is processed. Foreign key to Location.LocationID.
	ActualStartDate	datetime	x		Actual start date.
	ActualEndDate	datetime	x		Actual end date.
	ActualResource Hrs	dec imal(9, 4)	x		Number of manufacturing hours used.
	ActualCost	money	x		Actual manufacturing cost.
	ModifieDate	datetime		Default: getdate()	Date and time the record was last updated.

*Table 3.**Error! No text of specified style in document.**.5 Table Production.WorkOrder*

Production.WorkOrder is a table in the AdventureWorks2019 database that stores information about the work orders for the production of goods.

Table: Production.Location

Key	Name	Data type	Null	Attributes	Description
PK	LocationID	smallint		Identity	Primary key for Location records.
	Name	nvarchar(50)			Location description.
	CostRate	smallmoney		Default: 0.00	Standard hourly cost of the manufacturing location.
	Availability	decimal(8, 2)		Default: 0.00	Work capacity (in hours) of the manufacturing location.
	ModifiedDate	datetime		Default: getdate()	Date and time the record was last updated.

*Table 3.**Error! No text of specified style in document.**.6 Table Production.Location*

Table: Production.Location is a table in the AdventureWorks2019 database that stores information related to the physical location of a company's facilities where production activities are taking place.

Table: Production.ProductInventory

Key	Name	Data type	Null	Attributes	Description
PK, FK	ProductID	int			Product identification number.

PK, FK	LocationID	smallint			Inventory location identification number.
	Shelf	nvarchar(10)			Storage compartment within an inventory location.
	Bin	tinyint			Storage container on a shelf in an inventory location.
	Quantity	smallint		Default: 0	Quantity of products in the inventory location.
	rowguid	uniqueidentifier		Default: newid()	ROWGUIDCOL number uniquely identifying the record. Used to support a merge replication sample.
	ModifiedDate	datetime		Default: getdate()	Date and time the record was last updated

*Table 3.**Error! No text of specified style in document.**.7 Table Production.ProductInventory*

Production.ProductInventory is a table in the AdventureWorks2019 database that contains information about the inventory levels of each product in each inventory location.

Challenges: Managing the manufacturing process presents a number of difficulties for AdventureWorks 2019's manufacturing department. Among the principal difficulties are:

Production orders and work orders can be difficult to schedule, particularly when there are competing priorities or little resources available.

Inventory Management: It can be challenging to control inventory levels and make sure there is always enough raw material on hand to satisfy manufacturing demands.

It can be difficult to ensure that completed products fulfill quality requirements when there are problems with the quality of the raw materials or the manufacturing processes.

Supply Chain Management: Coordinating with suppliers and managing the supply chain may be challenging, particularly when there are delays or problems with delivery.

3.3 Business Requirements Analysis

3.3.1 Report on Costs of Production

An important statistic that businesses use to assess the effectiveness of their manufacturing process is the cost of production. Companies can find areas of inefficiency and implement the required changes to increase profitability by assessing the cost of production. An examination of AdventureWorks' manufacturing expenses is provided in this paper. Businesses may find inefficient regions and put cost-controlling measures in place by assessing production costs and segmenting them by product, subcategory, and category. Businesses receive useful data from the cost analysis report that they may use to decide wisely about manufacturing procedures and pricing schemes. Businesses are able to anticipate production costs more precisely in the cost analysis report, which helps them organize their budget and resources more efficiently. Businesses may create pricing plans that maximize profitability by analyzing the expenses associated with each product, subcategory, and category.

3.3.2 Report on the production line

It gives management levels a broad perspective of the effectiveness and performance of the manufacturing process. Managers may more easily monitor progress and control output by using the dashboard's different charts and graphs that display critical data. Customers may view the state and caliber of the business, which can increase client happiness and draw in new clients. They can also learn about AdventureWorks' capability for accurately and quickly processing requests. Additional insight into the performance of the manufacturing process is provided by the tables in this report that provide production figures and predictions for each product over time. This information enables managers to spot patterns and implement the appropriate corrective actions to increase productivity and efficiency. Businesses may benefit greatly from this work order report by allowing management to make educated decisions, maximize production output, and more.

3.3.3 Report on Inventory Status

A report on inventory gives a thorough analysis of the trends, expenses, and quantities of inventory. Businesses may pinpoint areas for development and enhance their inventory management procedures by analyzing this data. The inventory report offers information on the expenses related to keeping and storing goods. Businesses may take action to lower expenses and increase profitability by identifying areas of high cost. Businesses may decide wisely on inventory levels, purchases, and stocking with the use of inventory reports. Better decisions are made as a result, and resources are used more effectively. Businesses may make sure they have the things consumers need when they need them by using effective inventory management. Higher client retention and satisfaction result from this. An early warning system for items that are going low or out of stock is also provided by the inventory report. Businesses may make sure they always have enough inventory on hand to fulfill client demand by setting reorder points and safety stock levels. In conclusion, an inventory report may offer useful information on inventory levels, expenses, and trends, assisting firms in enhancing their inventory management procedures, cutting costs, and reaching more informed judgments. Businesses may enhance customer happiness, boost profitability, and remain competitive with the help of this information.

3.4 IT requirements Analysis (IT & Infrastructure)

Define the project's objectives and scope: This entails determining the business opportunity or problem that the data warehouse is meant to solve, as well as the targeted results and advantages.

Analyze and rank the requirements: This entails looking over the requirements, finding any conflicts or contradictions, and deciding which needs should be given top priority. It is necessary to divide the requirements into functional and non-functional categories.

Create a data warehouse architecture by creating a data warehouse architecture.

Create a data model: This entails creating the data model that will be included into the data warehouse to serve the organization's reporting and analytics requirements.

Create a data model: This entails creating the model that will be utilized in the data warehouse to fulfill the reporting and analytics requirements of the firm.

CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

4.1 Designing Data Warehouse

4.1.1 Bus Matrix

Business Process	Common dimension				
	Date	Product	Product Category	Inventory	Cost
Analyzing Product Sales	X	X	X		
Monitoring Production Output	X	X			X
Inventory Management	X	X		X	
Monitoring Production Costs and Product Costing	X	X			X
Analyzing Warehouse Management Efficiency	X			X	X
Analyzing Products that have been in Inventory for Too Long	X	X		X	
Optimizing Product-related Costs	X	X	X	X	

Table 4.1 Bus matrix

4.1.2 Master Data

Object	Description
ProductID	Unique identifier for each product
ProductSubcategoryID	Foreign key linking to the product subcategory table
ProductCategoryID	Unique identifier for each product category
LocationID	Unique identifier for each location
ScrapReasonID	Unique identifier for each scrap reason
OperationSequence	The sequence of the operation for the work order

Name	Name of the product or product category or location or scrap reason
Color	Color of the product
Class	Product class
Style	Product style

Table 4.2 Master Data

4.1.3 Transaction Data

Object	Description
OrderQty	Product quantity to build
StockedQty	Quantity built and put in inventory
ScrappedQty	Quantity that failed inspection
StartDate	Start date of the work order
EndDate	End date of the work order
PlannedCost	Estimated manufacturing cost
ActualCost	Actual manufacturing cost
SafetyStockLevel	Minimum inventory quantity
ReorderPoint	Inventory level that triggers a purchase order or work order
StandardCost	Standard cost of the product
SellStartDate	Date the product was available for sale

Table 4.3 Transaction Data

4.1.4 Fact and dimension tables

4.1.4.1 DimProduct

SELECT ProductID, ProductSubcategoryID, Name, Color, Size, Weight, Class, Style
FROM Production.Product

Column Name	Data Type
ProductID	int

ProductSubcategoryID	int
Name	nvarchar(50)
Color	nvarchar(15)
Size	nvarchar(5)
Weight	decimal(8, 2)
Class	nvarchar(2)
Style	nvarchar(2)

Table 4.4 DimProduct

4.1.4.2 DimProductCategory
SELECT ProductCategoryID, Name
FROM Production.ProductCategory

Column Name	Data Type
ProductCategoryID	int
Name	nvarchar(50)

Table 4.5 DimProductCategory

4.1.4.3 DimProductSubcategory
SELECT ProductSubcategoryID, ProductCategoryID, Name
FROM Production.ProductSubcategory

Column Name	Data Type
ProductSubcategoryID	int
ProductCategoryID	int

Name	nvarchar(50)

Table 4.6 DimProductSubcategory

4.1.4.4 DimLocation

SELECT [LocationID],[Name]
FROM [Production].[Location]

Column name	Data Type
LocationID	int
Name	nvarchar(50)

Table 4.7 DimLocation

4.1.4.5 DimScrapReason

SELECT ScrapReasonID, Name
FROM [Production].[ScrapReason]

Column name	Data Type
ScrapReasonID	int
Name	nvarchar(50)

Table 4.8 DimScrapReason

4.1.4.6 DimDate

Column Name	Data Type
DateKey	varchar(20)
Date	datetime

FullDate	char(10)
DayOfMonth	varchar(2)
DayName	varchar(9)
DayOfWeek	char(1)
DayOfWeekInMonth	varchar(2)
DayOfWeekInYear	varchar(2)
DayOfQuarter	varchar(3)
DayOfYear	varchar(3)
WeekOfMonth	varchar(1)
WeekOfQuarter	varchar(2)
WeekOfYear	varchar(2)
Month	varchar(2)
MonthName	varchar(9)
MonthOfQuarter	varchar(2)
Quarter	char(1)
QuarterName	varchar(9)
Year	char(4)

YearName	char(7)
MonthYear	char(10)
MMYYYY	char(6)
FirstDayOfMonth	date
LastDayOfMonth	date
FirstDayOfQuarter	date
LastDayOfQuarter	date
FirstDayOfYear	date

Table 4.9 DimDate

4.1.4.7 FactWorkOrder

SELECT ProductID, ScrapReasonID, WorkOrderID, OrderQtyStockedQt , ScrappedQty, StartDate, convert(varchar(20), EndDate,112) as EndDate FROM Production.WorkOrder.

