Date Warehouse -
POC using Tableau's SuperStore Dataset

TABLE OF CONTENTS

Subject Area Selection

For this proof-of-concept Data Warehouse, the selected subject area is **Sales and Profit Analysis** for a retail organization.
 The data source used is the **Tableau Superstore dataset**, which represents a fictitious retail company that sells office supplies, furniture, and technology products across multiple regions and customer segments.

This dataset is ideal for a data warehouse project because it:

- Contains clearly defined **business processes** (orders, shipments, and sales transactions)
- Includes **time-based data** (Order Date, Ship Date) for trend analysis.
- Provides **dimensional attributes** such as Product, Customer, Region, and Ship Mode, suitable for developing a **star schema**.
- Offers **measurable facts** (Sales, Profit, Quantity, Discount) that can drive business intelligence reports and dashboards.

Key Stakeholders

Stakeholder	Role / Interest
Sales Managers	Monitor sales performance by region, category, and segment.
Marketing Team	Identify profitable customer segments and products for targeted promotions.
Operations / Logistics Team	Analyze shipping modes, delivery times, and efficiency.
Finance Department	Track profit margins, discount impact, and revenue trends.
Executive Management	View summarized KPIs and strategic dashboards for decision-making.

Business Vision and Objectives

The vision of this Data Warehouse is to **enable data-driven decision-making** by integrating operational sales data into a unified analytical system.

The warehouse aims to transform raw transactional data into **actionable insights** that support performance monitoring, profitability analysis, and strategic planning.

Key Objectives:

1. Consolidate sales transactions from operational systems into a centralized Data Warehouse.
2. Support **multi-dimensional analysis** (by product, region, customer, and time).
3. Provide accurate and consistent **key performance indicators (KPIs)** such as:
    ○ Total Sales and Profit
    ○ Profit Margin %
    ○ Average Discount
    ○ Quantity Sold per Segment
4. Enable fast and reliable **Business Intelligence (BI)** reporting using **SSRS** and **Tableau**.
5. Facilitate forecasting, trend analysis, and identification of underperforming areas.

Insights and Expected Outcomes

From this Data Warehouse, the company can gain the following insights:

1. **Sales Performance Analysis**
    ○ Identify top-performing regions, categories, and customer segments.
    ○ Compare monthly and yearly sales growth trends.

2. **Profitability Insights**

    ○ Analyze the correlation between discounts and profit margins.
    ○ Detect products or subcategories with low profitability despite high sales.

3. **Customer Segmentation Analysis**

    ○ Discover the most valuable customer segments (Consumer, Corporate, Home Office).
    Understand geographic distribution of customers and regional profitability.

4. **Operational Efficiency**

    ○ Evaluate average shipping times across different ship modes.
    ○ Measure on-time vs. delayed shipments (based on Order Date and Ship Date).

5. **Strategic Planning**

- Support business decisions such as pricing strategies, promotional offers, and inventory optimization.
- Enable "what-if" analyses in Tableau to simulate the impact of discount policies.

Conclusion

The project successfully demonstrated the end-to-end design and implementation of a **Retail Data Warehouse** using **Microsoft SQL Server, SSIS, SSRS, and Tableau**.
 A star schema model was used to organize data into dimension and fact tables, allowing efficient data analysis and reporting.

Using **SSIS**, data from the operational system was cleansed, transformed, and loaded into the warehouse. The dimensional modeling approach simplified complex queries and improved analytical performance.

SSRS reports were developed to provide operational insights such as:

- Regional sales and profit performance

- Product category profitability

- Customer segment sales contribution

- Shipping mode analysis

Additionally, the **Tableau Dashboard** integrated all these analyses into an interactive visual interface, helping business users derive insights easily.
 Key insights from the BI analysis include:

- **West region** recorded the highest total sales and profit.

- **Consumer segment** contributed the majority of overall revenue.

- **Technology category** was the most profitable product segment.

- **Standard Class** was the most used shipping mode, with the highest order count.

- Sales and profit showed a steady upward trend from 2014 to 2017.

The project successfully achieved its objectives by integrating ETL, reporting, and visualization tools to enable data-driven decision-making.
 This proof-of-concept establishes a scalable foundation for further BI enhancements, predictive analytics, and real-time reporting capabilities.

