UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEM

FINAL PROJECT

TOPIC: Operational Performance and Supply Chain Management

Subject: Business Intelligence and Decision Support Systems

Group:

Lecturer:

Ho Chi Minh,

Members of Group

No.	Student ID	Full name	Performance
1	K214162155	Nguyễn Phúc Thịnh	100%
2			100%
3			100%
4			100%
5			100%

Acknowledgements

Commitment

List of Figures

List of Tables

Table of content

Abstract

Chapter 1. About Project
This chapter introduces the research context, specifies the objectives, scope, and subjects of the research, the expected outcomes, as well as the tools and programming languages that will be used. The research focuses on inventory and demand management, optimizing purchasing operations, improving production performance, and enhancing sales results. The project utilizes the AdventureWorks database to develop key performance indicators (KPIs) and create analytical reports through Azure and Power BI. Azure tools such as Data Lake Storage Gen 2, Azure Data Factory, Azure Databricks, Azure Synapse Analytics, and Power BI will be applied to execute the research tasks.

Chapter 2. Related Research and Theoretical Basis
This chapter presents the theoretical basis necessary for the research and evaluates related studies. It provides a scientific foundation and synthesizes previous research results, serving as a basis for the development and implementation of the current research. Additionally, this chapter outlines the concepts of the services that the project will use, providing an overall view of each implementation phase of the project.

Chapter 3: Business Requirements Analysis
This chapter begins with an introduction to the bicycle company AdventureWorks Cycles and identifies the company's requirements for inventory and demand management based on KPIs. It provides a detailed description of the AdventureWorks database modules, focusing on Production, Purchasing, and two tables from Sales. Finally, this chapter describes the process of selecting and preparing the necessary data from AdventureWorks, ensuring that the data is ready for subsequent analyses.

Chapter 4: Integrating Data and Building Data Warehouse
This chapter describes the data ingestion processes into the data lake from the SQL database to the bronze, silver, and gold data layers. It explains the rationale for choosing the galaxy schema and provides detailed descriptions of the dimension and fact tables in the data warehouse model. Finally, this chapter presents the results after building and loading the data into the data warehouse.

Chapter 5: Visualization By Power BI
The final chapter presents the research results through data visualization using Power BI. It illustrates how the data is transformed into intuitive charts and reports to support decision-making processes. The KPIs will be highlighted to help managers grasp the company's situation.

Chapter 1. About Project

1.1 Business Case:

In the competitive landscape of today's markets, businesses must prioritize efficient inventory management and effective demand management. These factors are crucial for maintaining cost-efficiency and ensuring customer satisfaction. Traditional systems often struggle to adapt to the dynamic nature of market demands and supply chain variables, leading to inefficiencies. Adopting a cloud-based solution such as Azure presents a strategic approach to addressing these challenges, leveraging the latest in technology to enhance operational capabilities.

Effective inventory and demand management is essential for minimizing capital locked in excess inventory and preventing stock outs that lead to lost sales. The dynamic nature of consumer demands and supply chain challenges requires a system that not only accommodates real-time changes but also integrates seamlessly with existing operational frameworks to enhance decision-making processes. The need for a solution that offers scalability, flexibility, and real-time data access underscores the importance of this business case.

The Azure Cloud Platform offers a compelling suite of features that make it an ideal choice for businesses aiming to enhance inventory and demand management. Central to its strengths is enhanced data management, where Azure enables efficient handling and integration of data. This allows businesses to gain a holistic view of inventory levels and product requirements across all channels, thereby maintaining optimal inventory levels that are responsive to market demands. Additionally, Azure's cloud-based model offers significant cost efficiency by reducing the need for large capital investments in IT infrastructure. Businesses benefit from Azure’s scalable services, paying only for the resources they use, which is particularly advantageous in fluctuating market conditions.

Moreover, Azure enhances operational agility through real-time data processing and advanced analytical tools, allowing businesses to rapidly adapt to market changes. This capability is essential for efficiently managing inventory and adjusting procurement strategies. The platform’s scalability and flexibility further empower businesses, enabling them to adjust operations based on demand patterns and explore new markets without the limitations of physical infrastructure. Finally, Azure's commitment to security and regulatory compliance ensures that sensitive customer and operational data are protected, meeting industry standards and fostering a secure business environment. These features collectively make Azure a robust and strategic choice for businesses looking to optimize their inventory management and effectively manage product demand.

Implementing Azure for enhanced inventory and demand management is a strategic initiative that aligns with the goals of reducing operational costs, avoiding inventory mismanagement, and improving customer satisfaction. This business case is pivotal for companies looking to leverage technology to streamline operations and respond adeptly to market changes. Azure's comprehensive suite of tools and scalable solutions provides a robust platform for businesses to optimize their inventory management practices and effectively manage product demand. The adoption of Azure not only supports day-to-day operational needs but also contributes to long-term business growth and sustainability in a competitive marketplace. This approach not only optimizes resource allocation but also positions the company for future success by enhancing operational flexibility and responsiveness.

1.2 Objectives

Based on the business context and current challenges regarding inventory management and demand, the specific objectives for managing operational performance and optimizing the supply chain are defined as follows:

(1) Optimize Purchasing Operations:

- Cost Reduction: Reduce procurement costs by selecting and collaborating with reliable suppliers, thereby enhancing the financial efficiency of purchasing operations. This is monitored through the KPI "Purchasing Cost per Unit," reflecting the cost per product purchased.

- Enhanced Order Processing Efficiency: Reduce the time from order creation to receipt, enhancing efficiency and rapid response, aligning with the flexible response model of Azure. This is measured by the "Order Processing Time" KPI, emphasizing the system's ability to adapt to rapid market changes.

(2) Improve Production Performance:

- Optimize Production Scheduling: Improve production schedules to minimize downtime and increase productivity, aligned with Azure's advanced analytical tools, helping to more accurately predict production needs. The related KPI is "Production Cycle Time," tracking the operational efficiency of production time.

- Reduce Defect Rate: Focus on reducing the rate of defective or scrapped products through process and quality improvements, tracked by the "Scrap Rate" KPI, assessing the effectiveness of quality management and waste reduction.

(3) Enhance Sales Outcomes:

- Effective Inventory Management: Improve the ability to predict sales trends and optimize inventory management through the predictive technology and data analysis support of Azure, minimizing costs and avoiding overstocking or stockouts. This relates to the "Order Fulfillment Rate" KPI, reflecting effective coordination between departments.

- Reduce Return Rate: Enhance product quality to reduce the return rate, thereby increasing customer satisfaction and brand reputation, monitored through the "Return Rate" KPI.

These objectives aim to create a robust framework for enhancing operational effectiveness across purchasing, production, and sales, compatible with the flexibility and scalability of the Azure platform. Implementing Azure not only supports daily operational needs but also contributes to long-term business growth and sustainability in a competitive market.

1.3 Objects and Scopes

1.3.1 Objects

Inventory of Bicycle-Related Products in the AdventureWorks Dataset: Statistics and analysis of the quantity and cost of inventory products. The goal is to control inventory products to balance supply and demand, reduce storage costs, and ensure timely response to market demand.

Product Demand Management: Managing the demand for bicycle products, reflected through orders and consumer trends. The goal is to use historical sales data to forecast demand, helping to adjust production and manage inventory effectively to meet market demand.

1.3.2 Scopes

- **Space**: University of Economics and Law
- **Time**: 15/04/2024 - 18/05/2024

1.4 Expected output of the project

This project will utilize the Adventure Works Model Production database to develop key performance indicators (KPIs) that are aligned with our strategic objectives for

optimizing operational performance and supply chain management. By leveraging this rich data source, we will create comprehensive reports and analytical dashboards that provide insights into purchasing efficiencies, showcased through the "Purchasing Cost per Unit" and "Order Processing Time" KPIs. Similarly, improvements in production will be evident from the enhanced "Production Cycle Time" and reduced "Scrap Rate," highlighting better scheduling and quality management. Additionally, the outputs will include analysis of the "Order Fulfillment Rate" and "Return Rate" KPIs, reflecting advancements in inventory management and product quality. These integrated insights will support our use of Azure's technology to enhance operational responsiveness to market demands and elevate customer satisfaction, thereby fostering long-term growth and sustainability in a competitive marketplace.

1.5 Usage services

In this study, the team executed tasks using services provided by the Azure platform, such as Data Lake Storage Gen 2, Azure Data Factory, Azure Databricks, Azure Synapse Analytics, and Power BI.

Data Lake Storage Gen 2: In our project, it is organized according to a 3-tier model (Bronze, Silver, and Gold), with each tier representing a different level of data cleansing. The Bronze tier is where data is ingested without any special cleansing process. The Silver tier is where data undergoes basic cleansing to remove duplicates and verify integrity. The Gold tier is the highest level, where data has been fully cleansed and is ready to be loaded into a Data Warehouse or other data analytics platforms. This tiered approach optimizes the data processing workflow, providing a flexible database to support the project's data analysis and management needs.

Azure Data Factory (ADF): Using ADF (Azure Data Factory) to create pipelines automates the data transfer through the three tiers of the data lake. These pipelines act as a control mechanism, transferring data from the Bronze tier, through Silver, and finally to Gold. This ensures that data is always up-to-date and ready for reporting and analysis.

Azure Databricks: Azure Databricks is used to perform complex data transformation tasks. This platform enables tasks such as scientific computing, machine learning, and data optimization with high scalability. The transformation of data from the Bronze tier to Silver is primarily carried out here.

Azure Synapse Analytics: After the data has been transformed and optimized, Azure Synapse Analytics is used to create efficient views and queries. The data is then pushed to the SQL Serverless pool. This allows for analysis on a large volume of data without the need to manage server resources.

Power BI: Finally, Power BI is used to visualize and create dashboards from the refined and optimized data. These reports and dashboards provide a visual and insightful view of the data, enabling managers and stakeholders to make decisions based on accurate and up-to-date information.

Chapter 2. Related Research and Theoretical Basis

2.1 Related Researches

- A Data Warehouse design for the detection of fraud in the Supply Chain by using the Benford's law

This study was conducted with the aim of developing a data warehouse design to support forensic analysis using Benford's Law to detect fraud in supply chain management. The researchers tested this design by applying it to two datasets from an active supply chain database, and evaluated the effectiveness of the method in identifying fraud.

Highlights of this study include the innovative data warehouse design that allows for the reuse of stored procedures to analyze data without the need for reinstallation, thereby minimizing the time and effort required for analysis. The study also effectively leverages Benford's Law to analyze the first digits of the data, a method proven to be effective in indicating anomalies that may be related to fraudulent behavior. Furthermore, although the study focuses only on inventory and warranty management, the proposed data warehouse design has the potential to be broadly applied to other processes in the supply chain, enhancing its generalizability and applicability.

However, this study also has several limitations. Its scope is primarily limited to inventory and warranty management, which may not be sufficient to reflect other aspects of the supply chain. The effectiveness of Benford's Law may also be influenced by the quality and nature of the input data, potentially reducing the accuracy of the method in certain cases. To confirm the feasibility and effectiveness of the data warehouse design, further research is needed on other processes in the supply chain and with larger datasets.

Although further exploration is necessary, this study provides a significant step forward in developing tools capable of detecting fraud in supply chain management, expanding the applicability of data warehouses in the field of digital forensics and data analysis.

- Data Warehouse success lead towards Supply Chain Efficiency

This study was conducted to assess the impact of data warehouse success on supply chain efficiency in companies in Indonesia, focusing on factors such as system quality, information quality, service quality, and relationship quality. The study indicates that the success of a data warehouse is positively related to supply chain efficiency, thereby enhancing the productivity and operational effectiveness of companies.

The highlights of this study include the emphasis on the role of system quality, information quality, service quality, and relationship quality in ensuring the success of a

data warehouse, which in turn improves supply chain efficiency. Findings from the study show that when these factors are well-managed, they not only contribute to the success of the data warehouse but also promote overall business efficiency. Although the study provides valuable insights, it has certain limitations such as the geographic scope being limited to Indonesia and data collection primarily from warehouse staff, which may not fully reflect the influencing factors from other departments within the organization. The study suggests that companies need to ensure a robust data warehouse system to drive supply chain efficiency, and further research is needed to verify the feasibility and effectiveness of data warehouse design on a broader scale and in different contexts.