Column Name	Data Type
ProductID	int
WorkOrderID	int
ScrapReasonID	int
OrderQty	int
StockedQty	int

ScrappedQty	int
StartDate	datetime
EndDate	varchar(20)

Table 4.10 FactWorkOrder

4.1.4.8 FactSequenceCost

SELECT ProductID, work.LocationID, WorkOrderID, OperationSequence,
ActualResourceHrs, lo.CostRate, PlannedCost, ActualCost, ActualEndDate
convert(varchar(20), ActualEndDate,112) as ActualEndDate
FROM Production.WorkOrderRouting work
JOIN Production.Location lo
ON work.LocationID = lo.LocationID

Column Name	Data Type
ProductID	int
LocationID	int
WorkOrderID	int
OperationSequence	int
ActualEndDateKey	varchar(20)
ActualResourceHrs	decimal(9, 4)
CostRate	money
PlannedCost	money

ActualCost	money
ScheduledStartDate	datetime
ScheduledEndDate	datetime
ActualStartDate	datetime

Table 4.11 FactSequenceCost

4.1.4.9 FactInventory

SELECT inv.LocationID, pr.ProductID, ProductSubcategoryID, MakeFlag, SafetyStockLevel, ReorderPoint, StandardCost,Quantity, convert(varchar(20), SellStartDate,112) as SellStartDate
FROM Production.ProductInventory inv
JOIN Production.Product pr
ON inv.ProductID = pr.ProductID

Column Name	Data Type
LocationID	int
ProductID	int
MakeFlag	bit
SafetyStockLevel	int
ReorderPoint	int
StandardCost	money
Quantity	int

SellStartDate	varchar(20)

Table 4.12 FactInventory

4.1.5 Data Warehouse model

Figure 4.1 Relationship between fact and dimension tables

4.2 ETL processes

4.2.1 Create Database

As the Staging layer, create a database to house operational data.
The author team uses the AdventureWorks2019 database as the Staging layer by uploading the whole AdventureWorks2019 database to the cloud as depicted in image x. This means that the everyday operations of the company (or occasionally, depending on the business choice) will be uploaded to and stored in this database.

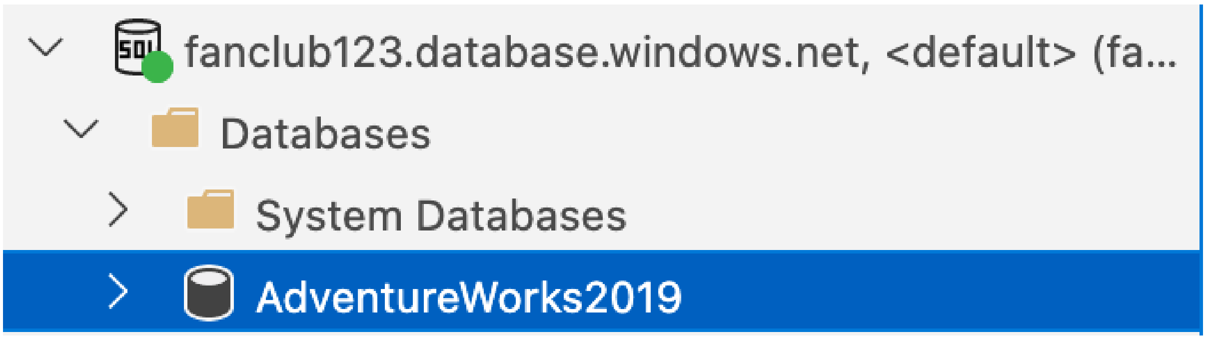


Figure 4.2 Database AdventureWorks2019

The team also creates scripts for the Dim and Fact tables they have created during this stage.

The author group then developed the DW database, as seen in picture x. The whole Dim and Fact tables for the Data Warehouse design are stored in this database.
At the staging layer, the team first constructed empty Dim and Fact tables with attribute data types that were compatible with the source data.

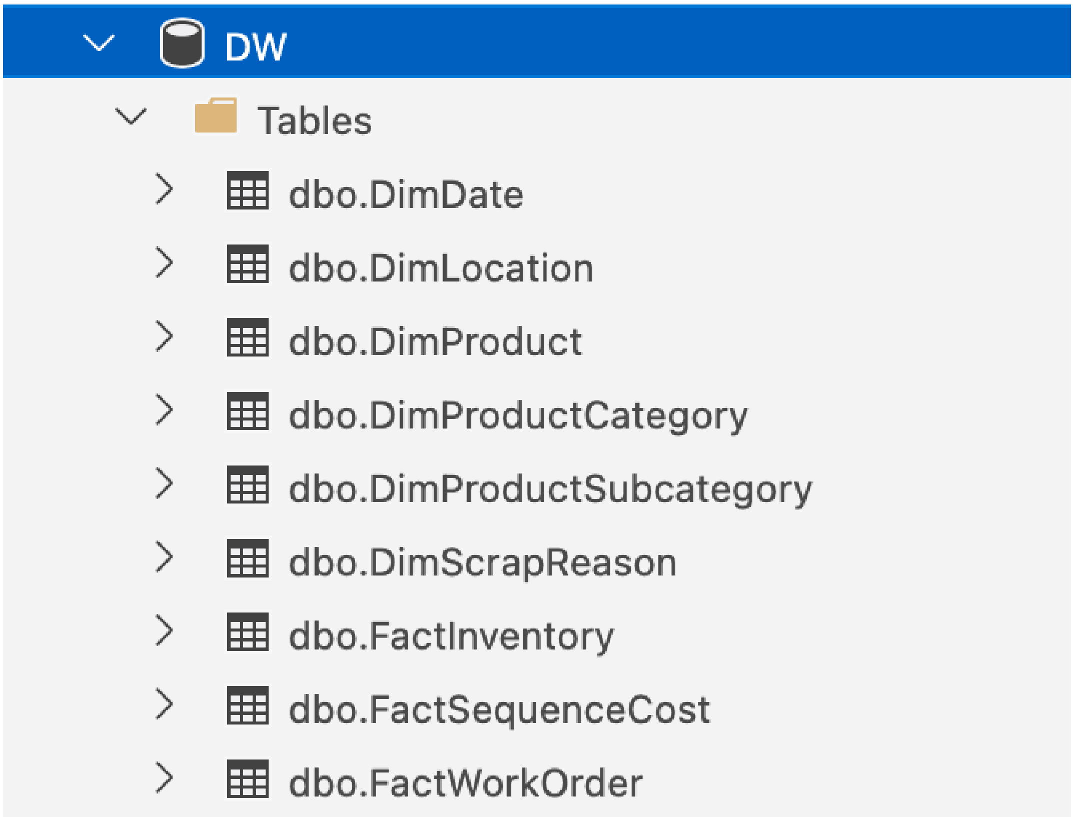


Figure 4.3: Database DW

4.2.2 Dimension Table's ETL Process

4.2.2.1 Generate DimDate

We also require a Dim Date table that corresponds to the business data activity period of 2008–2019 in addition to the aforementioned tables. The DimDate table that is produced will resemble this figure x

	DateKey	Date	FullDate	DayOfMonth	DayName	DayOfWeek	DayOfWeekInMonth
2008-01-14 00:00:00.000		008-01-01 00:00:00.000	01/01/2008	1	Tuesday	3	1
2	20080102	2008-01-02 00:00:00.000	01/02/2008	2	Wednesday	4	1
3	20080103	2008-01-03 00:00:00.000	01/03/2008	3	Thursday	5	1
4	20080104	2008-01-04 00:00:00.000	01/04/2008	4	Friday	6	1
5	20080105	2008-01-05 00:00:00.000	01/05/2008	5	Saturday	7	1
6	20080106	2008-01-06 00:00:00.000	01/06/2008	6	Sunday	1	1
7	20080107	2008-01-07 00:00:00.000	01/07/2008	7	Monday	2	1
8	20080108	2008-01-08 00:00:00.000	01/08/2008	8	Tuesday	3	2
9	20080109	2008-01-09 00:00:00.000	01/09/2008	9	Wednesday	4	2
10	20080110	2008-01-10 00:00:00.000	01/10/2008	10	Thursday	5	2
11	20080111	2008-01-11 00:00:00.000	01/11/2008	11	Friday	6	2
12	20080112	2008-01-12 00:00:00.000	01/12/2008	12	Saturday	7	2
13	20080113	2008-01-13 00:00:00.000	01/13/2008	13	Sunday	1	2

Figure 4.4: DimDate

4.2.2.2. Create an ADF pipeline for the remaining Dim tables.

According to figure x, the author team builds a Pipeline for each Dim table.

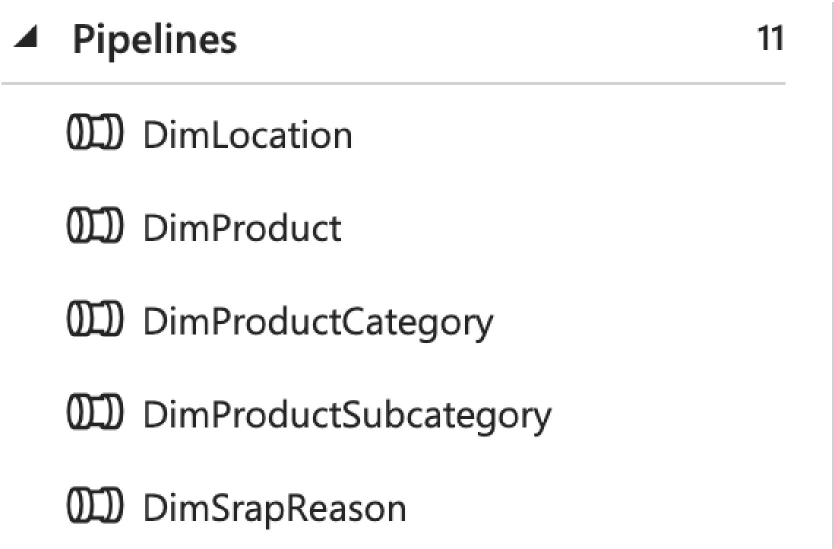


Figure 4.5: Pipelines for remaining Dim

The author team will use the prepared query to retrieve the Source for each Dim table from the Staging database (AdventureWorks2019) that has been uploaded to the server. Create a link service pointing to the staging database concurrently as figure x