Appendix

Appendix A – Table Creation Scripts

```sql
CREATE TABLE DimCustomer (
    CustomerID NVARCHAR(20) PRIMARY KEY,
    CustomerName NVARCHAR(100),
    Segment NVARCHAR(50)
);

CREATE TABLE DimProduct (
    ProductID NVARCHAR(20) PRIMARY KEY,
    Category NVARCHAR(50),
    SubCategory NVARCHAR(50),
    ProductName NVARCHAR(150)
);

CREATE TABLE DimRegion (
    Region_ID INT IDENTITY(1,1) PRIMARY KEY,
    Country NVARCHAR(100),
    State NVARCHAR(100),
    City NVARCHAR(100),
    Region NVARCHAR(50),
    PostalCode VARCHAR(50)
);

CREATE TABLE DimShipMode (
    ShipModeKey INT IDENTITY(1,1) PRIMARY KEY,
    ShipMode NVARCHAR(50)
);

CREATE TABLE DimDate (
    DateKey INT PRIMARY KEY,
    Date DATE,
    Year INT,
    Month INT,
    Day INT
);
```

```sql
CREATE TABLE FactSales (
    FactID INT IDENTITY(1,1) PRIMARY KEY,
    RowID INT,
    OrderID NVARCHAR(20),
    OrderDateKey INT,
    ShipDateKey INT,
    ShipModeKey INT,
    CustomerID NVARCHAR(20),
    ProductID NVARCHAR(20),
    Region_ID INT,
    Sales FLOAT,
    Quantity INT,
    Discount FLOAT,
    Profit FLOAT,
    FOREIGN KEY (ShipModeKey) REFERENCES DimShipMode(ShipModeKey),
    FOREIGN KEY (Region_ID) REFERENCES DimRegion(Region_ID),
    FOREIGN KEY (CustomerID) REFERENCES DimCustomer(CustomerID),
    FOREIGN KEY (ProductID) REFERENCES DimProduct(ProductID),
    FOREIGN KEY (OrderDateKey) REFERENCES DimDate(DateKey),
    FOREIGN KEY (ShipDateKey) REFERENCES DimDate(DateKey)
);
```

Appendix B – Sample ETL SQL Queries (SSIS Source Commands)

For DimCustomer

```sql
SELECT DISTINCT Customer_ID, Customer_Name, Segment
FROM dbo.Orders;
```

For DimProduct

```sql
SELECT DISTINCT Product_ID, Category, Sub_Category, Product_Name
FROM dbo.Orders;
```

For DimRegion

```sql
SELECT DISTINCT Country, State, City, Region, PostalCode
FROM dbo.Orders;
```

For DimShipMode

```sql
SELECT DISTINCT Ship_Mode FROM dbo.Orders;
```

For DimDate

```sql
SELECT DISTINCT
    CONVERT(INT, FORMAT(Order_Date,'yyyyMMdd')) AS DateKey,
    Order_Date AS Date,
    YEAR(Order_Date) AS Year,
    MONTH(Order_Date) AS Month,
    DAY(Order_Date) AS Day
FROM dbo.Orders
UNION
SELECT DISTINCT
    CONVERT(INT, FORMAT(Ship_Date,'yyyyMMdd')) AS DateKey,
    Ship_Date AS Date,
    YEAR(Ship_Date) AS Year,
    MONTH(Ship_Date) AS Month,
    DAY(Ship_Date) AS Day
FROM dbo.Orders;
```

For FactSales

```sql
SELECT
    Row_ID,
    Order_ID,
    Order_Date,
    Ship_Date,
    Ship_Mode,
    Customer_ID,
    Product_ID,
    Region_ID,
    Sales,
    Quantity,
    Discount,
    Profit
FROM dbo.Orders;
```

Appendix C – SSRS

Report 1 – Sales and Profit by Region

Purpose: Compare total sales and profit across regions.
Visualization Type: Column Chart or Bar Chart

```sql
SELECT
    r.Region AS [Region],
    SUM(f.Sales) AS [Total Sales],
    SUM(f.Profit) AS [Total Profit]
FROM FactSales f
JOIN DimRegion r ON f.RegionKey = r.RegionKey
GROUP BY r.Region
ORDER BY [Total Sales] DESC;
```

Report 2 – Sales and Profit by Customer Segment

Purpose: Show which customer segment (Consumer, Corporate, Home Office) contributes most.
Visualization Type: Pie Chart

```sql
SELECT
    c.Segment AS [Customer Segment],
    SUM(f.Sales) AS [Total Sales],
    SUM(f.Profit) AS [Total Profit]
FROM FactSales f
JOIN DimCustomer c ON f.CustomerID = c.CustomerID
GROUP BY c.Segment
ORDER BY [Total Sales] DESC;
```

Report 3 – Top 10 Products by Sales

Purpose: Display the best-performing products by sales.
Visualization Type: Horizontal Bar Chart

```sql
SELECT TOP 10
    p.ProductName AS [Product Name],
    p.Category AS [Category],
    p.SubCategory AS [Sub-Category],
    SUM(f.Sales) AS [Total Sales],
    SUM(f.Profit) AS [Total Profit]
FROM FactSales f
JOIN DimProduct p ON f.ProductID = p.ProductID
GROUP BY p.ProductName, p.Category, p.SubCategory
ORDER BY [Total Sales] DESC;
```

Report 4 – Sales and Orders by Ship Mode

Purpose: Compare total sales and number of orders by shipping method.
Visualization Type: Column Chart

```sql
SELECT
    sm.ShipMode AS [Ship Mode],
    SUM(f.Sales) AS [Total Sales],
    COUNT(f.OrderID) AS [Order Count]
FROM FactSales f
JOIN DimShipMode sm ON f.ShipModeKey = sm.ShipModeKey
GROUP BY sm.ShipMode
ORDER BY [Total Sales] DESC;
```