- Advancing Logistics 4.0 with the Implementation of a Big Data Warehouse: A Demonstration Case for the Automotive Industry

This study presents the implementation of a Big Data Warehouse (BDW) to promote the Logistics 4.0 movement in the automotive industry. The document details an approach to improving the data analytics infrastructure of organizations, enabling them to effectively harness large volumes of data through Big Data and Machine Learning technologies. The research provides a detailed architecture for storing, managing, and processing data using technologies such as Hadoop and Spark, and illustrates this application through a real-world case study at a multinational organization.

While this study offers numerous benefits such as enhanced data analytics systems and practical insights from a specific application case, it also poses significant challenges. Implementing a BDW system requires substantial technical and financial resources, which not all organizations, particularly smaller ones or those with limited resources, can afford. Additionally, the complexity of the architecture and technologies can increase training and deployment costs, as well as the demand for high technical expertise.

In summary, this study provides a deep insight into the potential of Big Data to improve logistics in Industry 4.0, while also highlighting the cost and technical barriers. It is crucial for organizations to carefully weigh the benefits against the costs when deciding to invest in such technologies and be prepared to face the challenges during the implementation process to fully leverage the capabilities of Big Data.

- The Snowflake Elastic Data Warehouse

This study introduces Snowflake Elastic Data Warehouse, a highly scalable, flexible, multi-user, transactional, and secure data warehouse system with full SQL support and integrated extensions for semi-structured and schema-less data. Snowflake is offered as a pay-per-use service on the Amazon cloud, allowing users to upload data to the cloud and manage and query it instantly using familiar tools and interfaces.

Snowflake is designed to completely separate storage and compute capabilities, enabling flexible and immediate scaling of compute resources without affecting data availability or the performance of concurrent queries. It is highly fault-tolerant, with no downtime during hardware or software upgrades, and provides extremely high data durability. Additionally, the system efficiently supports semi-structured data with automatic schema detection and columnar storage.

While Snowflake offers significant benefits, it also has certain limitations. Specifically, being designed to run entirely on the cloud may not suit organizations with extremely stringent data security requirements, as data must be stored outside the organization's physical premises. Moreover, although Snowflake provides auto-scaling based on usage demands, costs can become substantial as data scales up rapidly, particularly with large, continuous data loads.

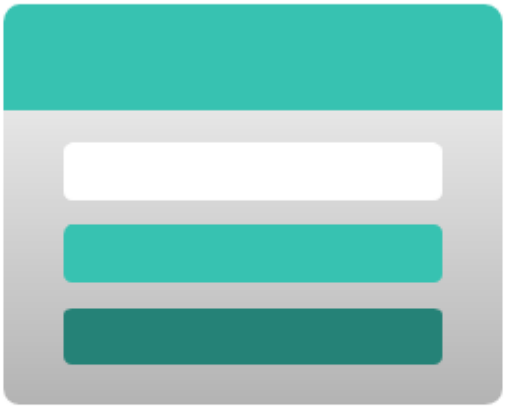
Snowflake represents a significant advancement in providing data warehouse solutions for the cloud era, offering high flexibility, scalability, and efficiency without requiring users to delve deeply into system configuration. This presents considerable advantages for organizations looking to leverage big data and analytics in a cloud environment.

2.2 Theoretical Basic

2.2.1 Overview of Azure Services

Azure Storage Account

Azure Storage Account is a comprehensive cloud storage service providing scalable, durable, and secure storage solutions for a wide range of data types, including blobs, files, tables, and queues. It is designed to accommodate both structured and unstructured data, crucial for diverse business applications ranging from simple data storage to complex data analytics platforms (Microsoft, n.d.).



Azure Storage Account

Figure 2.1 - Azure Storage Account's icon

Key Features:
- *Durability:* The service ensures data safety through redundant storage, replicating data across multiple datacenters, thus guaranteeing high availability and disaster recovery capabilities (Microsoft, n.d.).
- *Scalability:* Azure Storage Account can dynamically scale to meet the growing data needs of an organization, ensuring that storage capacity is neither a bottleneck nor an excess cost factor (Microsoft, n.d.).
- *Security:* It incorporates advanced security features to protect data against unauthorized access and threats, making it a reliable choice for enterprises concerned with data security (Microsoft, n.d.).

Use for Data Lake: Given its ability to handle large volumes of unstructured data, Azure Storage Account serves as an ideal foundation for data lakes. This capability supports the storage of massive datasets in their native format, which is essential for big data analytics applications (Microsoft, n.d.).

Azure SQL Database

Azure SQL Database is a managed, cloud-based database service built on SQL Server technology. It offers high SQL compatibility and optimized performance for SQL-based

applications, making it a prime choice for organizations looking to migrate or extend their databases to the cloud (Microsoft, n.d.). Some benefits of it can be list down here:
- ***Ease of Use:*** The service simplifies database management by providing automated updates, backups, and scalability options, allowing teams to focus more on application development rather than database administration (Microsoft, n.d.).
- ***Scalability:*** It offers the flexibility to scale resources up or down based on demand, thereby aligning operational costs with actual usage (Microsoft, n.d.).
- ***Security:*** Azure SQL Database provides robust security features, including automated threat detection and secure network connectivity, essential for protecting sensitive business data (Microsoft, n.d.).

Store Structured Data: This service is particularly suited for applications that require structured data storage, efficient querying capabilities, and reliable performance, such as customer relationship management and enterprise resource planning systems (Microsoft, n.d.).

Azure Synapse

Azure Synapse Analytics is an integrated analytics service that merges big data and data warehousing into a single service. This powerful platform enables the storage, processing, and analysis of large datasets, facilitating quick insights and informed decision-making (Microsoft, n.d.). Some key features:
- ***Supporting Big Data Analytics:*** It can process large volumes of data swiftly, providing valuable insights through analytical models and business intelligence tools (Microsoft, n.d.).
- ***Real-time Analytics:*** The service supports real-time analytics, which is critical for operational reporting and timely decision-making (Microsoft, n.d.).
- ***Inventory and Product Demand Management****:* Azure Synapse plays a pivotal role in managing inventory and product demand by analyzing large datasets, helping businesses optimize their operations and improve profitability (Microsoft, n.d.).

Azure Synapse offers two distinct types of SQL Pools - SQL Serverless Pool and Azure Synapse offers two distinct types of SQL Pools—SQL Serverless Pool and SQL Dedicated Pool. Each type serves specific purposes based on performance needs, cost, scalability, and reliability, making them suitable for different aspects of data handling and querying within cloud environments.

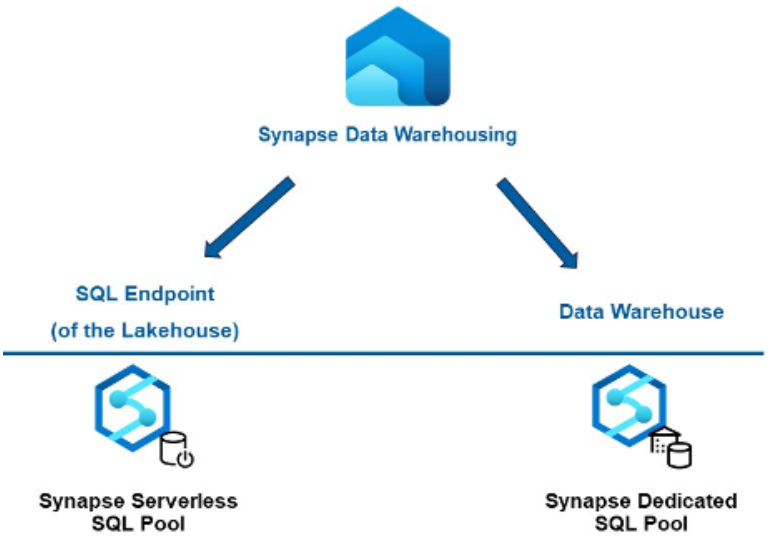


Figure 2.2 - SQL Serverless Pool and SQL Dedicated Pool

SQL Serverless Pool:
- Flexibility: Allows querying data from various sources and serving data to applications.
- Advantages: User-friendly, cost-effective, and automatically scalable. Ideal for less complex querying needs and for dynamically querying and accessing data.
- Disadvantages: Less suitable for applications requiring high query performance or high data reliability.

SQL Dedicated Pool:
- Performance: Provides high performance for complex queries.
- Reliability: Offers very high data reliability, essential for critical data operations.

Table 2.1 - Comparison of SQL Serverless Pool with SQL Dedicated

Feature	**SQL Serverless Pool**	**SQL Dedicated**
Scalability	Automatically scales	Requires manual configuration
Cost	Pay-as-you-go	Pay for storage capacity
Performance	Good for simple queries	High performance for complex queries
Reliability	High	Very high

Our team will utilize the SQL Serverless Pool to create views from the data stored that reference the structured and processed data in the Gold layer. This approach leverages the

flexibility and cost-effectiveness of the SQL Serverless Pool, making it well-suited for our project needs, enabling efficient data access and manipulation for analysis and reporting.By integrating SQL Serverless Pool in this manner, we aim to enhance our project's efficiency and flexibility in data management and analysis, achieving cost-effective scalability without compromising on the ability to meet dynamic data querying needs.

Power BI

Power BI is a business analytics service that provides interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards (Microsoft, n.d.). Some integration and features:

- Integration with Azure Services: It integrates seamlessly with various Azure services, enhancing the analytics capabilities of Azure Synapse, Azure SQL Database, and more (Microsoft, n.d.).

- Data Visualization and Analytics: Power BI enables users to convert data into intuitive graphical reports and dashboards, facilitating an easier understanding of complex data (Microsoft, n.d.).

- Business Decision Making: By providing deep insights into data, Power BI assists in making informed business decisions that can significantly impact an organization's success (Microsoft, n.d.).

Azure Databricks for Big Data Transformation

Azure Databricks is a cloud-based big data analytics platform optimized for the Microsoft Azure cloud services platform. It provides a collaborative environment using Databricks notebooks that support multiple languages. Azure Databricks integrates deeply with Spark, offering enhanced functionality and allowing for robust big data processing.

Big Data and File Formats

Big Data refers to massive datasets that are complex and difficult to process using traditional data processing software. For effective storage and processing, we utilize specialized file formats:

- Parquet: A columnar storage file format optimized for fast retrieval and efficient data compression. In our data storage system, Parquet is primarily used in the Bronze and Gold layers, where it ensures efficient storage and quick access to historical data.

- Delta Lake: Builds upon Parquet to add ACID transactions, enabling reliable data integrity through Atomic, Consistent, Isolated, and Durable (ACID) properties. Delta Lake is used in the Silver layer, where data undergoes various transformations and requires robust consistency and version control.

Apache Spark and PySpark

- Apache Spark: An open-source unified analytics engine for large-scale data processing. Spark provides high-performance data processing with capabilities for in-memory computing and real-time processing.
- PySpark: The Python API for Apache Spark, allowing data scientists and analysts to write Spark code using Python. This is particularly useful for teams familiar with Python, enabling them to leverage Spark's capabilities without needing to switch to Scala or Java.

Utilization in Azure Databricks
In our project, Azure Databricks is leveraged for its seamless integration with these technologies. It enhances the efficiency of transforming and processing Big Data - Data Transformation: Using PySpark in Azure Databricks, we can perform complex data transformations more effectively compared to Synapse Notebooks or Azure Data Factory (ADF). This is due to the powerful in-memory processing capabilities of Spark, which are well-suited for handling the large volumes of data typical of our project.

By adopting Azure Databricks combined with Parquet, Delta Lake, and Spark (via PySpark), our project enhances the agility and efficiency of data operations, supporting robust Big Data analytics in a scalable cloud environment.

2.2.2 Understanding Data Lakes and Data Warehouses

Data Lakes
A data lake is a centralized repository designed to store all types of data, whether structured, semi-structured, or unstructured. By storing data in its raw form, data lakes allow for flexibility in the types of data stored and the ways in which it can be analyzed (IBM, n.d.).
- Flexibility: They are capable of storing data in a variety of formats, making them highly adaptable to changing business needs (IBM, n.d.).
- Scalability: Data lakes support scaling data storage to petabytes and beyond, which is essential for big data applications (IBM, n.d.).
- Cost-effectiveness: Generally, data lakes are more cost-effective to scale compared to traditional data warehousing solutions, especially when handling large volumes of unstructured data (IBM, n.d.).

Data Warehouse:
Data warehouses are structured repositories that store processed and filtered data specifically structured for query and analysis. This structured approach facilitates efficient data retrieval and is crucial for supporting business intelligence and decision-making processes (Gartner, n.d.).