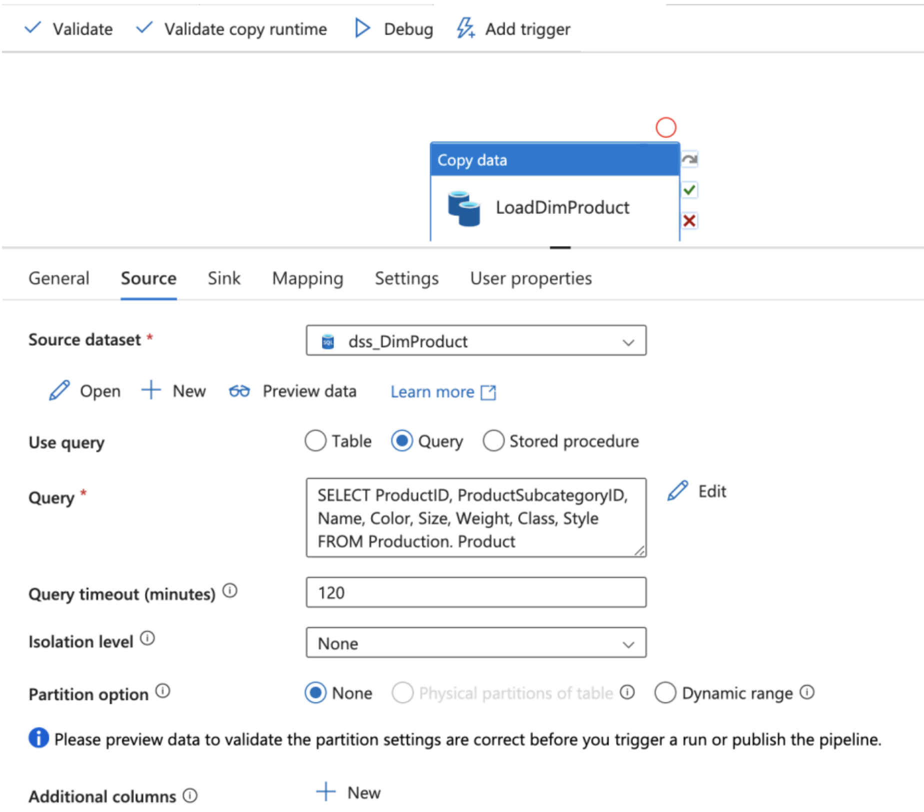


Figure 4.6: Load data for DimProduct

The group builds a link service to the server-created DW database on the Sink tab. Similar to figure 4.6, the remaining dim tables will also be used, but each dim's query script will be different.

4.2.3 Fact Table's ETL Process

Create an ADF pipeline for Fact table. The author team first develops a Pipeline, which is then dragged using the Data Flow activity, as seen in figure 4.7.

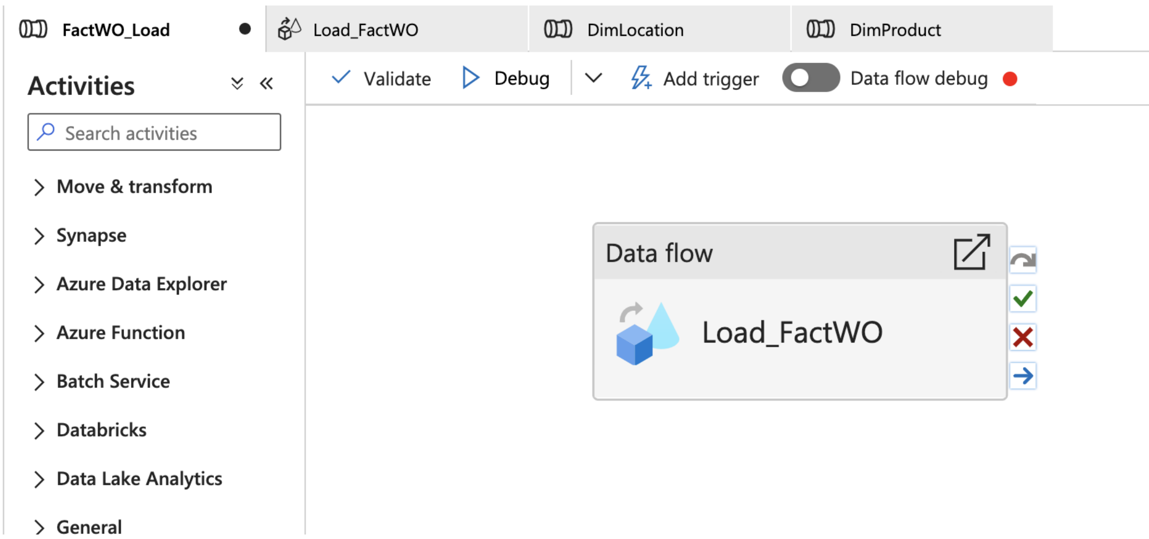


Figure 4.7: Load data to Fact_WorkOrder

Each appropriate Fact table in this stage is related to its own Dims. Figure 4.7 is shown brief screenshot of the dataflow of a Fact_WorOrder table

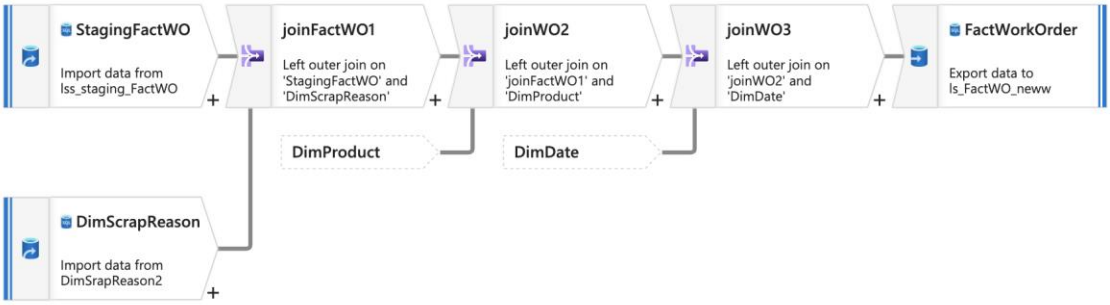


Figure 4.8: Dataflow

The team also ran a query to get data for Fact from the staging layer, same like it did for Dim tables as figure 4.8

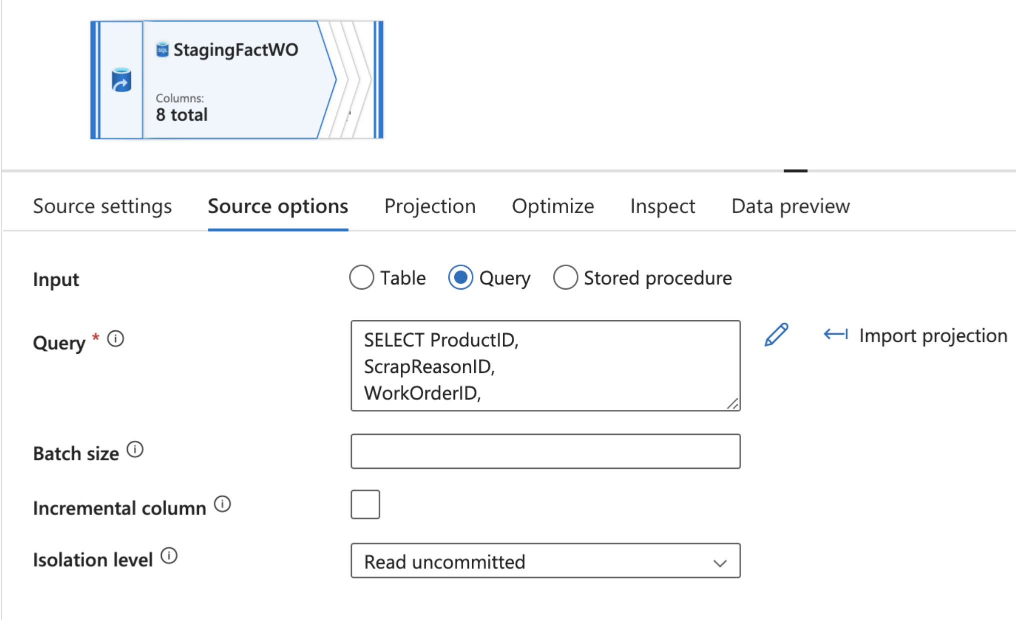


Figure 4.9: Putting information into the Fact_WorkOrder table

Before adding the data to the final Fact table in the DW database, the author team also utilizes the "JOIN" operation to connect the data of Fact and its associated Dim simultaneously.

As a result, the process of fully loading all data to serve as the input data for the data warehouse of the tac group has been finished.

4.3 Incremental Load

We'll be using data from DimProduct to create a new DimProductIncre table, which contains 504 rows, to serve the Incremental Load process for this research. Aspects like ProductKey, ProductID, ProductSubcategoryID, Name, Color, Size, Weight, Class, and

Style are included in this table. The ProductKey, commonly known as the old watermark, will be used to identify new or modified rows.

	ProductKey	ProductID	ProductSubcategoryID	Name	Color	Size	Weight	Class	Style
1	1	1	NULL	Adjustable Race	NULL	NULL	NULL	NULL	NULL
2	2	2	NULL	Bearing Ball	NULL	NULL	NULL	NULL	NULL
3	3	3	NULL	BB Ball Bearing	NULL	NULL	NULL	NULL	NULL
4	4	4	NULL	Headset Ball Bearings	NULL	NULL	NULL	NULL	NULL
5	5	316	NULL	Blade	NULL	NULL	NULL	NULL	NULL
6	6	317	NULL	LL Crankarm	Black	NULL	NULL	L	NULL
7	7	318	NULL	ML Crankarm	Black	NULL	NULL	M	NULL
8	8	319	NULL	HL Crankarm	Black	NULL	NULL	NULL	NULL
9	9	320	NULL	Chainring Bolts	Silver	NULL	NULL	NULL	NULL
10	10	321	NULL	Chainring Nut	Silver	NULL	NULL	NULL	NULL
11	11	322	NULL	Chainring	Black	NULL	NULL	NULL	NULL
12	12	323	NULL	Crown Race	NULL	NULL	NULL	NULL	NULL
13	13	324	NULL	Chain Stays	NULL	NULL	NULL	NULL	NULL
14	14	325	NULL	Decal 1	NULL	NULL	NULL	NULL	NULL
15	15	326	NULL	Decal 2	NULL	NULL	NULL	NULL	NULL
16	16	327	NULL	Down Tube	NULL	NULL	NULL	NULL	NULL
17	17	328	NULL	Mountain End Caps	NULL	NULL	NULL	NULL	NULL

Query ... | fanclubso16.database.window... | fanclubso16 (66) | DW_fanclubso16_0505 | 00:00:00 | 504 rows

Figure 4.10: DimProduct table

We will write the appropriate SQL code for the incremental load process after determining the use factor for the old watermark.

