Architectural Diagram :

A diagram of a data flow

AI-generated content may be incorrect.

In this architecture we collect sales, products, return data from GitHub using Azure Data Factory. Data Lands in data lake in raw form and is processes using Data Bricks for transformation. Then it is stored in Azure Synapse for analytical queries and visualized using power BI.

DataSet :

<https://www.kaggle.com/datasets/ukveteran/adventureworks?resource=download&select=AdventureWorks_Products.csv>

<https://www.kaggle.com/datasets/ukveteran/adventure-works?resource=download&select=AdventureWorks_Returns.csv>

<https://www.kaggle.com/datasets/ukveteran/adventure-works?resource=download&select=AdventureWorks_Sales_2015.csv>

**Dataset Overview**

The **Products.csv** file typically includes the following columns:

| **Column Name** | **Description** |
| --- | --- |
| ProductKey | Unique identifier for each product. |
| ProductAlternateKey | Alternate product key or product code. |
| ProductSubcategoryKey | Foreign key to the product subcategory. |
| WeightUnitMeasureCode | Unit of measure for weight. |
| SizeUnitMeasureCode | Unit of measure for size. |
| EnglishProductName | Product name in English. |
| SpanishProductName | Product name in Spanish. |
| FrenchProductName | Product name in French. |
| StandardCost | Standard cost of the product. |
| FinishedGoodsFlag | Indicates if the product is a finished good (1) or a work in progress (0). |
| Color | Product color. |
| SafetyStockLevel | Safety stock level for the product. |

The **AdventureWorks\_Returns.csv** file typically includes the following columns:

| **Column Name** | **Description** |
| --- | --- |
| ReturnID | Unique identifier for the return transaction |
| SalesOrderID | Foreign key linking to the original sales order |
| ProductKey | Identifier for the returned product (links to Products dataset) |
| Quantity | Number of units returned |
| ReturnReasonKey | Key indicating the reason for return (links to Return Reasons table) |
| ReturnDate | Date the return was processed |
| CustomerID | Identifier for the customer who returned the product |
| StoreID | Identifier for the store processing the return |

**Overview of the Sales\_2015.csv Dataset**

**🔹 Typical Columns (Fields):**

| **Column Name** | **Description** |
| --- | --- |
| OrderDate | Date the order was placed |
| StockDate | Date the product was shipped or stocked |
| OrderNumber | Unique identifier for the order |
| ProductKey | Foreign key to identify the product |
| CustomerKey | Foreign key to identify the customer |
| TerritoryKey | Foreign key for sales territory |
| OrderLineItem | Line number in the order (useful for multi-item orders) |
| OrderQuantity | Number of units ordered |
| UnitPrice | Price per unit |
| TotalProductCost | Cost incurred to manufacture or acquire the product |
| SalesAmount | Total sale value = OrderQuantity × UnitPrice |
| TaxAmt | Tax amount for the order |
| Freight | Shipping or freight charges |
| Region | Sales region (e.g., Northwest, Southwest, etc.) |

Data Pipeline Creation:

Ingest data into the chosen source:

Data Factory : Data integration service that enables you to create, schedule and manage pipelines and efficient data movement and transformation between various sources and destinations in Azure.

Data lake Gen 2 : Data lake solution that combines the capabilities of a data lake with a power of azure blob storage allow to store and analyse large volumes of structured and unstructured data.

Azure Data bricks: Data Bricks is a unified platform built on top of Apache spark designed to help data engineers and data scientists collaborate on big data processing and machine learning tasks.it provides tools for data exploration, data processing and building machine learning models.

Synapse analytics: it combines big data and data warehousing into a single integrated form, allowing organization to analyze and process large volumes of data for BI and data analytics purposes.

**Creating Resource Group and various resources in Azure**

**A screenshot of a computer

AI-generated content may be incorrect.**

The product-supplier data is situated in the containers

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Creating Pipeline in Azure Data Factory

A screenshot of a computer

AI-generated content may be incorrect.

Creating Linked Services

A screenshot of a computer

AI-generated content may be incorrect.

Creating various Datasets - Http to ADLS

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Creating a pipeline in ADF, To Store the data in Appropriate destination

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A computer screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Data Cleaning & Transformation in Databricks

A screenshot of a computer program

AI-generated content may be incorrect.

A computer screen shot of a program code

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A black background with orange text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A computer screen shot of text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A close-up of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.