- *Star Schema:* Characterized by a central fact table surrounded by dimension tables, this schema simplifies queries and is optimal for handling basic reporting needs (Gartner, n.d.).

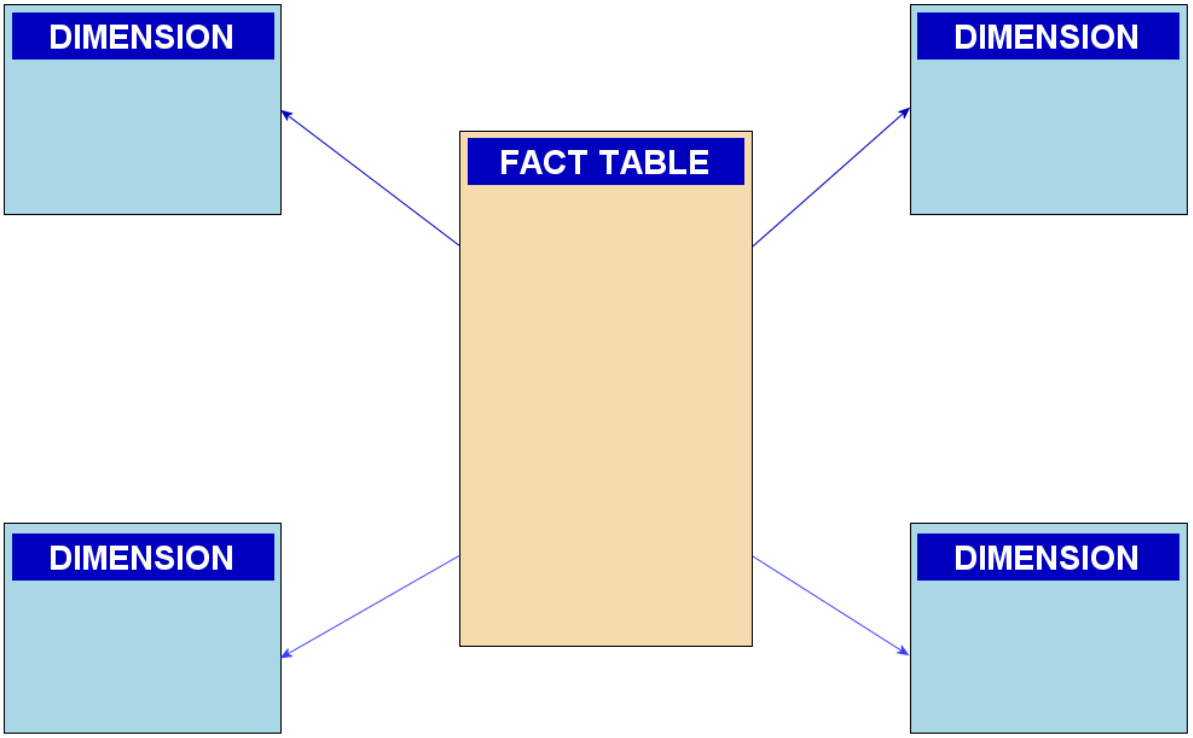


Figure 2.3 - Example of Star-Schema

- *Snowflake Schema:* Involving normalization of dimension tables into multiple related tables, this schema reduces data redundancy and can support more complex queries (Gartner, n.d.).

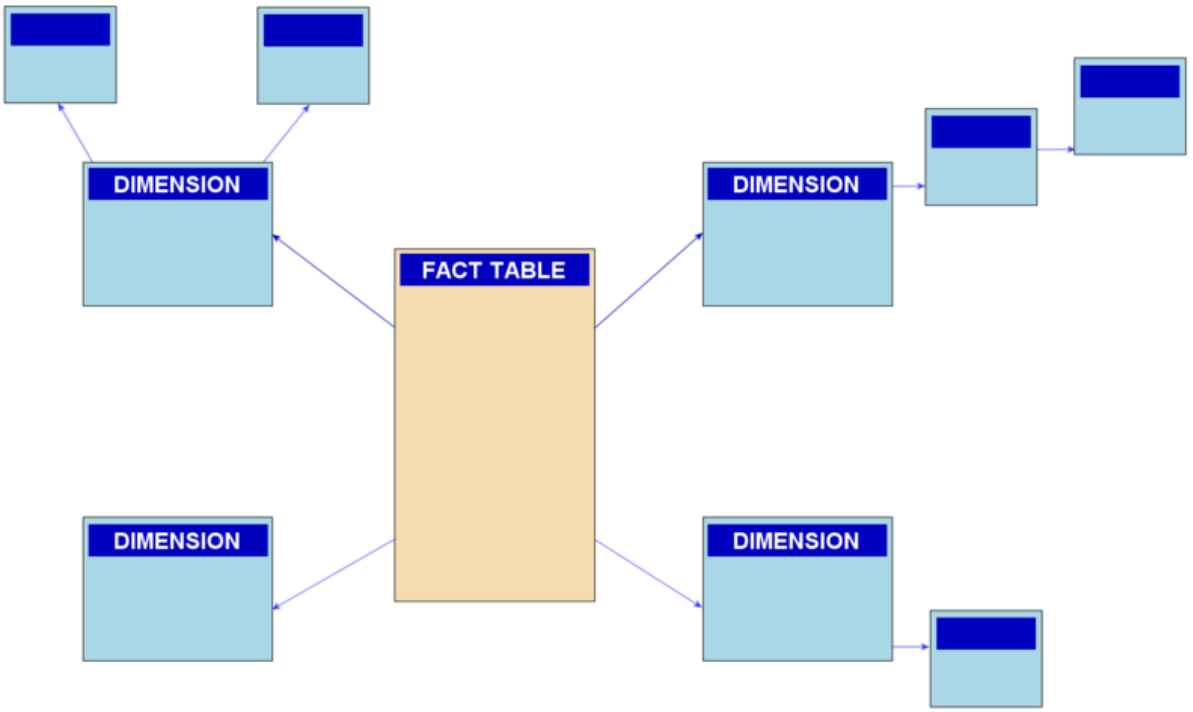


Figure 2.4 - Example of Snowflake-Schema

- *Galaxy Schema:* is a data warehousing model designed to store data thematically. Each theme is stored in a separate table, with columns optimized for specific queries. It helps enhance query performance and simplifies data management but sometimes can be more challenging to design and maintain than other data models. In our project, the Data Warehouse is designed using a Galaxy Schema with 3 Facts tables.

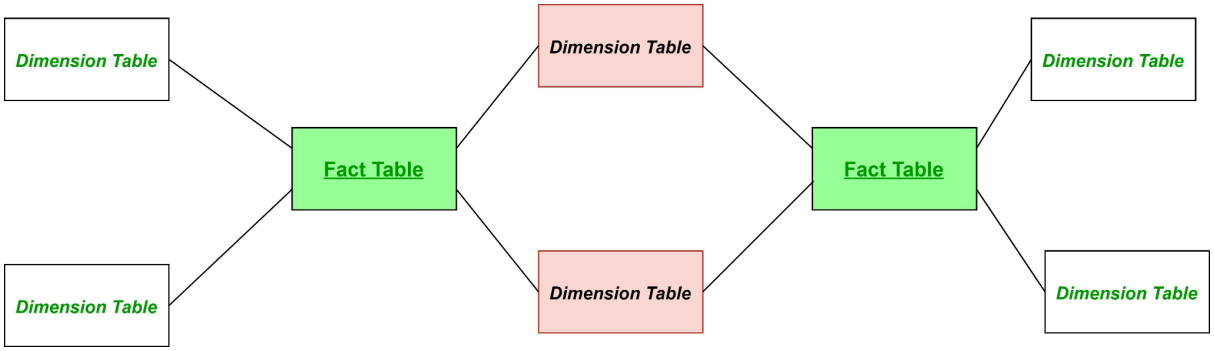


Figure 2.5 - Example of Galaxy-Schema

Role in Supporting Business Decisions: Data warehouses are integral to strategic business processes as they provide clean, organized, and easily retrievable data, essential for making timely and accurate business decisions (Gartner, n.d.).

Chapter 3. Business Requirements Analysis

3.1 Introduce the Business

Adventureworks Cycles is a prominent entity within the AdventureWorks dataset, specializing in bicycle-related products and accessories. Renowned for its professionalism and diverse offerings in the sports and recreational retail sector, Adventureworks Cycles primarily serves customers through retail outlets and online platforms, providing a wide range of bicycles, from street bikes to mountain bikes and racing bikes, along with essential accessories such as helmets, lights, and locks.

With a widespread distribution network, Adventureworks Cycles operates retail stores nationally and globally, offering doorstep delivery services through established transportation networks or partnerships with external shipping providers. Their customer base is diverse, ranging from individual consumers to organizations and businesses, seeking not only personal purchases but also gifts or essential items for sports events, advertising, or outdoor activities.

In essence, Adventureworks Cycles is more than just a bicycle retail company; it is a trusted destination for the sports and outdoor equipment community, offering a comprehensive range of products and services to meet the diverse needs of its customers.

3.2 Business requirements

Based on the context and objectives outlined in the beginning, Adventureworks Cycles needs to implement specific business requirements to *achieve operational efficiency* and *optimize the supply chain.*

To streamline purchasing operations, the company needs to reduce costs by *establishing relationships* with reliable suppliers and *enhancing financial efficiency*. Additionally, they also need to *improve order processing efficiency* by reducing the time from order creation to receipt.

In improving production performance, Adventureworks Cycles should *optimize production scheduling* to minimize downtime and increase productivity. Moreover, they should focus on *reducing the rate of defective or scrapped products* through process and quality improvements.

For enhancing sales outcomes, the company needs to improve forecasting accuracy and *optimize inventory management* to minimize costs and avoid stockouts or excess

inventory. Additionally, they should work on improving product quality to *reduce return rates* and increase customer satisfaction.

3.3 Data Description

3.3.1 Introduction to the AdventureWorks Database

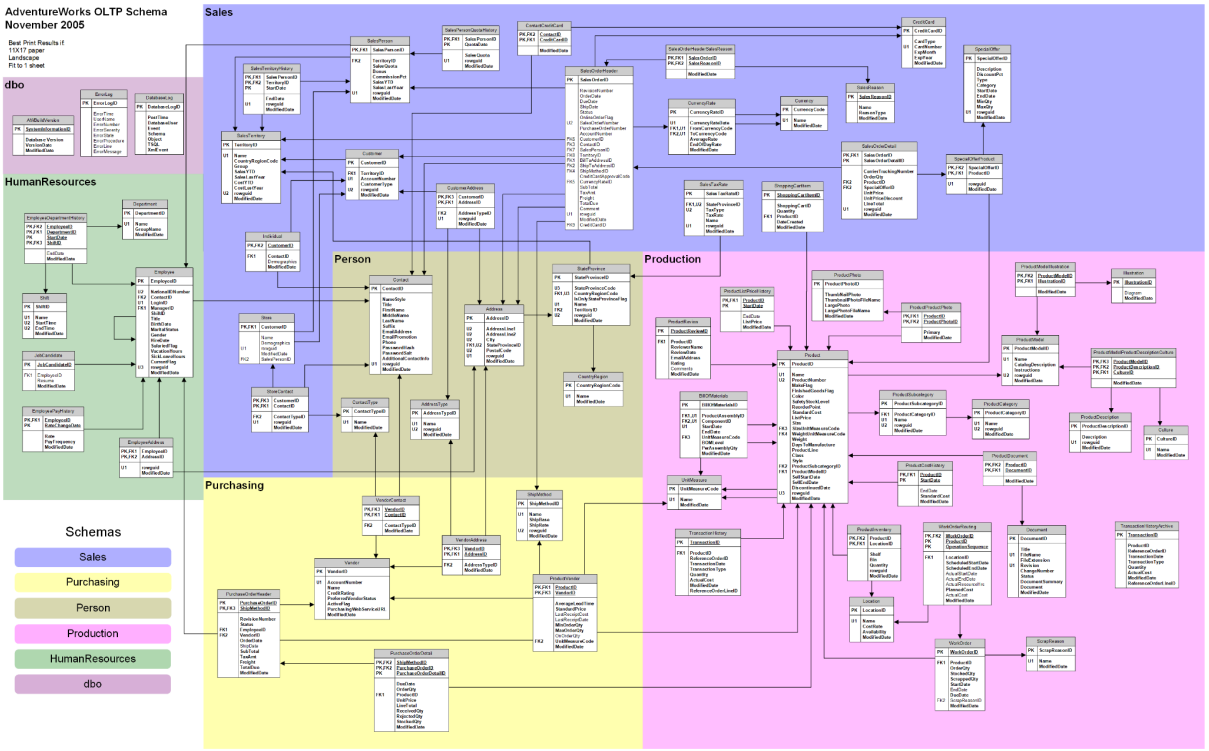


Figure 3.1 - AdventureWorks

AdventureWorks is a sample database created by Microsoft to facilitate learning and experimentation with SQL Server. It simulates the operations of a manufacturing and retail company that sells bicycles, including business processes such as production, purchasing, sales, inventory management, and more.