- Create watermarktable

```
create table watermarktable
(
WatermarkValue int,
)
```

Figure 4.11: Create Watermark table

To record the high-watermark value from the last data load, we build a Watermark table. The matching WatermarkValue in this table, which has the data type Int and saves the value since we have selected ProductKey as the identifying factor, stores the value.

- Get newwatermark value

```
Select MAX(ProductKey) as ProductKey from DimProductIncre
```

Figure 4.11: Get newwatermark value

For the next data loading procedure, we utilize this code fragment to get the new high-watermark value of the DimProductIncre table. To be more precise, this number will be utilized to only import data rows that are fresher than the watermark value into the target database.

- Create procedure to convert data

```sql
create schema desti

create proc desti.ProductIncre (
    @oldwatermark int,
    @newwatermark int)
as
begin
    select * from DimProductIncre
    where (ProductKey > @oldwatermark) and (ProductKey <= @newwatermark)
end
```

Figure 4.12: Create desti.ProductIncre procedure

This code creates a procedure used to perform the data loading process from the source table into the destination database. Specifically, it only retrieves the data rows with ProductKey values within the range of oldwatermark and newwatermark (meaning only retrieving the rows of data newer than the old watermark and not retrieving the data rows that have already been loaded into the destination database in the previous data loading process).

- Create procedure to update data

```sql
CREATE PROCEDURE update_watermark @ProductKey int
AS
BEGIN
    UPDATE watermarktable
    SET [WatermarkValue] = @ProductKey
END
```

Figure 4.13: Create update_watermark procedure

This code generates a procedure that may be used to update the latest watermark value of the data loading process into the watermarktable control table. The new value of WatermarkValue will be assigned with the value of the @ProductKey parameter, which is the new high-watermark value of the data loading process.

Updating this watermark value is necessary to ensure that the next data loading process will only load data rows that are newer than the latest watermark value, avoiding data duplication and ensuring data integrity in the target database.

- Insert data to test Incre Load

```sql
INSERT INTO DimProductIncre ( ProductID, ProductSubcategoryID, Name, Color, Size, Weight, Class, Style)
VALUES (1000,6,'Multiverse-2002 Blue 50','Blue','50',20.42,'L','U')
INSERT INTO DimProductIncre ( ProductID, ProductSubcategoryID, Name, Color, Size, Weight, Class, Style)
VALUES (1001,6,'Multiverse-2002 Red 40','Red','40',27.77,'L','W')
```

Figure 4.14: Insert data

This code snippet performs the insertion of new data rows into the source database table, which is DimProductIncre. These data rows will be prepared for the Incremental Load process into the destination database.
After all the necessary parts have been prepared, we will proceed to build the process and perform the Incremental Load on Azure.

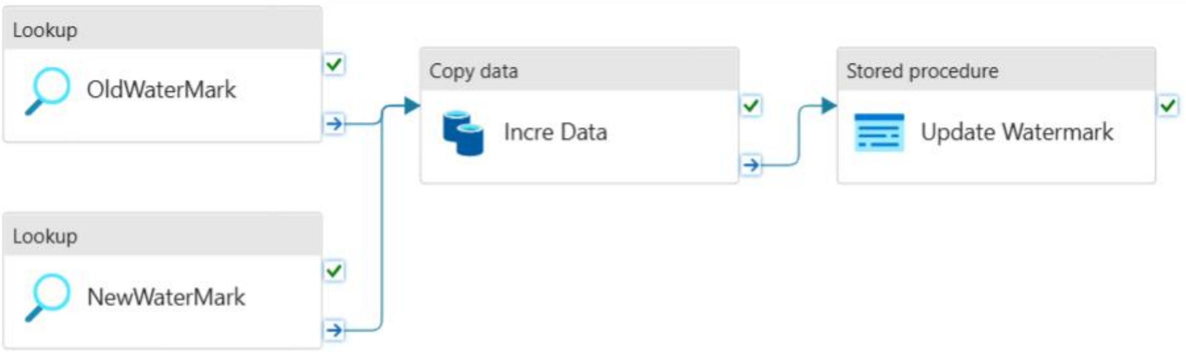


Figure 4.15 Incremental load Process

Figure 4.14 depicts our procedure, which has four fundamental steps: Step 1 creates a look up for the old watermark; Step 2 creates a look up for the new watermark; Step 3 creates a process to transfer newly updated data from the source to the destination database; Step 4 updates the watermark for future executions. Please see the parts below for a more detailed explanation.

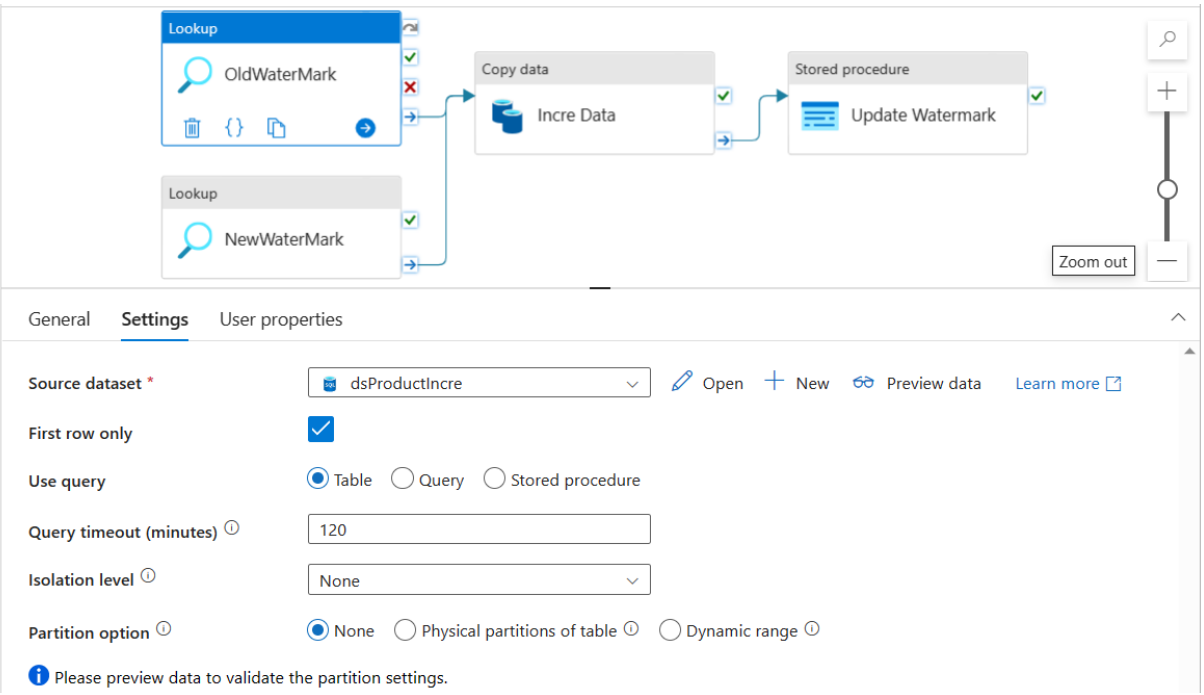


Figure 4.16: Set up Look up 1: Old watermark

At this step, we will use the code snippet in Figure 15: Create Watermark table, and perform an Insert operation to insert values into the watermarktable.

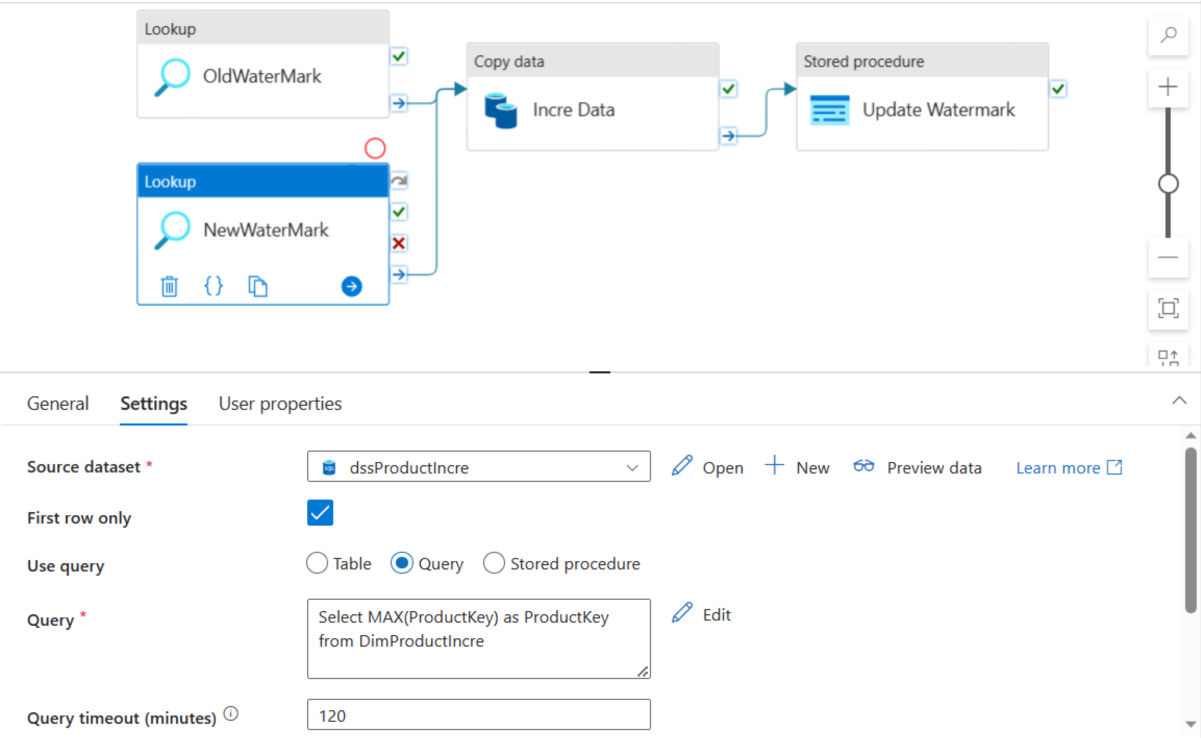


Figure 4.17: Set up Look up 2: New watermark

At this step, we will apply similar settings and add a query as shown in Figure 16: Get newwatermark value.