The AdventureWorks database is widely used in courses, tutorials, and real-world projects for purposes such as:
- *Understanding SQL Server concepts:* Table structures, data querying, data updating, etc.
- *Practicing SQL skills:* Writing complex queries, generating reports, analyzing data, etc.
- *Evaluating SQL Server tools and features:* Integration capabilities, security, performance, etc.
- *Developing SQL Server applications*: Inventory management systems, sales systems, reporting systems, etc.

Structure of the AdventureWorks Database: The database includes several key modules:

- *Production:* Manages production details including products, production processes, raw materials, etc.
- *Purchasing:* Manages purchasing including suppliers, purchase orders, invoices, etc.
- *Sales:* Manages sales including customers, orders, sold products, etc.
- *Inventory:* Manages warehouse inventory including stock levels, minimum and maximum levels, etc.
- *Human Resources:* Manages personnel including employees, departments, payroll, etc.
- *Sales and Marketing***:** Manages sales and marketing including advertising campaigns, leads, etc.
- *Finance:* Manages financial aspects including bank accounts, invoices, financial reports, etc.

These modules are interconnected through foreign keys to form a comprehensive data model for the company's business operations.

3.3.2 Detailed Description of Selected Modules

Production Module

The Production module is essential for managing the details of products throughout their lifecycle, from design to market release. It includes the following tables:

- Product: Central to inventory management, storing critical product details like ID, name, and number.
- ProductModel: Manages different models within product lines, capturing unique design specifications.
- ProductDescription: Provides rich, detailed descriptions for marketing and customer understanding.
- ProductModelProductDescriptionCulture: Bridges product models with their descriptions in various languages, vital for global reach.
- ProductCategory and ProductSubcategory: Facilitate organized categorization, improving search ability and reporting.
- ProductPhoto and ProductProductPhoto: Manage and link product images to specific products, essential for sales and marketing efforts.
- ProductDocument: Stores documents related to products, including description files, user manuals, and other technical documents. This table serves as a link between documents and their corresponding products, providing detailed information and guidance for both customers and sales staff.

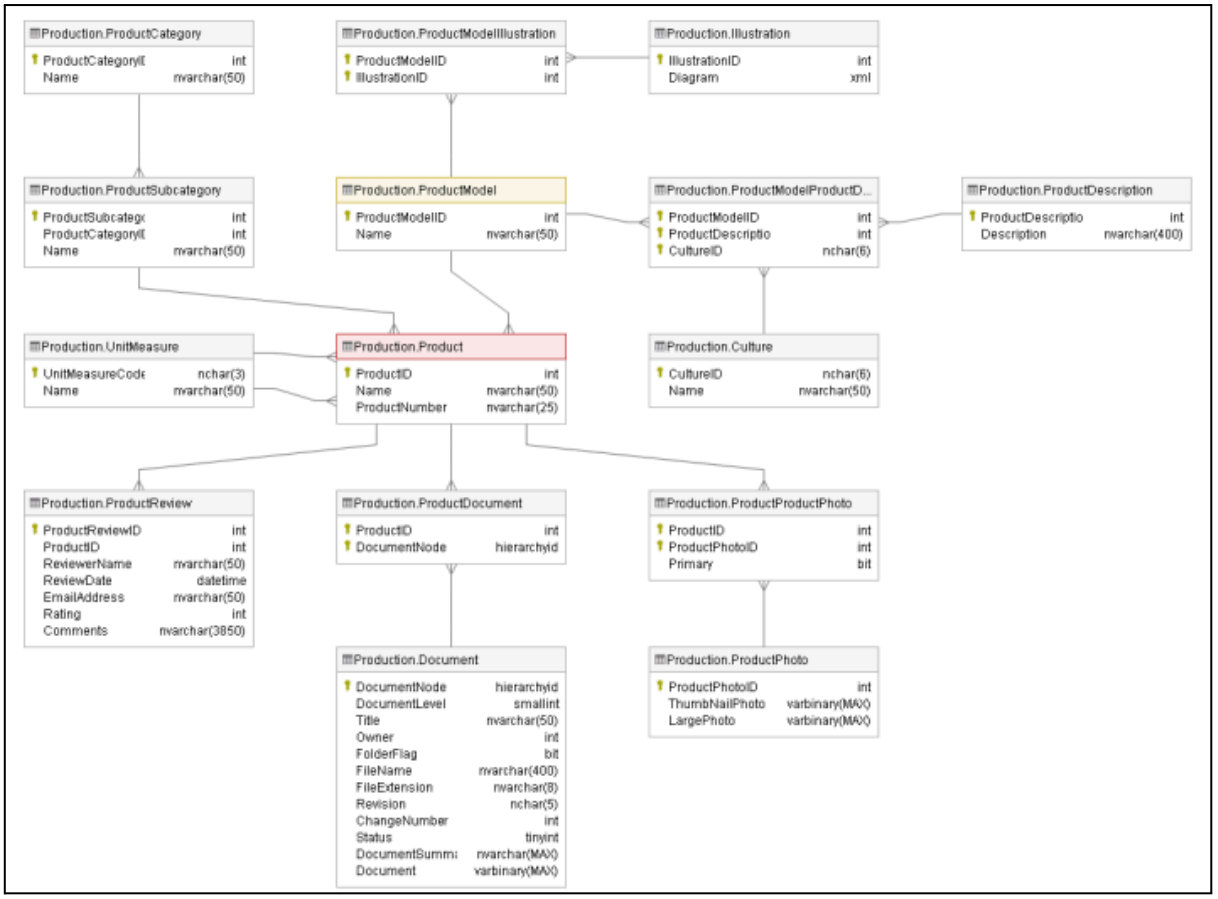


Figure 3.2 - Production Module in AdventureWorks

Purchasing Module

The Purchasing module streamlines the procurement process, from vendor interactions to order fulfillment. Key tables include:

- Vendor: Records supplier details, fundamental for procurement operations.
- PurchaseOrderHeader and PurchaseOrderDetail: Track purchase orders at both summary and detail levels, essential for inventory management and financial accounting.
- ShipMethod: Details shipping methods, integrating logistics into the procurement process.
- ProductVendor: Links products to vendors, providing a basis for sourcing and pricing strategies.

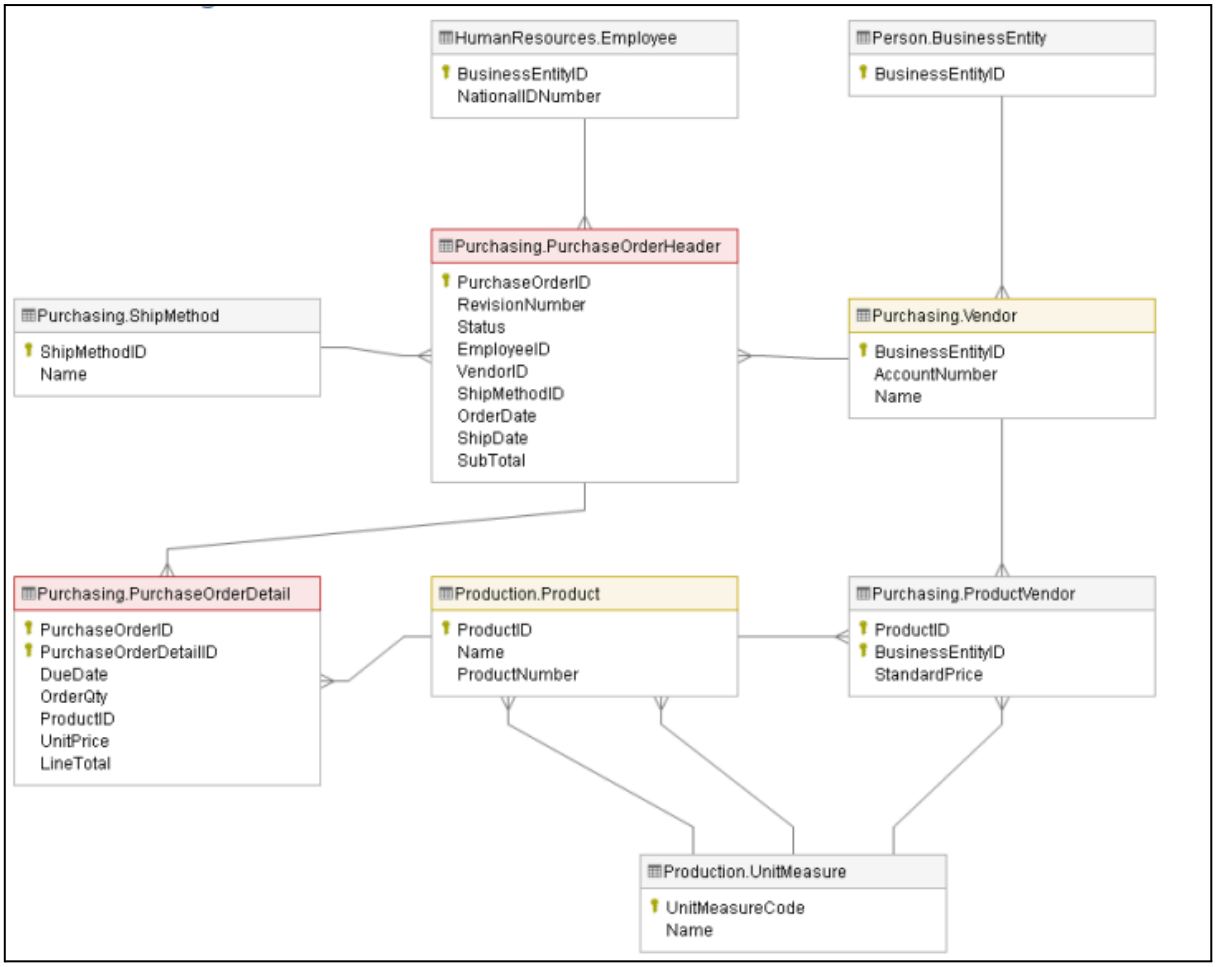


Figure 3.3 - Purchasing Module in AdventureWorks

Sales Module

The Sales module handles everything from customer management to order processing and sales analytics. It includes:

- SalesOrderHeader: Stores header information for sales orders, such as customer details, total amounts, and status.
- SalesOrderDetail: Contains line item details for sales orders, crucial for detailed sales tracking and inventory management.
- Customer: Manages comprehensive customer information, supporting CRM and sales strategies.
- SalesPerson and Territory: Track sales employee details and manage sales territories, respectively, essential for performance management and regional sales strategies.
- SpecialOffer and SalesReason: Manage promotional offers and the reasons for sales discounts or promotions, supporting strategic sales initiatives.

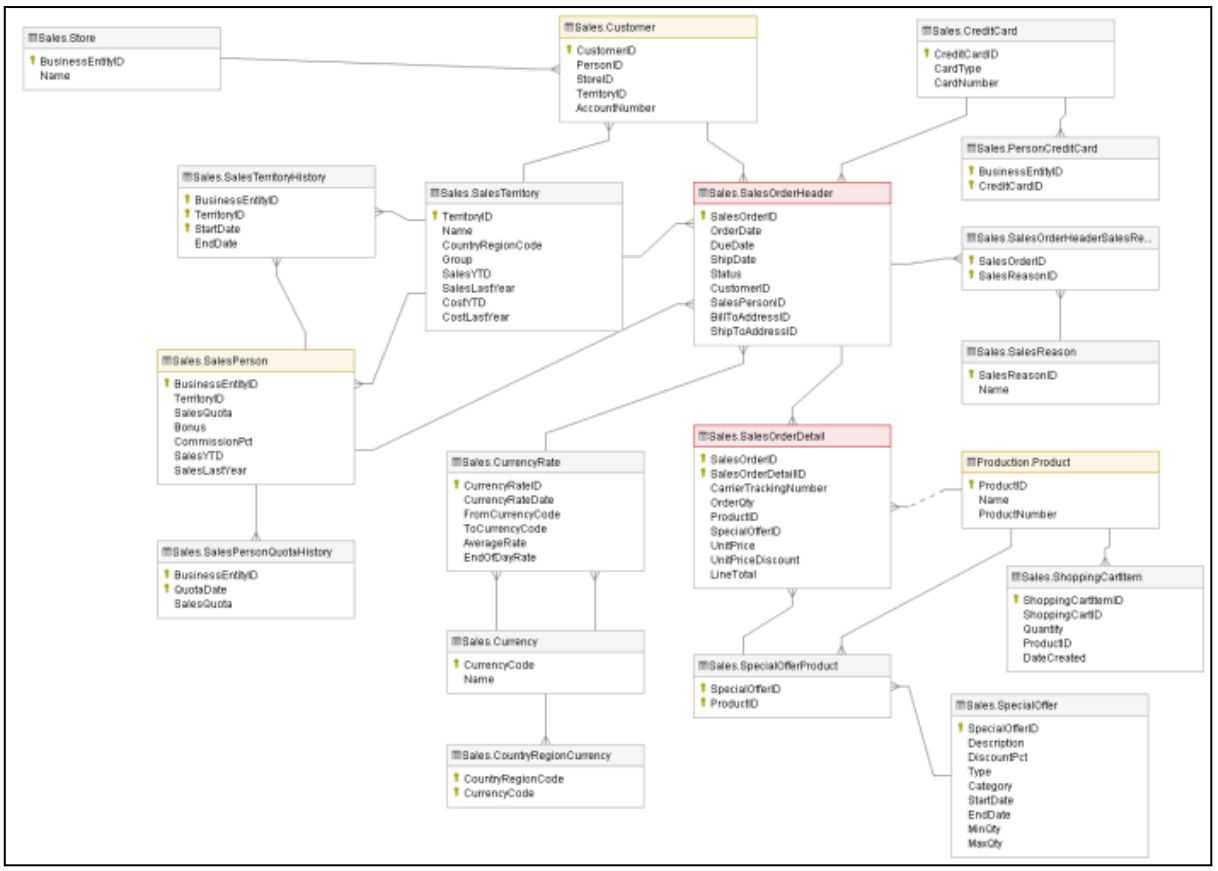


Figure 3.4 - Sales Module in AdventureWorks

3.4 Preparing Data

To serve the given purposes and the future ETL process, we need to determine the necessary data. This will avoid the excess of irrelevant data. Since our goal is to manage performance and the supply chain, the data will primarily be selected from the Production and Purchasing modules, along with two tables from Sales. Below is a detailed description of the data fields that will be selected.

Table 3.1 - Selected Data In Module Sales

Column	Table	Type	Description
SalesOrderID	SalesOrderDetail	int	Primary key. Foreign key to SalesOrderHeader.SalesOrderID.
SalesOrderDetailID	SalesOrderDetail	int	Primary key. One incremental unique number per product sold. Identity / Auto increment column.
ProductID	SalesOrderDetail	int	Product sold to customer. Foreign key to Product.ProductID
OrderQty	SalesOrderDetail	smallint	Quantity ordered per product.
UnitPrice	SalesOrderDetail	money	Selling price of a single product
LineTotal	SalesOrderDetail	money	Per product subtotal. Computed as UnitPrice * (1 - UnitPriceDiscount) * OrderQty.
SalesOrderID	SalesOrderHeader	int	Primary key
OrderDate	SalesOrderHeader	datetime	Dates the sales order was created.
DueDate	SalesOrderHeader	datetime	Date the order is due to the customer.
ShipDate	SalesOrderHeader	datetime	Date the order was shipped to the customer.

Table 3.2 - Selected Data In Module Purchasing

Column	Table	Type	Description
PurchaseOrderID	PurchaseOrderDetail	int	Primary key. Foreign key to PurchaseOrderHeader.PurchaseOrderID
PurchaseOrderDetailID	PurchaseOrderDetail	int	Primary key. One line number per purchased product. Identity / Auto increment column
OrderQty	PurchaseOrderDetail	smallint	Quantity ordered.
ProductID	PurchaseOrderDetail	int	Product identification number. Foreign key to Product.ProductID.
UnitPrice	PurchaseOrderDetail	money	Vendor's selling price of a single product.
LineTotal	PurchaseOrderDetail	money	Per product subtotal. Computed as OrderQty * UnitPrice.
ReceivedQty	PurchaseOrderDetail	decimal(8, 2)	Quantity actually received from the vendor
RejectedQty	PurchaseOrderDetail	decimal(8, 2)	Quantity rejected during inspection
StockedQty	PurchaseOrderDetail	decimal(9, 2)	Quantity accepted into inventory. Computed as ReceivedQty - RejectedQty.
PurchaseOrderID	PurchaseOrderHeader	int	Primary key. Identity / Auto increment column
VendorID	PurchaseOrderHeader	int	Vendor with whom the purchase order is placed. Foreign key to Vendor.BusinessEntityID.
ShipMethodID	PurchaseOrderHeader	int	Shipping method. Foreign key to ShipMethod.ShipMethodID.
OrderDate	PurchaseOrderHeader	Datetime	Purchase order creation date.

ShipDate	PurchaseOrderHeader	Datetime	Estimated shipment date from the vendor.
SubTotal	PurchaseOrderHeader	money	Purchase order subtotal.
TaxAmt	PurchaseOrderHeader	money	Tax amount
TotalDue	PurchaseOrderHeader	money	Total due to vendor. Computed as Subtotal + TaxAmt + Freigh
ShipMethodID	ShipMethod	int	Primary key for ShipMethod records. Identity / Auto increment column
Name	ShipMethod	nvarchar (50)	Shipping company name
ShipBase	ShipMethod	money	Minimum shipping charge
ShipRate	ShipMethod	money	Shipping charge per pound
BusinessEntityID	Vendor	int	Primary key for Vendor records. Foreign key to BusinessEntity.BusinessEntityID
Name	Vendor	nvarchar (50)	Company name
CreditRating	Vendor	tinyint	1 = Superior, 2 = Excellent, 3 = Above average, 4 = Average, 5 = Below average

Table 3.3 - Selected Data In Module Production

Column	Table	Type	Description
LocationID	Location	smallint	Primary key for Location records. Identity / Auto increment column
Name	Location	nvarchar(50)	Location description
CostRate	Location	smallmoney	Standard hourly cost of the manufacturing location.
ProductID	Product	int	Primary key for Product records. Identity / Auto increment column
Name	Product	nvarchar(50)	Name of the product.
SafetyStockLevel	Product	smallint	Minimum inventory quantity
StandardCost	Product	money	Standard cost of the product.
ListPrice	Product	money	Selling price
ProductLine	Product	nchar(2)	R = Road, M = Mountain, T = Touring, S = Standard
ProductSubcategoryID	Product	int	Product is a member of this product subcategory. Foreign key to ProductSubCategory.ProductSubCategoryID.
ProductCategoryID	Productcategory	int	Primary key for ProductCategory records. Identity / Auto increment column
Name	Productcategory	nvarchar(50)	Category description.
ProductSubcategoryID	ProductSubcategory	int	Primary key for ProductSubcategory records. Identity / Auto increment

			column
ProductCategoryID	ProductSubcategory	int	Product category identification number. Foreign key to ProductCategory.ProductCategoryID.
Name	ProductSubcategory	nvarchar(50)	Subcategory description.
ScrapReasonID	ScrapReason	smallint	Primary key for ScrapReason records. Identity / Auto increment column
Name	ScrapReason	nvarchar(50)	Failure description
WorkOrderID	WorkOrder	int	Primary key for WorkOrder records. Identity / Auto increment column
OrderQty	WorkOrder	int	Product quantity to build.
StockedQty	WorkOrder	int	Quantity built and put in inventory.
ScrappedQty	WorkOrder	smallint	Quantity that failed inspection.
ScrapReasonID	WorkOrder	smallint	Reason for inspection failure.
StartDate	WorkOrder	datetime	Work order start date.
EndDate	WorkOrder	datetime	Work order end date.
DueDate	WorkOrder	datetime	Work order due date.
WorkOrderID	WorkOrderRouting	int	Primary key. Foreign key to WorkOrder.WorkOrderI
ProductID	WorkOrderRouting	int	Primary key. Foreign key to Product.ProductID
ActualResourceHrs	WorkOrderRouting	decimal (9, 4)	Number of manufacturing hours used.
ActualCost	WorkOrderRouting	money	Actual manufacturing cost.
OperationSequence	WorkOrderRouting	smallint	Primary key. Indicates the

			manufacturing process sequence.
LocationID	WorkOrderRouting	smallint	Manufacturing location where the part is processed. Foreign key to Location.LocationID.

Chapter 4. Integrating Data and Building Data Warehouse

4.1 Design Data Warehouse Model

4.1.1 Bus Matrix

Bussiness Process	Vendor	Product	Scrap Reason	ShipMethod	Location	Date
Production Cycle Mangement		X	X	X	X	X
Purchase Processing Management	X	X		X		X
Product Analyst	X	X				X
Cost Units Management		X			X	X

Figure 4.1 - Bus Matrix

The business matrix you provided is a strategic tool that outlines the interaction between various business processes and the corresponding data dimensions within a data warehouse. This matrix serves as a clear guide, highlighting which data dimensions are utilized by specific business processes such as Production Cycle Management, Purchase Processing Management, Product Analysis, and Cost Units Management.

Each row in the matrix represents a different business process, while the columns represent data dimensions such as Vendor, Product, Scrap Reason, ShipMethod, Location, and Date. The presence of an "X" in a cell indicates that the business process utilizes data from the associated dimension. For example, Production Cycle Management involves a comprehensive array of dimensions including Product, Scrap Reason, ShipMethod, Location, and Date, suggesting a need to track the entire lifecycle of production, from manufacturing to shipping, and including details like scrap reasons and production dates.

This matrix is incredibly beneficial for several reasons. Firstly, it provides clarity and focus by delineating the data needs aligned with specific business requirements, which aids in system design and ensures that data structures are tailored to support these needs efficiently. Secondly, it supports strategic business decisions by offering a structured framework that outlines data interactions and dependencies, enhancing capabilities in

forecasting, budgeting, and operational planning. Moreover, it facilitates effective communication across different departments, ensuring all stakeholders have a clear understanding of data flows and how they impact business operations.

4.1.2 Galaxy Schema of Supply Chain

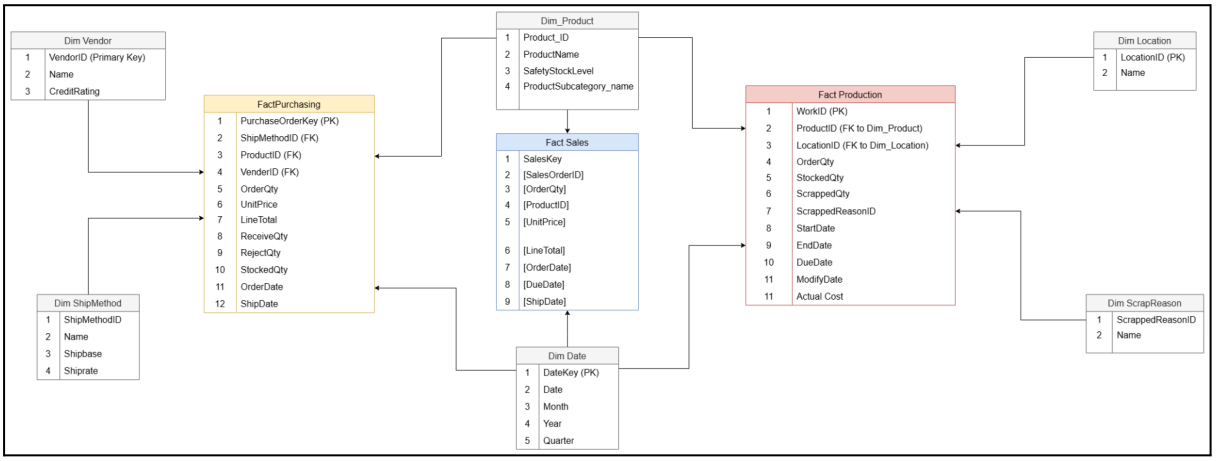


Figure 4.2 - Galaxy Schema of Supply Chain Management

This is our galaxy schema, which is a variant of the star schema used in data warehouse design. This type of schema is ideal for multi-faceted data analysis and querying. Below is a detailed description of the schema and its benefits to businesses, as well as reasons for choosing the galaxy schema for this module.