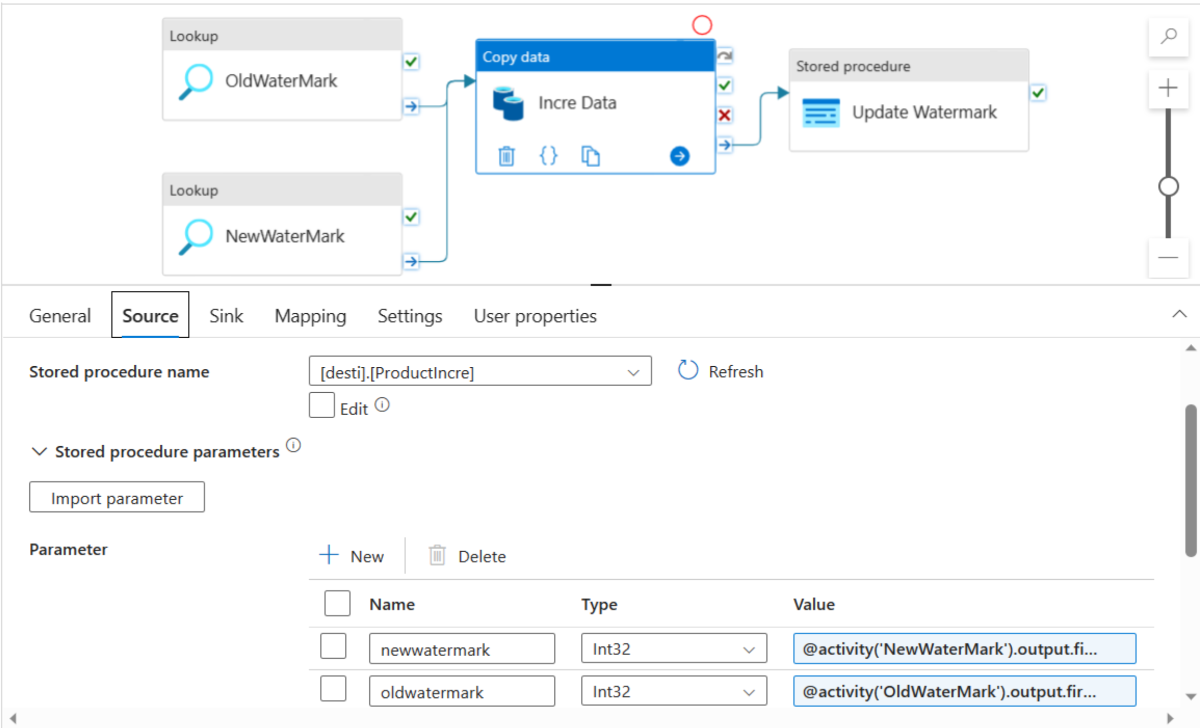


Figure 4.18: Set up converting data

In this step, we will transfer new data from the source database to the destination database. Before that, we need to create a procedure according to Figure 4.16: Create desti.ProductIncre procedure. And especially, pay attention to the source setup step, we have to set up the following parameters:

- @activity('NewWaterMark').output.firstRow.ProductKey
- @activity('OldWaterMark').output.firstRow.WatermarkValue

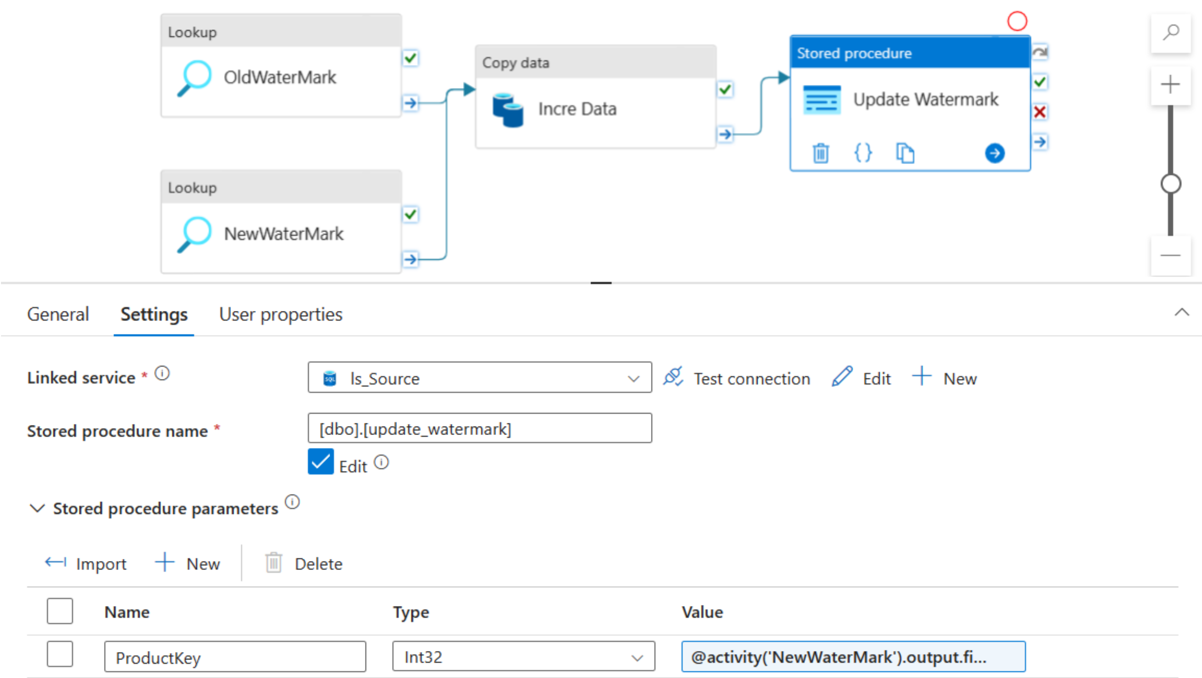


Figure 4.19: Set up updating watermark

In the final step, we will create a stored procedure to update the latest watermark for subsequent loads. Before that, we will also create a procedure as shown in Figure 4.17: Create update_watermark procedure. Then, we will continue to set up the parameter: @activity('NewWaterMark').output.firstRow.ProductKey for the process.

After completing the Incremental Load Process, we will test it with the data that has been prepared beforehand. Then we select Trigger Now and wait for the result.

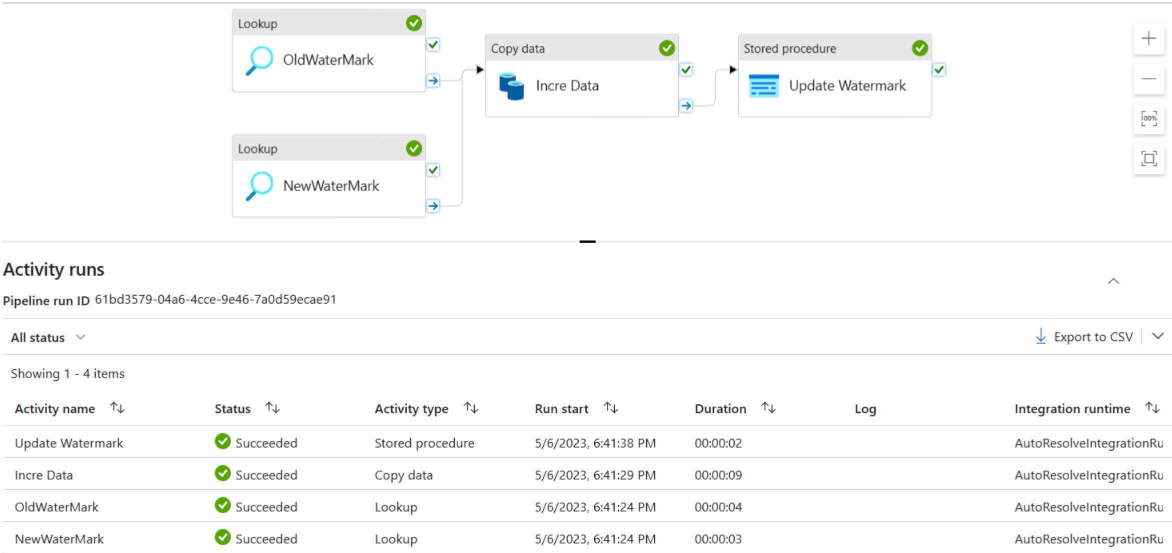


Figure 4.20: Successful Trigger

	ProductKey	ProductID	ProductSubcategoryID	Name	Color	Size	Weight	Class	Style
491	491	986	1	Mountain-500 Silver,...	Silver	44	28.13	L	U
492	492	987	1	Mountain-500 Silver,...	Silver	48	28.42	L	U
493	493	988	1	Mountain-500 Silver,...	Silver	52	28.68	L	U
494	494	989	1	Mountain-500 Black,...	Black	40	27.35	L	U
495	495	990	1	Mountain-500 Black,...	Black	42	27.77	L	U
496	496	991	1	Mountain-500 Black,...	Black	44	28.13	L	U
497	497	992	1	Mountain-500 Black,...	Black	48	28.42	L	U
498	498	993	1	Mountain-500 Black,...	Black	52	28.68	L	U
499	499	994	5	LL Bottom Bracket	NULL	NULL	223.00	L	NULL
500	500	995	5	ML Bottom Bracket	NULL	NULL	168.00	M	NULL
501	501	996	5	HL Bottom Bracket	NULL	NULL	170.00	H	NULL
502	502	997	2	Road-750 Black, 44	Black	44	19.77	L	U
503	503	998	2	Road-750 Black, 48	Black	48	20.13	L	U
504	504	999	2	Road-750 Black, 52	Black	52	20.42	L	U
505	505	1000	6	Multiverse-2002 Blu...	Blue	50	20.42	L	U
506	506	1001	6	Multiverse-2002 Red...	Red	40	27.77	L	W

Query executed successfully. | fanclubso16.database.window... | fanclubso16 (62) | DW_fanclubso16_0505 | 00:00:00 | 506 rows

Figure 4.21: Increment Load Result

The process was successful and produced the desired result. The 2 new data rows added to the source database were loaded into the destination database according to the specified process.

CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION

5.1 Report and dashboard systems (structure)

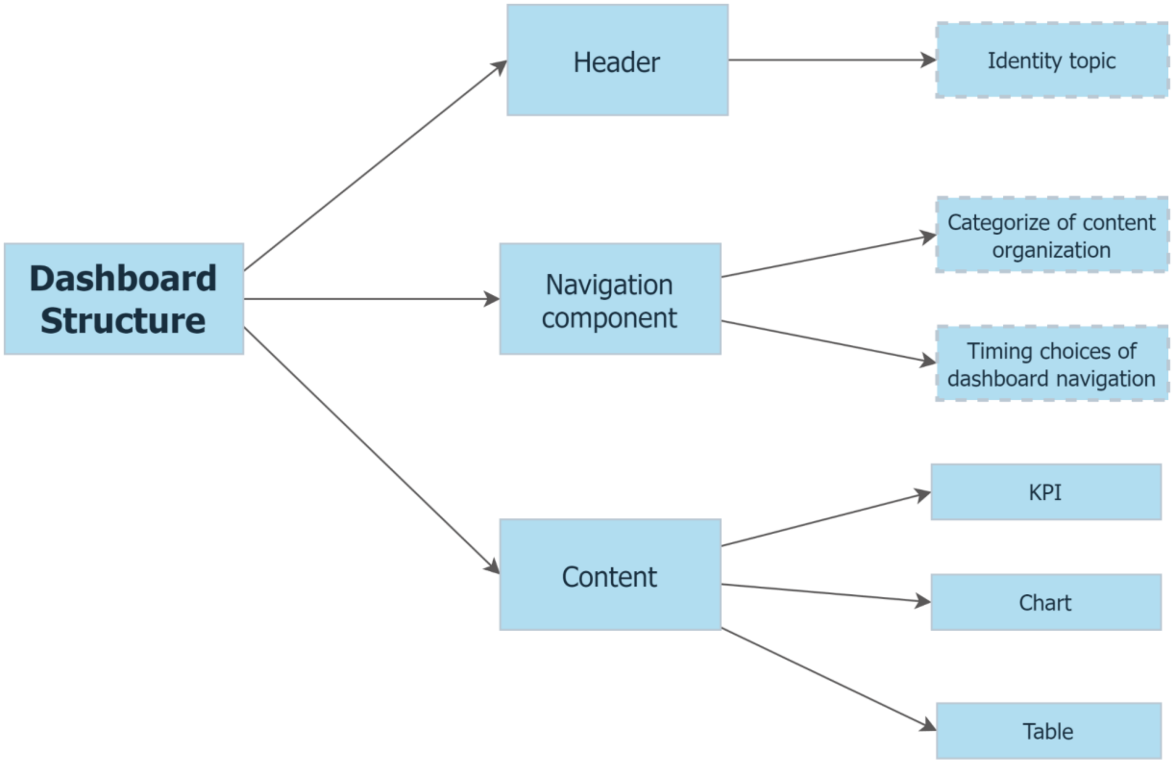


Figure 5.1: Dashboard structure

This dashboard template incorporates three basic components as in figure …. The first component is dashboard headers. The dashboard's theme is defined by the header. Above all, be sure to clarify for visitors what is being presented on the dashboard.

The second element is the selection of the navigational options and the categorization of content structure. In order to make navigating the dashboard easier, it could be possible to display the multitude of displayed indications on a single page, via a menu, or using tabs. It will also enable you to group indicators that have the same goal in order to maximize efficiency.

The dash-board's central element, or the content area that concentrates on indicators, is represented by the third component.There are three primary components to it. KPIs assist managers in monitoring operations, assessing performance, or assisting in making prompt choices. Charts give a better visual representation of what is going on within the company. Table, it is simple to compare the statistics in this table and even forecast future values, which makes it unique because it particularly depicts a significant event in the dashboard's theme.

5.2 Data analysis with Power BI

5.2.1 Dashboard Work Order

This dashboard provides an overview of the process data that creates product output in the manufacturing process to management levels. This dashboard's primary function is to assist managers in tracking progress and managing production output, manufacturing damage, and production orders.

The dashboard first displays the company's capacity for complete order fulfillment. We specifically create a line chart called "Manufacturing Efficiency" that displays the company's production performance from month to year. All data points are over 99.9%, which indicates that the capacity to accurately finish the volume of client orders is fairly excellent and stable. Customers can see the company's status and quality thanks to this.

Additionally, a bar chart titled "Scrap Ratio by Product Category" on the dashboard displays the typical product failure rate for each product category. The scrap ratio is a crucial metric for assessing the efficacy of a manufacturing process. It may be necessary to make adjustments to the production process if the scrap ratio is high in order to optimize efficiency and improve product quality. In this chart, this percentage is less than 1% for all Product Categories. This is excellent news for the business as a whole and the production process in particular.

Although the rate of product failure is quite low, there is nevertheless harm to the manufacturing process. More specifically, the "Total Scrap" card reveals that more than 11,000 faulty goods were produced in just three years. As a result, the dashboard also offers a "Top Scrap Reason" table, allowing managers to track and quickly identify typical mistakes occurring with their goods.

The "Production Quantity Overtime" chart also allows management levels to track the precise amount of goods produced over time. One can evaluate production growth patterns, peaks, variations, or potential issues by plotting production data over time using a light blue line. Additionally, because it can be challenging to determine how volatile the data is, we display the Moving Average with an orange line to help managers better understand the trend. We also present our prediction of the future production volume within a year by the dark blue line in order to aid in decision-making.

Finally, in the right corner of the dashboard as in …, a table displays production statistics and projections for each product throughout time.

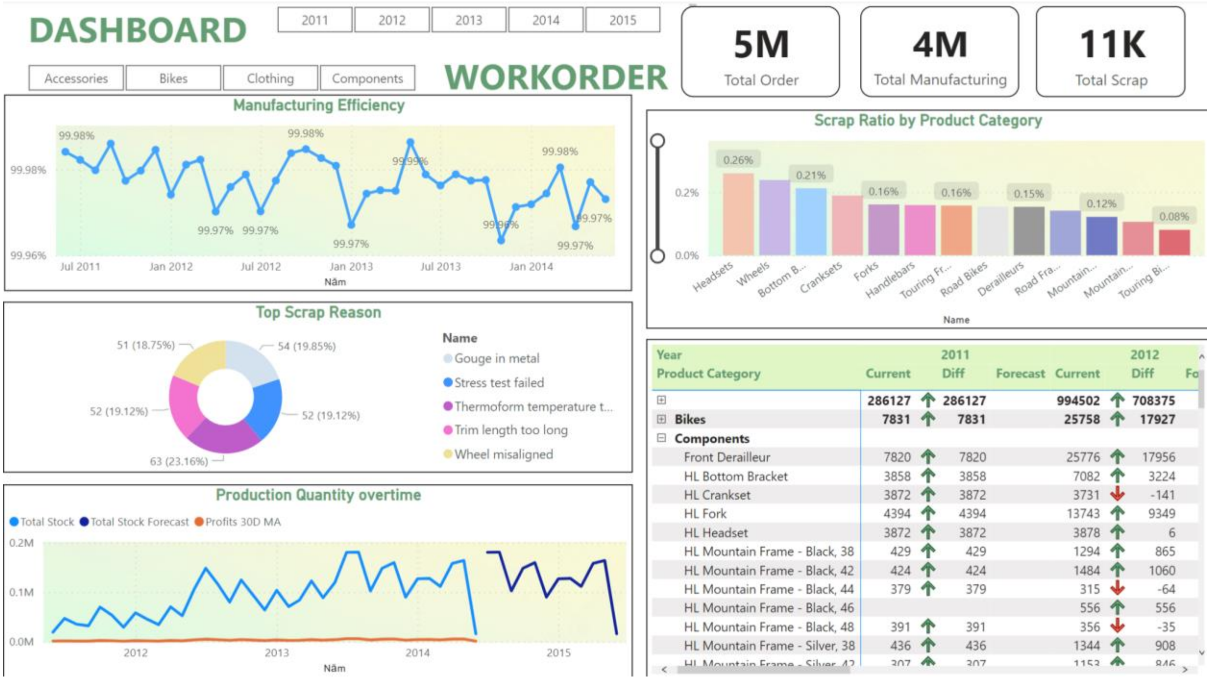


Figure 5.2: Dashboard Work Order

5.2.2 Dashboard Sequence Cost

The Sequence Cost Dashboard as figure 1 provides an overview of the production costs for each type of product, as well as the costs and labor hours required to satisfy each subcategory and category.

In this dataset, Actual Cost and Planned Cost have the same value. The first is the header of the dashboard, which displays the entire actual cost of manufacturing the product and the expected cost of making that product.

We can rely on the "Resource Hour" chart, which is shown below, to estimate the cost for the upcoming order by showing how long the product categorized by Subcategory will be produced. more exact. In addition to the "Manufacturing Efficient" graphic, it displays the total number of days needed to complete the product, broken down by subcategory, and the actual start date as opposed to the predicted start date. Whether or not there is a significant difference, modifications and changes will be made to policies as well as suitable production procedures to make sure everything can be done in a predictable manner and within the boundaries. great control

The top 5 manufactured products will have the largest cost when compared to the other products, as this chart will demonstrate for individuals who are interested in expenses. In order to identify the top items with high actual production costs and a lot of working hours from which strategies can be developed, it will also be integrated with the "Resource Hour" chart. If expenses frequently surge in comparison to other items, it may be possible to regulate as well as reevaluate the elements that contribute to this.

The last chart on the left, labeled "Forecast Cost," will help us gain a better understanding of the costs once more. It will show us how the costs have changed over time, and the moving average, which is an average line that represents the average cost values over a given period of time (in this case, 30 days), will make our predicted line more accurate, particularly the cost predictor for the entire year. We were able to anticipate production expenses for 2015 using the chart, so we may use that as a guide.

Finally, the data table in the right-hand corner gives us a thorough picture of the costs we incur by classifying the product's costs by SubCategory/Category and showing the costs from the previous year, this year, and the following year.