The schema includes several fact and dimension tables:

Fact Tables:

+ FactPurchasing: Contains data related to purchasing activities, such as order quantity, unit price, total line, quantities received, rejected, and stocked, as well as order and ship dates.
+ Fact Sales: Stores information about sales transactions including quantities sold, prices, total line amounts, and relevant dates like order, due, and ship dates.
+ Fact Production: Tracks production details, including quantities ordered, stored, scrapped, reasons for scrapping, and costs associated with production along with relevant production timelines.

Dimension Tables:

+ Dim Product: Details about products including names and subcategories.
+ Dim Vendor: Information regarding suppliers including names and credit ratings.
+ Dim ShipMethod: Shipping methods employed, detailed down to base and rate.
+ Dim Location: Locations relevant to storage or production.
+ Dim Date: A date dimension for temporal analysis.
+ Dim ScrapReason: Reasons for scrapping materials.

The galaxy schema provides substantial business benefits due to its comprehensive and efficient approach to data analysis. By connecting multiple fact tables with several dimension tables, it facilitates detailed multi-dimensional analysis. This depth of analysis supports more precise and informed decision-making across various departments. Additionally, its high performance allows for efficient and quick querying, significantly reducing data access times. This is crucial in today's fast-paced business environments where timely information can lead to competitive advantage.

Moreover, the galaxy schema offers remarkable scalability. Businesses can expand their data warehouse by adding new fact or dimension tables without major modifications to the existing structure, ensuring that the data architecture can evolve with the company's growth and changing needs. This flexibility is paired with the schema's ability to maintain data accuracy and reliability, providing a holistic and detailed view of operations. This comprehensive visibility into purchasing, production, and sales activities not only ensures data integrity but also enhances reliability in strategic planning and operational management.

Thus, utilizing a galaxy schema not only streamlines data management but also significantly contributes to strategic business operations, offering a clear competitive edge in analyzing market trends and improving operational efficiencies. This makes it an invaluable tool for businesses aiming to maintain agility and accuracy in their decision-making process

4.1.3 Reason for Choosing Galaxy Schema

The galaxy schema was chosen for this module due to its effective cross-table analytical capabilities. In this scenario, fact tables such as FactPurchasing, Fact Sales, and Fact Production are closely linked to dimension tables like Dim Product and Dim Location, allowing for detailed aggregate analysis across purchasing, sales, and production activities within a unified data structure.

Utilizing a galaxy schema enables businesses to access and analyze data swiftly and accurately, effectively supporting decision-making processes and business strategy development. The design caters to complex business environments where multifaceted relationships and interactions between different business activities need to be analyzed and understood comprehensively. This schema thus facilitates deeper insights into operational efficiencies and market trends, proving crucial for strategic planning and competitive analysis in dynamic business landscapes.

4.2 Data Flow

4.2.1 Overall process

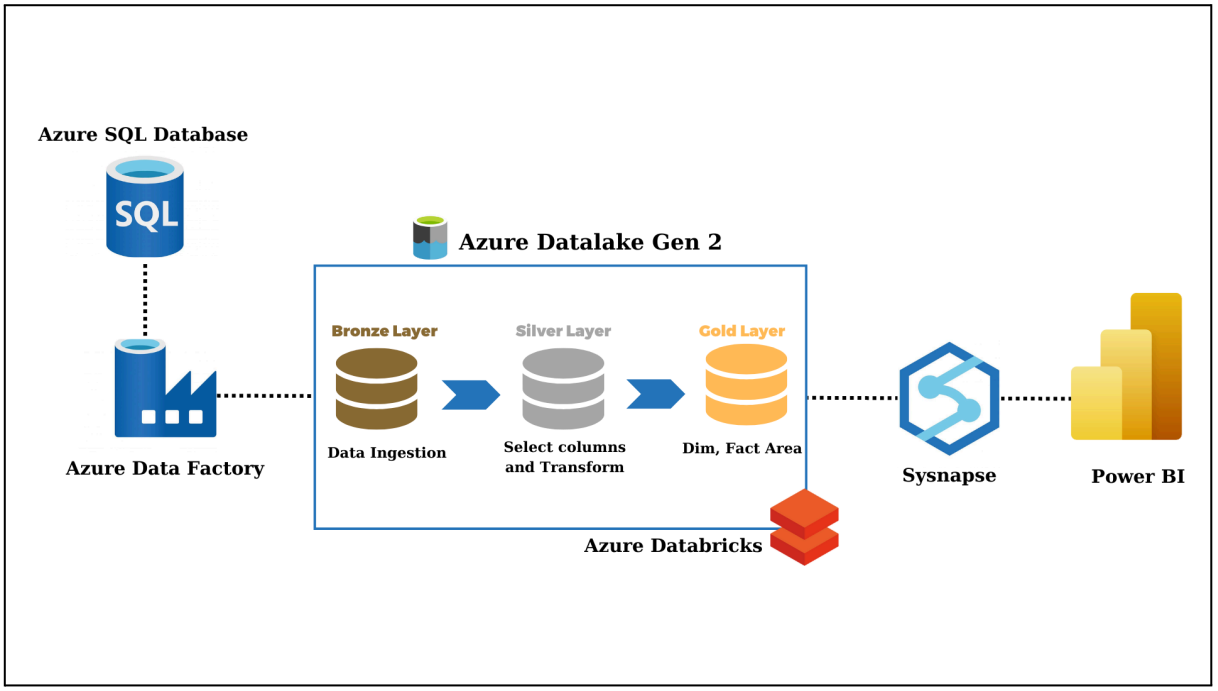


Figure 4.3 - Data Flow

Figure 4.3 illustrates a complete process that the data undergoes. The data passes through three major stages. First, the raw data is extracted from the Azure SQL Database to the Bronze layer of the Data Lake via ADF. From there, we select and remove unnecessary columns that do not align with our objectives. The retained data is then transformed in terms of data types and renamed, facilitating smoother transitions from the Silver layer to the Gold layer.

During the transition from Silver to Gold, the data is meticulously selected and carefully calculated to produce comprehensive Dim and Fact tables. However, these are merely individual tables without any established relationships at this stage, which will be discussed later. All data loading activities from Bronze to Silver, Silver to Gold, and these data transformations are performed on Azure Databricks. We utilize Databricks due to the assumed Big Data context, where Databricks is the most suitable service as it excels in handling large volumes of data.

Once the data is thoroughly cleansed and ready for analysis, we create views of the data from the Gold layer using SQL Serverless Pool in the Synapse space. Through Synapse, the data is visualized using Power BI to clarify the predefined metrics.

4.2.2 Ingest data from Database to Data Lake

Data in a business comes from various sources such as sales systems, customer management, and production departments. For a manager, it is crucial and essential to have a comprehensive grasp of information to make informed decisions. In this case, the best solution is one that can aggregate data from all sources into one place, allowing for analysis that reveals both the overall picture and specific details. In our project, Azure was chosen to achieve this. Azure Data Factory (ADF) is designed to aggregate data from numerous sources. These considerations are hypothetical as we are using the AdventureWorks dataset for this research. This dataset will be uploaded to Azure SQL Database to facilitate the subsequent data ingestion and loading process.

We proceed with the first step of the ETL process, which is loading the necessary data tables into the Data Lake. We use Data Lake because our data context is Big Data, making it essential to store them in the Data Lake. To ingest a large volume of data from various sources into the Data Lake, Azure Data Factory is required. Once the data has been uploaded to Azure SQL Database, necessary tables such as Product, WorkOrder, PurchaseOrderHeader, etc., will be extracted. The following are the operations performed with ADF.

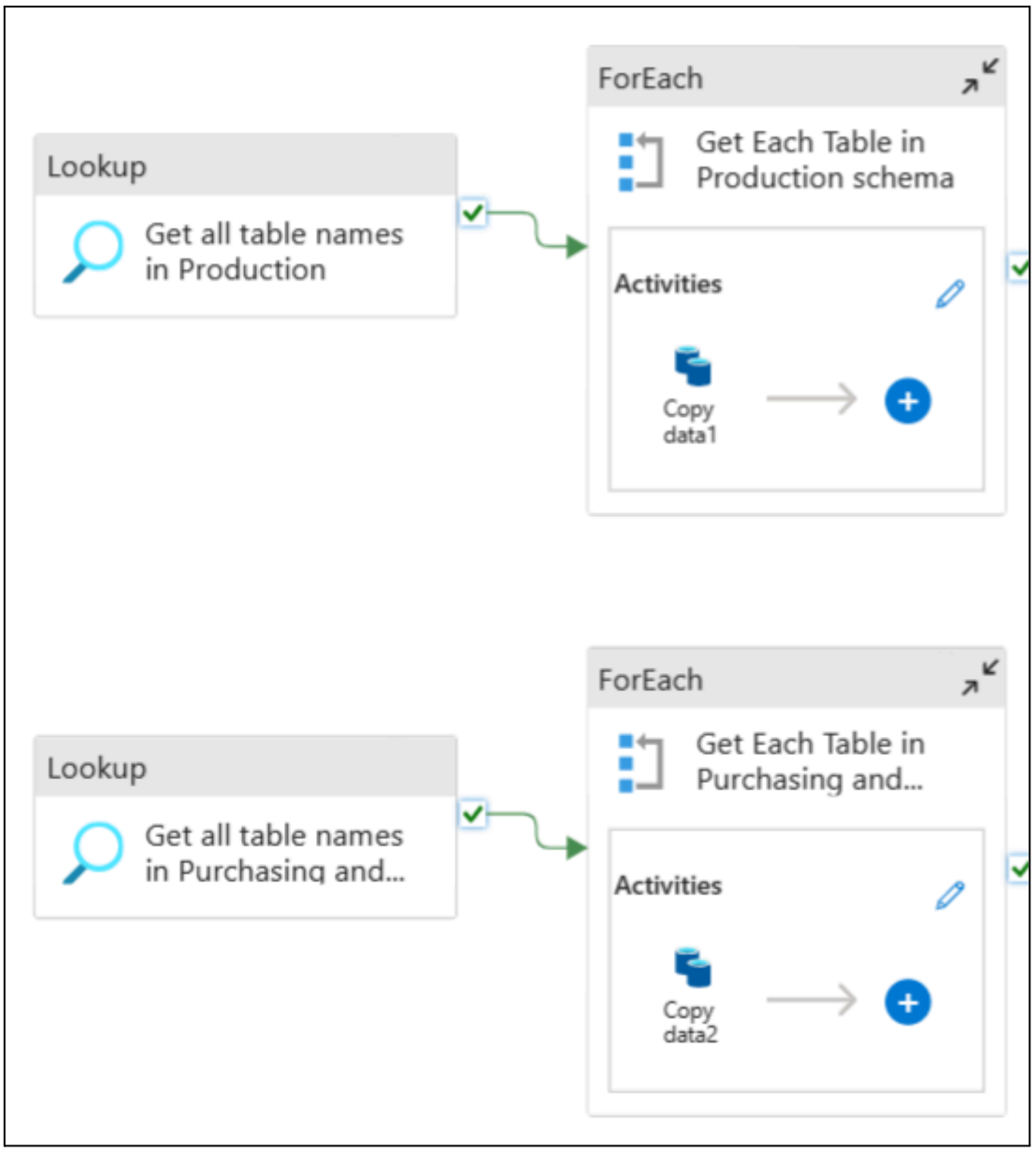


Figure 4.4 - Ingest data to Bronze

First, the active Lookup will be used to retrieve all the necessary table names using an SQL query. The input data is the AdventureWorks dataset stored in Azure SQL Database. Next, an SQL query will be used to extract all the required data tables for the topic "Operational Performance and Supply Chain Management."