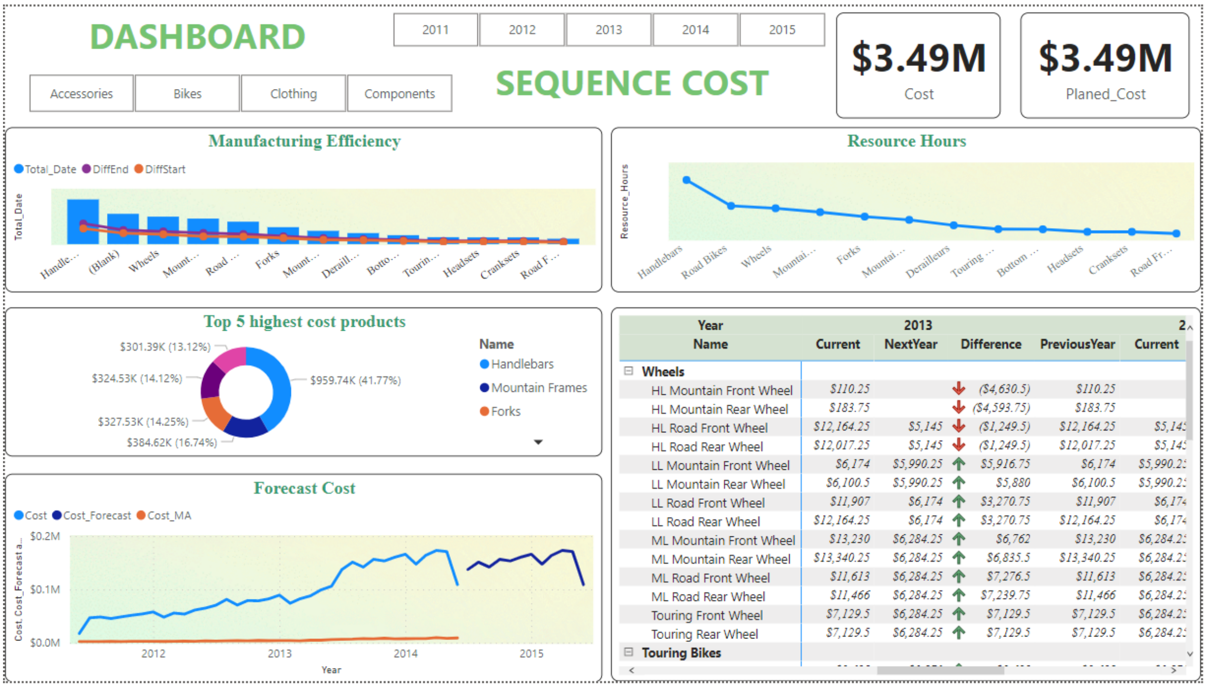


Figure 5.3: Dashboard Sequence Cost

5.2.3 Dashboard Inventory

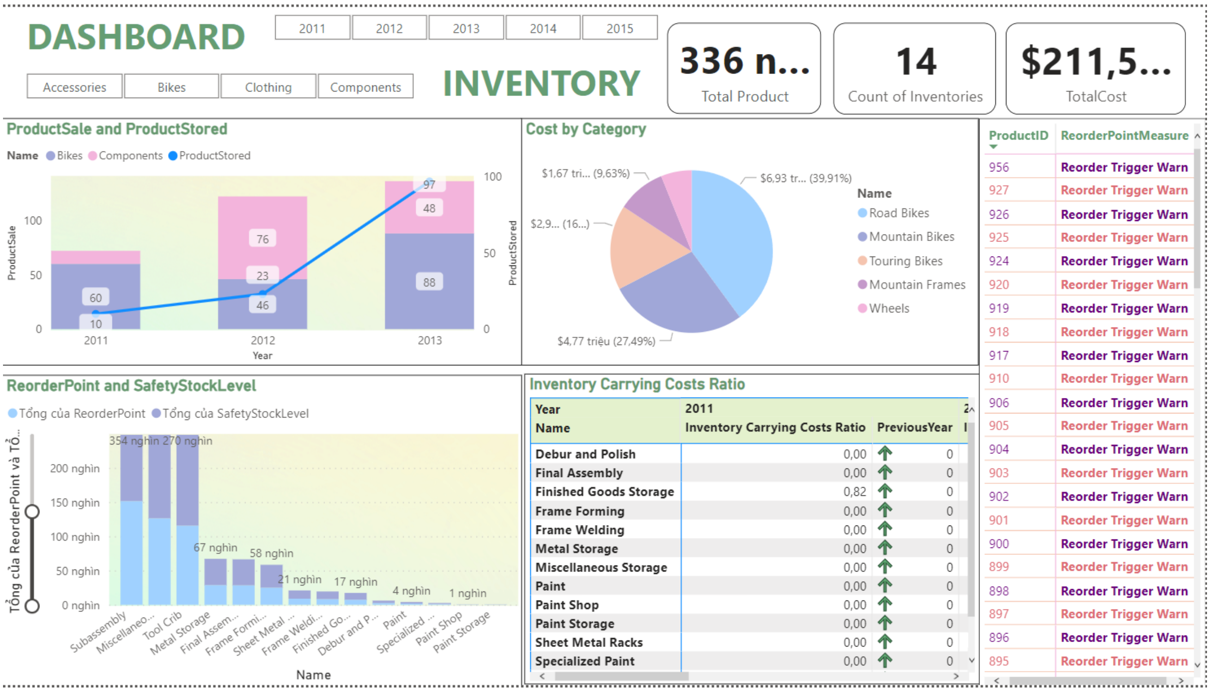


Figure 5.4: Dashboard Inventory

The dashboard Inventory shows the KPIs and the business features related to Inventory and Product from FactInventory. 336000 is the number of total products from all inventory in this database, and 14 is the number of inventories. With all that product and inventories, we received 211,500 thousand dollars for costs. It shows the trend of products that were sold and the number of products stored over the years, so we can have the insight of sales and holding inventories in these years. Inventory carrying Costs Ratio are the expenses associated with storing and maintaining inventory for a business. These costs include expenses such as warehousing, insurance, utilities, labor, equipment, taxes, and other expenses that are required to keep the inventory in a usable condition. The primary goal of inventory carrying cost management is to minimize the costs associated with inventory while maintaining sufficient stock levels to meet customer demand. By managing inventory levels effectively, a business can minimize the costs associated with inventory carrying and improve its profitability. With the number of products in inventories, we can group two categories: ReorderPoint and SafetyStockLevel in each inventory. With these KPI, we can know the number of products, the quantity of products in stock and control the quantity in stock to ensure the required quantity of products, maintain stock stock and enough products to sell to customers. To remain in business with inventory, we need to know if a product is nearly out-of-stock and there is no supply to ensure the quantity. Therefore, the table in the right of the dashboard shows a warning for a product if that product is nearly at the reorder point and needs to be refilled. The cost by category is shown by the top 5 group categories that have the highest costs overall and show it in detail.

5.3 Time series and Forecasting or Predictive model (Python)

Predicting future values provides firms a competitive edge in addition to helping them visualize business KPIs to enhance corporate decision-making. Additionally, we employ the Random Forest method, which mixes several decision trees to create a more potent prediction model, to address the issue of forecasting future values over time.
This method exhibits its strengths in prediction issues and is heavily employed in them. This approach may be used with any type of data, including time series data, which are defined by the reliance of the current value on earlier values. Because we can transform time series data into a supervised math problem, we can handle time series data as a supervised math problem. We can utilize upcoming observed values as output variables and previous observed values as input variables, to be more precise. The "sliding window" procedure enables the production of several fresh data samples for use in supervised mathematics. Using this method, we can forecast the value of future time series using traditional supervised machine learning models.
The number of things produced and production costs are the three issues for which we continue to create future value prediction models.

5.3.1 Production Quantity Forecasting

Businesses across a wide range of sectors value the ability to forecast product output. Businesses may better plan and allocate resources, increase their capacity to meet market demand, and minimize material, equipment, and supply waste by forecasting the amount of items to be produced. people and equipment.
Forecasting production output also assists companies in making crucial business choices, such as those involving investment, manufacturing growth, and hiring of human resources. If the projection is accurate, companies may take the appropriate actions that will boost production effectiveness, revenue growth, and profit.
The latest six months will be the test set, and the remaining data will be the training set, as shown in the figure... below, as we experiment with the prediction for this issue utilizing the Random Forest algorithm.

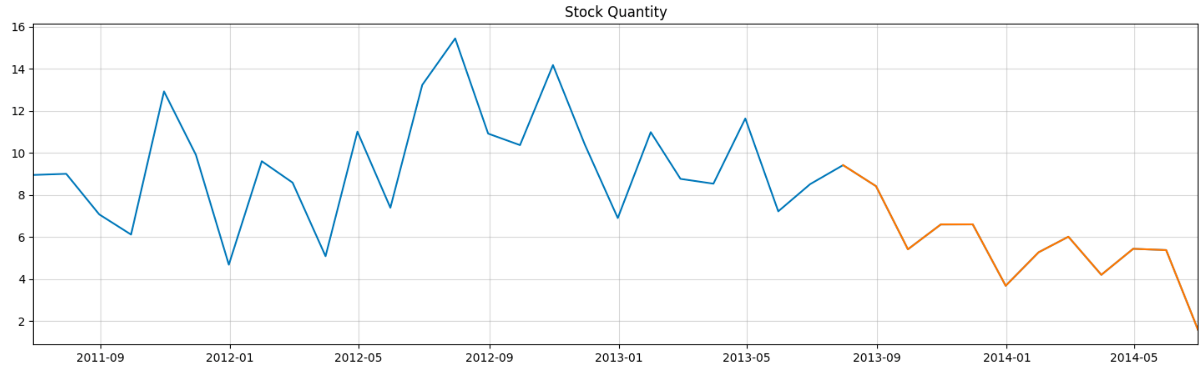


Figure 5.5: Production Quantity over time

Start the model training after normalizing and completing the input prerequisites. We discovered during the experiment that a six-step data delay is adequate, meaning that it is essential to take into account up to six prior values in order to forecast the future value. We achieve the same outcome as shown in the figure …..

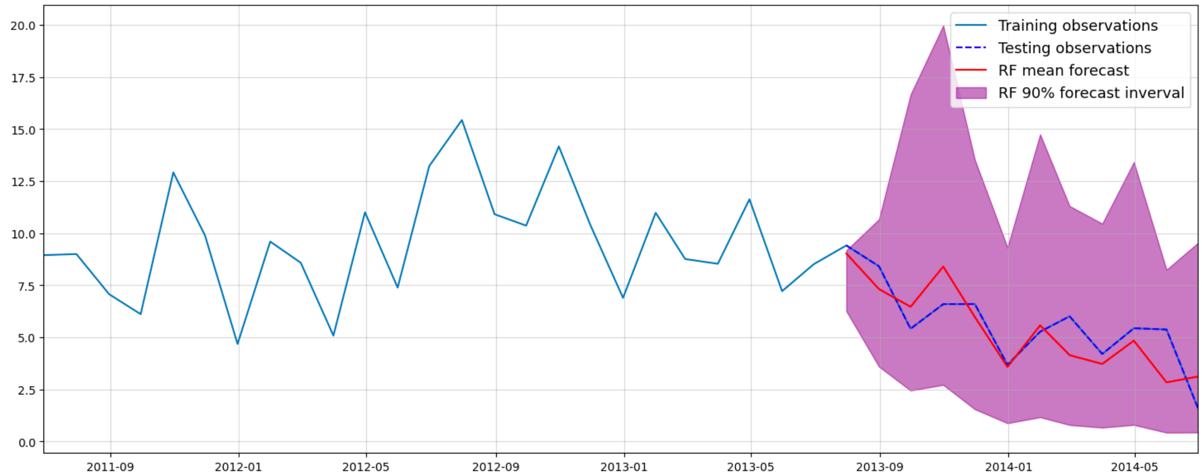


Figure 5.6 Predicting result

The test set's (red line) prediction results are pretty good, as can be shown. This outcome demonstrates that predictability is high and that the volatility of the data trend is quite consistent. In order to draw more precise conclusions from this model, let's look at the assessment findings based on the proper evaluation indicators:

Mean Square Error	**Root Mean Square Error**	**Mean Absolute Error**	**Mean Absolute Percentage Error**
1.588413344731723	1.2603227145186755	1.0312588517949557	0.01653443727184259

Table 5.1: Metrics evaluation.

The indications in the table are fairly accurate and demonstrate that the real model may be used to forecast reality. So let's attempt to anticipate the data for the upcoming 12 months outside of our dataset. The figure... below depicts the data's negative trend for the upcoming year.

Figure 5.7: Forecasting the next 12 months out of dataset.

5.3.2 Production Cost Forecasting

Predicting product production costs is essential to the business and product development process. Cost control and leakage prevention are also crucial if a company wants to grow sustainably. The more successfully firms can address issues like reducing waste, managing finances effectively, and allocating resources rationally, the more accurate the cost estimate.

Predicting product production costs generally helps businesses grow sustainably and compete in the market, as well as manufacturers calculate production costs.

We follow the same partition of data, training, and prediction procedures as the model in section 5.3.1.

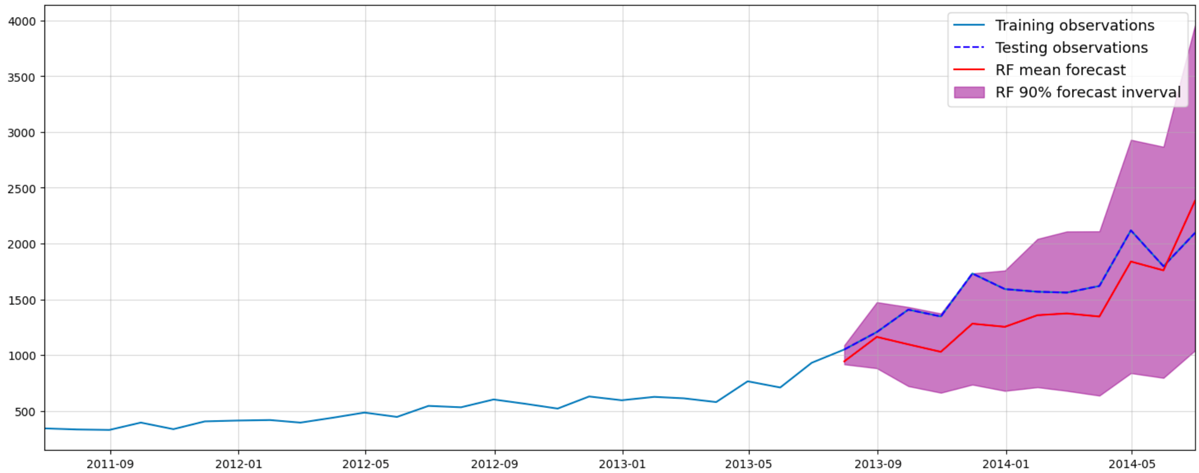


Figure 5.8: Predicting result

Mean Square Error	**Root Mean Square Error**	**Mean Absolute Error**	**Mean Absolute Percentage Error**
70258.9749547352 1	265.064095936690 76	236.701271598858 65	0.12540236306666 117

Table 5.2: Metrics evaluation

The figure … and table … show that the error between the actual value and the anticipated value is significantly different from the prior model. However, trend prediction still shares features with real data that are useful for making decisions.

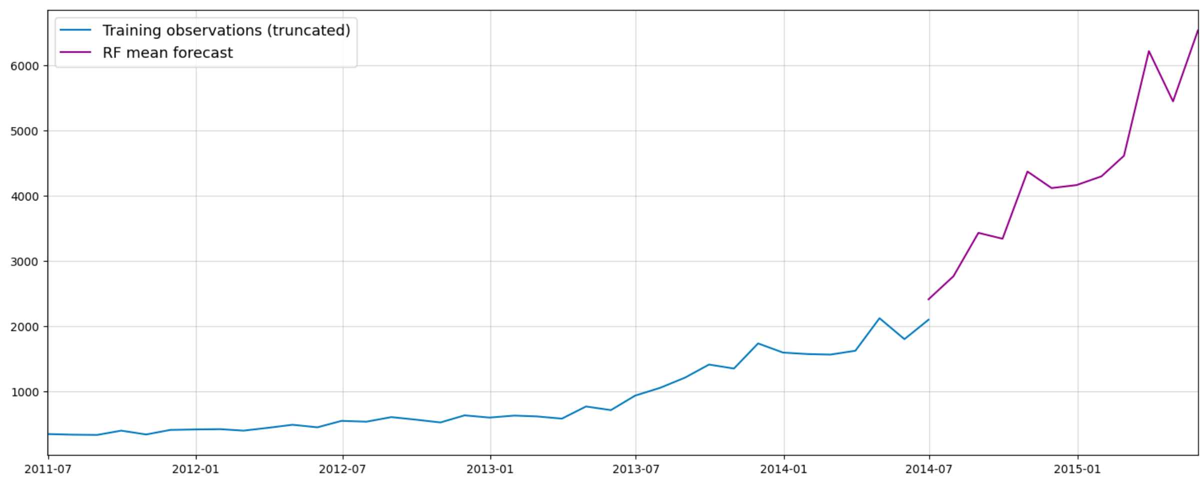


Figure 5.9: Forecasting the next 12 months out of dataset.

The data appears to have a strong increasing tendency for the upcoming year, as seen in the figure…. Managers will find it challenging since, according to the findings of the previous model, production volume appears to be down while expenses are rising. Additionally, this presents an analysis that calls for variable correlation. We could develop this subject more in the future.

5.4 Evaluation and Discussion

Dashboards are critical tools for organizations to use when making choices based on real-time data and information. Dashboards give a high-level overview of business activities and assist users in swiftly determining the present health of the firm. We expect that with such an easy-to-use design, it will assist businesses in making rapid decisions and assessing the present situation. Dashboards also enable users to monitor performance data and make required modifications to improve corporate success.

Furthermore, dashboards enable users to monitor risk data and implement risk-mitigation procedures. Simultaneously, the KPIs stated in this project are critical and truly significant in molding and making decisions for all levels of management. We showed them entirely on dashboards in a straightforward and simple manner.

With the original goal of creating a new and usable BI solution for Adventureworks' overall product manufacturing process. We believe that the information and indicators given and evaluated clearly in each dashboard will help to overcome the constraints of traditional process procedures and raise the company's operational efficiency.

We also did live future value prediction on the Power BI platform and created predictive models utilizing the Random Forest technique in Python, as previously described. However, the produced data are merely for reference and decision assistance, not for conclusive conclusions. The reason for this is because while Power BI is a robust platform for basic data visualization and analysis, it lacks in-depth prediction implementations about prediction. Assuming independent factors throughout time do not completely cover the model's influence and weight on company performance, we need to take the model into account in various contexts and variables for predictive models business in general. However, the findings, this results are still practical and should benefit businesses.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1 Results

First, by implementing the project of the subject, our team can apply knowledge from specialized subjects such as Databases, and Project Management Information System into reality. From there, grasp and better understand how to implement a BI solution into real business projects.

This project offers a different approach to typical manufacturing process management techniques. By replacing traditional storage techniques with a data warehouse designed expressly for this purpose. Furthermore, it delivers critical, basic analyses, measurements, and information for company assessment and decision-making. Simultaneously, displaying data using dashboards has helped to round out the overall image of our change proposal process. They are excellent instruments for successfully monitoring, evaluating, and managing resources in enterprises.

6.2 Limitations

Aside from the accomplishments, the project has the following limitations:

- Because our team did not completely comprehend the aims and challenges to be solved, we suggested goals that were too wide and difficult to accomplish with a BI solution, causing the first part of the project to be delayed.
- Furthermore, due to a lack of understanding of the Adventure Works 2019 data tables, our team struggled to create Dim and Fact tables and had issues when pouring data into these tables.
- The data warehouse has not integrated data relevant to warehouse storage and delivery.
- Because the project is still at the study and research stage, the execution of BI projects is still restricted, mostly to data representation; however, data analysis, data comparison, and data analysis are not yet in depth.

6.3 Future works

Creating and improving an enterprise BI solution takes time. As a result, we want each member of the group to be able to perfect themselves to the greatest extent possible. Complete the project's lacking prerequisites. In addition to the technical abilities required, we expect that in the near future, we will have the opportunity to experience genuine business difficulties in reality with the method chosen by the team, which helps determine which topics are truly vital and significant. From there, more appropriate indicators and indicators may be found to assist firms in management and decision-making.

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.

REFERENCES