Get all table names in Production module

```
SELECT table_schema, table_name
FROM information_schema.tables
WHERE table_type = 'base table'
  AND (
```

```
        (table_schema  =  'Production'  AND  table_name  IN
('Product',         'ScrapReason',         'ProductCategory',
'ProductSubcategory',         'Location',         'WorkOrder',
'WorkOrderRouting'))
)
```

Get all table names from Purchase and Sales module

```
SELECT table_schema, table_name
FROM information_schema.tables
WHERE table_type = 'base table'
  AND (
        (table_schema  =  'Purchasing'  AND  table_name  IN
('ShipMethod',        'Vendor',        'PurchaseOrderDetail',
'PurchaseOrderHeader'))
  OR
           (table_schema   =   'Sales'   AND   table_name   IN
('SalesOrderDetail', 'SalesOrderHeader'))
)
```

Once all the table names are retrieved, each of them will be iterated through using an active ForEach. Then, the active Copy Data will reference the AdventureWorks dataset to extract the tables we want to obtain.

4.2.3 Transformation with Azure Databricks

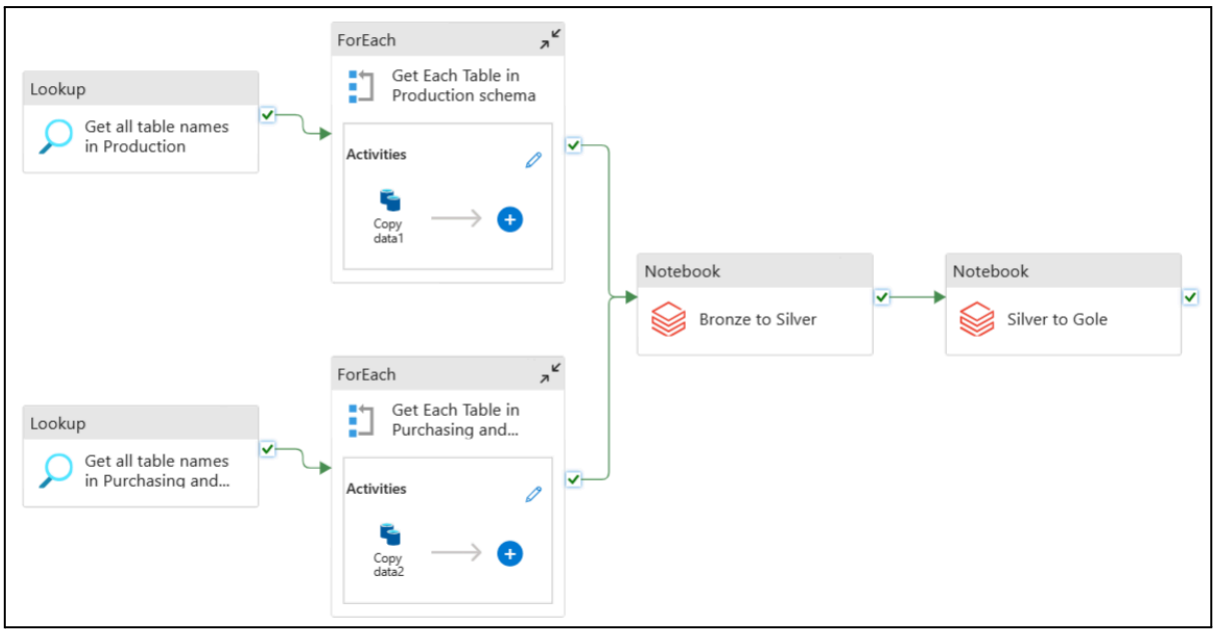


Figure 4.5 - Data Pipeline

Once the tables containing the necessary data are in the Bronze layer, we will use PySpark in Azure Databricks to process the data through the Silver and Gold layers sequentially. In the transition from Bronze to Silver, we will select the necessary columns from the data tables and perform some data transformations, such as converting columns to datetime format. This is an important step in cleaning and standardizing the data before further in-depth analysis. Once the input data is deemed sufficient for our objectives at the Silver layer, it will be sliced, merged, and joined to create Dimension (Dim) and Fact tables. These tables are crucial for analysis and reporting. However, at this step, the Dim and Fact tables will not yet have any relationships such as primary keys or foreign keys. Establishing these relationships will be done in the final step when the data is imported into Power BI. After completing the above process, the cleaned and standardized data is stored in the Gold layer with the Fact and Dim table names as designed in section 4.1.

4.2.4 Create View in Synapse WorkSpace

In this phase, we use SQL Serverless Pool in Azure Synapse Analytics to create views from the data processed and stored in the Gold layer. Creating these views facilitates easy querying and data analysis without needing to move or copy the data to another system. The views also provide an abstraction layer, allowing end users to access the data without needing to understand the complex structure of the database.

First, we connect to Azure Synapse Analytics using SQL Serverless Pool to connect to Azure Data Lake Gen 2, where the Gold layer data is stored. Then, we write SQL queries to create views based on the data tables in the Gold layer. These views aggregate and

transform the data, making it ready for analysis and reporting. For example, we create a view for the Fact Sales table as follows:

```
CREATE VIEW dim_Product
AS
SELECT *
FROM OPENROWSET(
                                                                    BULK
'https://adlsbamosfinal.dfs.core.windows.net/datalake/Gold/Di
m_Product/*.parquet',
    FORMAT = 'PARQUET'
) AS [result];
```

The benefits of using SQL Serverless Pool include allowing direct data querying from the Data Lake without needing to move the data, thereby increasing the speed of data access and analysis. The pay-per-query model optimizes costs compared to solutions requiring data storage. Easy scalability and the ability to manage large datasets without performance concerns are also strong points. Additionally, the created views can easily integrate with Power BI to generate interactive reports and dashboards.

This process is illustrated in the diagram, showing how data from the Gold layer is processed and stored in Azure Data Lake Gen 2, then using Azure Synapse Analytics with SQL Serverless Pool to create views. These views enable end users to query and analyze data easily and efficiently, with Power BI being used to visualize the data and generate detailed reports. This process ensures that data is stored, processed, and analyzed efficiently, providing businesses with crucial information for strategic decision-making.

Chapter 5. Visualization By Power BI

5.1 Create Relationship in Power BI

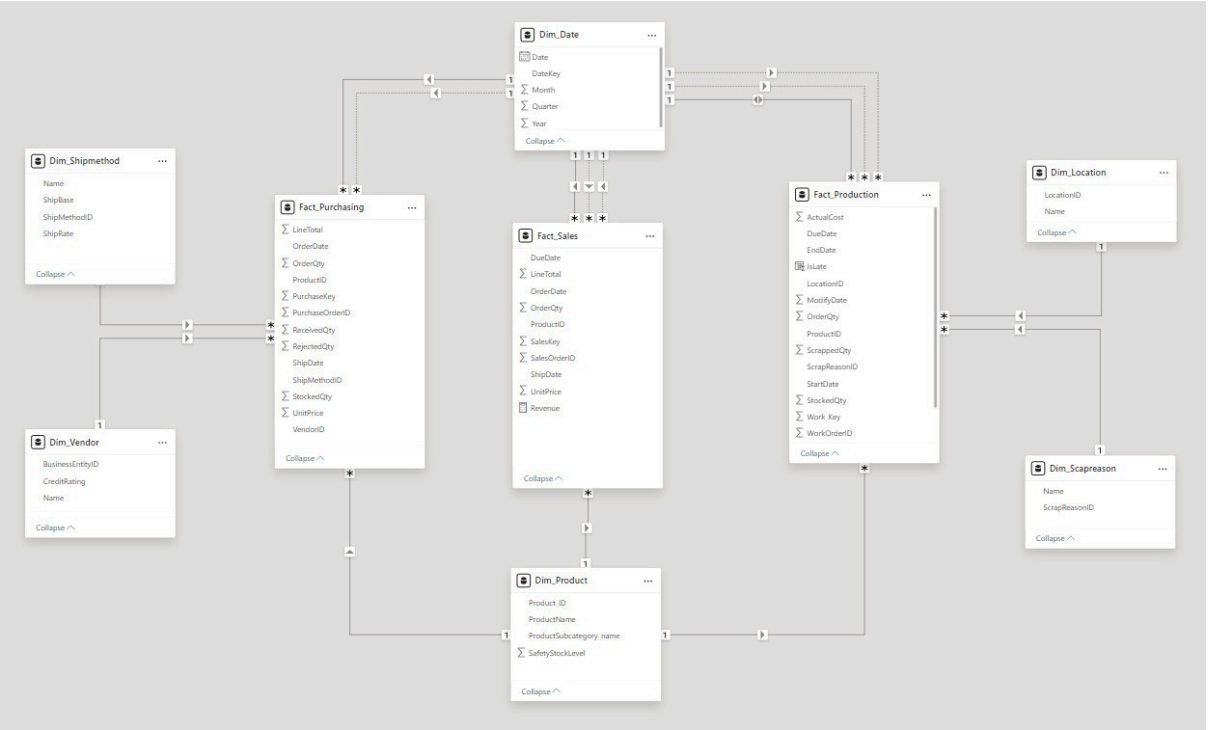


Figure 5.1 - Galaxy Schema of Data Warehouse

To create a Power BI report with data located on Azure Data Lake (Data Storage), we cannot directly connect Power BI to the Data Lake. Instead, we use Azure Synapse Analytics as the connection hub. Within Synapse, there are two types of SQL pools: SQL serverless pool and SQL dedicated pool. With the SQL dedicated pool, we have to move data from the Data Lake into the dedicated pool, which is labor-intensive. Therefore, we use the SQL serverless pool to create views directly from the data in the Data Lake, allowing Power BI to efficiently query data through Synapse.

After the data is retrieved, we create relationships between the dimension (dim) and fact tables in Power BI. This step further clarifies the BI solution we are building, based on the Data Warehouse architecture.

This is our approach to building a BI solution with the support of Azure services.

5.2 Overview Dashboard

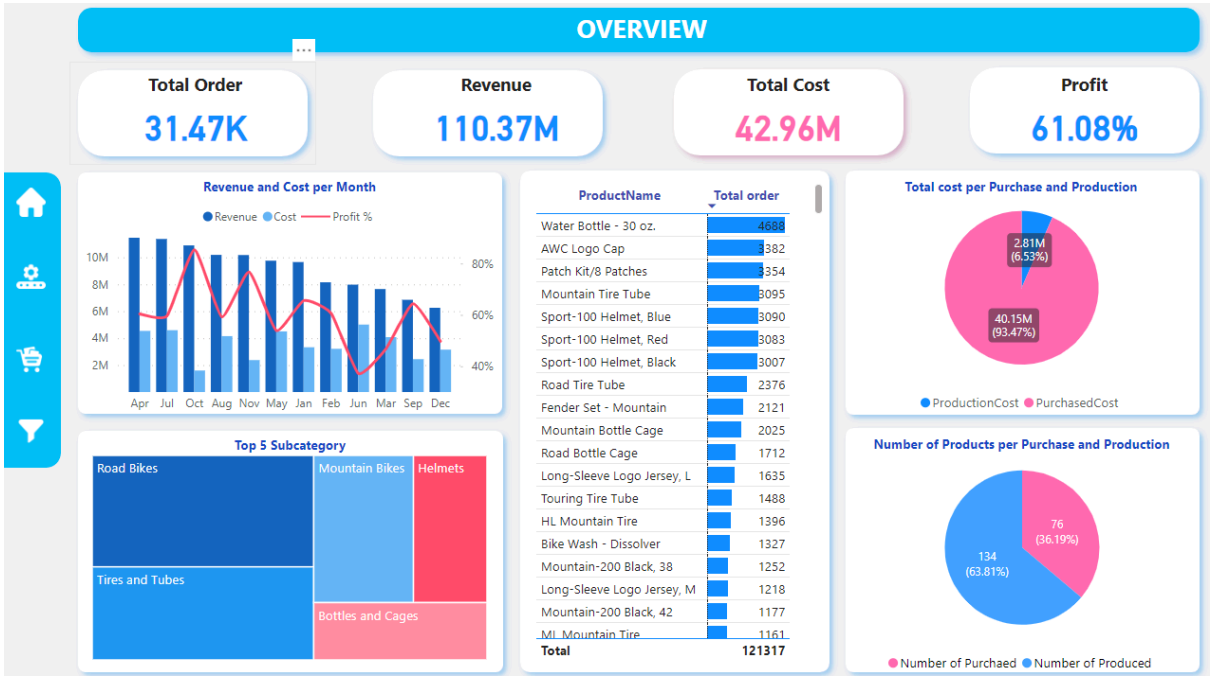


Figure 5.2 - Overview Dashboard

AdventureWorks is a company that manufactures and retails outdoor sports products, including bicycles, clothing, and accessories. The company operates a complex database with multiple data tables, reflecting various aspects of business operations such as manufacturing, sales, finance, human resources, and inventory. We have developed a Data Warehouse for managing operational performance and the supply chain, and this Overall Dashboard is designed based on the Data Warehouse to provide managers with a comprehensive overview of manufacturing and purchasing activities related to sales.

Manufacturing and purchasing play a crucial role in ensuring a steady supply of products to the market. Efficient manufacturing operations help us maintain product quality and meet customer demands promptly. If the production process is disrupted or inefficient, it can lead to product shortages, negatively affecting sales. Conversely, stable and efficient production enhances brand reputation, increases customer satisfaction, and thereby boosts sales.

Good management of supply sources and raw material costs is essential for optimizing production costs. If we can negotiate and purchase raw materials at favorable prices with guaranteed quality, it will help reduce production costs, thereby increasing profits. Moreover, maintaining good relationships with suppliers ensures a continuous supply of raw materials, supporting smooth production processes.

With this Overall Dashboard, it provides a comprehensive view of the company's operational effectiveness, thereby helping the leadership make more accurate strategic decisions. Having access to key metrics such as revenue, costs, and order quantities allows managers to easily monitor and assess the effectiveness of various business operations, helping to identify potential issues and improvement opportunities, thereby promoting sustainable development of the company.

Let's look at the main components of the Dashboard:
Starting with the Revenue and Total Cost at the top, we see that AdventureWorks generated impressive revenue of $110.37 million while the total cost was $42.96 million. This provides us with a strong profit margin of 61.08%, proving the effectiveness of our operations and cost management strategy. Additionally, the Total Number of Orders shows that we processed 31.47 thousand orders, reflecting high demand for our products.

Moving to more detailed charts of Revenue and Costs over time allows us to track financial performance throughout the year. The blue columns represent monthly revenue, the red line represents costs, and the profit margin is also highlighted. This chart helps us identify seasonal trends and fluctuations in sales and costs.

The Product Name and Total Orders bar chart lists the products with the highest order quantities. Clearly, the Water Bottle - 30 oz. leads with 4,688 orders, followed by other popular items like the Logo AWC Cap and Patch Kit. This information is crucial for inventory planning and marketing strategy.

We also have two pie charts illustrating Total Costs for Purchase and Production and Product Quantities by Purchase and Production. These charts provide detailed analysis of costs and product quantities, helping to understand better resource allocation and production efficiency.

Finally, the Top 5 Product Categories chart displays the leading product categories, with Road Bikes and Mountain Bikes being significant contributors to our revenue.

As we can see from the overview, manufacturing plays a very important role in maintaining a supply of products and meeting customer demand. Understanding the factors affecting the production process not only helps us optimize costs but also improves performance and product quality. Therefore, we will explore deeper into production metrics, including production costs, the quantity of products manufactured, and the defect rate through a detailed Production Dashboard. This information will provide a comprehensive view of production efficiency, helping us identify improvement opportunities and enhance our competitive edge in the market.

5.3 Production Dashboard

Figure 5.3 - Production Dashboard

In this section, we will delve into the analysis of the components of the Production Dashboard to provide a detailed view of AdventureWorks' manufacturing operations.

The Production Dashboard helps track and assess the performance of the manufacturing process, with key metrics including the rate of delayed orders, defect rate, average production time, and total work orders. Notably, the delayed order rate is 54%, indicating that over half of the orders are not completed on time. This poses a significant challenge for the company in improving manufacturing efficiency and meeting customer demands promptly. The defect rate is 0.19%, which, although relatively low, still requires attention to ensure even better product quality.

On average, it takes about 15.51 days to complete the production of a product. This information helps us better understand the time required to complete a manufacturing process. Additionally, the total number of work orders reached 43,000, reflecting the workload and production capacity of the company. These metrics not only help track manufacturing performance but also provide crucial data for improving workflow.

The "Stock Average Days Production per Subcategory" chart shows the average number of days required to produce products across different subcategories. For example, producing Mountain Bikes takes an average of 15.71 days, Mountain Frames 15.62 days, and Touring Bikes 15.56 days. This information is invaluable for managing time and

planning production more effectively. By knowing the average production time for each type of product, the company can adjust production schedules to optimize resources and minimize waiting times.

Next, the "Total of Scrapped by Location" chart displays the total number of scrapped products at various production locations. For example, the Subassembly location has the highest number of scrapped products. This helps identify steps in the production process that need to be inspected and improved to minimize waste. Understanding where product faults commonly occur enables the company to focus on improving critical steps in the production process.

Finally, the "Reason of Scrapped" pie chart analyzes the main reasons for product scrapping, such as Trim length, Thermoforming issue, and Color inconsistency. These reasons provide insights into the quality issues the company is facing. Consequently, the leadership can implement specific solutions to reduce the rate of scrapped products.

For detailed reporting, we can see specifics such as the number of products in inventory, production costs, and defect rates for each subcategory. For instance, Handlebars have an inventory of 282,132, a production cost of $959,744.50, and a defect rate of 0.18%. The combination of these data helps identify areas where cost optimization and product quality improvement are needed. Through detailed analysis of each subcategory, the company can detect weaknesses and devise measures to improve manufacturing efficiency.

Going in-depth into Production as described above will help the leadership and production management at AdventureWorks gain a comprehensive view of manufacturing performance. This will enable them to make strategic decisions aimed at improving processes, reducing error rates, and optimizing production time. Monitoring these metrics and charts will help the company maintain and enhance product quality, better meet customer demands, and enhance the efficiency of manufacturing operations.

5.4 Purchasing Dashboard

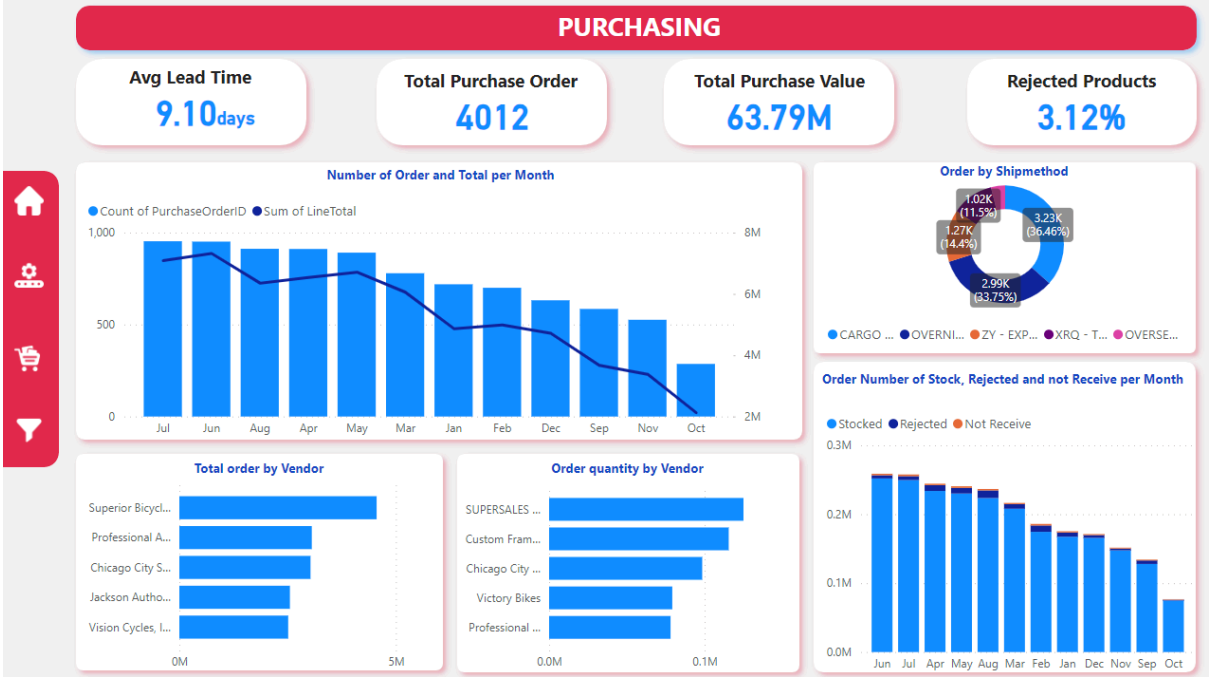


Figure 5.4 - Purchasing Dashboard

After understanding the metrics and performance in manufacturing, we now turn to the detailed aspects of Purchasing. Purchasing activities are an integral part of ensuring a continuous supply of raw materials necessary for production.

Effective management of the purchasing process not only helps reduce material costs but also ensures quality and continuity of supply. This plays a crucial role in maintaining and enhancing production efficiency. In this section, we will examine closely the metrics related to purchasing costs, quantity of materials purchased, and relationships with suppliers.

The Purchasing Dashboard helps track and assess the performance of the purchasing process, with key metrics including the order rejection rate, total purchase orders, average cycle time, and total purchase value. Notably, the order rejection rate is 3.12%, indicating that a small number of orders do not meet quality requirements. This suggests the need for stricter inspections of suppliers to ensure the quality of goods. The total number of purchase orders is 4012, reflecting the large volume of work that the purchasing department must handle.

The average cycle time is 9.10 days, which helps us understand the time needed to complete a purchase order from placement to receipt. The total purchase value reached $63.79 million, indicating a significant investment in purchasing raw materials and goods.

The "Order Quantity by Vendor" chart shows the number of orders per supplier. For instance, SUPERSALES INC. is the largest supplier with the highest number of orders, followed by Custom Frames and Chicago City Sales. This information helps the company identify key suppliers and evaluate their performance. By knowing the main suppliers, the company can focus on building and maintaining good relationships with them to ensure a stable and quality supply.

Moving to the "Order by Shipmethod" chart, we see the shipping methods used. For example, CARGO TRANSPORTATION accounts for 36% of total orders, followed by OVERNIGHT EXPRESS at 35.75%. This information is crucial for optimizing the shipping process and minimizing shipping costs. By analyzing the shipping methods, the company can adjust its shipping strategy for maximum efficiency.

The "Order Number of Stock, Rejected, and Not Receive per Month" chart displays the number of orders in stock, rejected, and not received monthly. This chart shows a decreasing trend in the number of stock and rejected orders over time, but it also indicates the need for improvements in the process to ensure goods are received fully and on time. Understanding these trends helps the company adjust its purchasing and inventory processes to meet production demands efficiently.

Finally, the "Total Order by Vendor" chart shows the total number of orders per supplier. For example, Superior Bicycles has the largest total number of orders, followed by Professional Athletic and Chicago City Sales. This information helps the company assess the performance of suppliers and identify key suppliers to focus on cooperation.

This detailed analysis helps the leadership and purchasing management of AdventureWorks gain a comprehensive view of purchasing performance. As a result, they can make strategic decisions to improve the purchasing process, reduce the order rejection rate, and optimize cycle time. Monitoring these metrics and charts will help the company maintain and enhance the efficiency of its purchasing activities, ensuring a stable supply of raw materials and better meeting production needs.

REFERENCES

SIGMOD16, 5 March 2024, https://dl.acm.org/doi/10.1145/2882903.2903741.

“AdventureWorks sample databases - SQL Server.” *Microsoft Learn*, 9 May 2024, https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver16&tabs=ssms.

“Azure SQL Database documentation - Azure SQL.” *Microsoft Learn*, https://learn.microsoft.com/en-us/azure/azure-sql/database/?view=azuresql.

“Azure Synapse Analytics - Azure Synapse Analytics.” *Microsoft Learn*, https://learn.microsoft.com/en-us/azure/synapse-analytics/.

“Configure AdventureWorks for Business Intelligence solutions - SharePoint Server.” *Learn Microsoft*, 20 January 2023, https://learn.microsoft.com/en-us/sharepoint/administration/configure-adventureworks.

Cornelia Kraus, and Raul Valverde. “A DATA WAREHOUSE DESIGN FOR THE DETECTION OF FRAUD IN THE SUPPLY CHAIN BY USING THE BENFORD’S LAW.” *American Journal of Applied Sciences*, 7 7 2014, https://thescipub.com/pdf/ajassp.2014.1507.1518.pdf.

“Create an Azure Storage account - Training.” *Microsoft Learn*, https://learn.microsoft.com/en-us/training/modules/create-azure-storage-account/.

“Definition of Data Warehouse.” *Gartner*, https://www.gartner.com/en/information-technology/glossary/data-warehouse.

“Power BI - Data Visualization.” *Microsoft*, https://www.microsoft.com/en-us/power-platform/products/power-bi.