Data Warehouse Project – Milestone 1

Guided by: Prof. Joseph Kinn

Team Alpha: Sakshi Singh, Amulya Walimbe, Reeya Patra, Shivani Pandeti
Course: IST 722
Date: 09/29/25

---

1. Executive Summary

The goal of this project is to design and implement a data warehouse for **Global Superstore** data using **Snowflake** as the data platform and **Tableau/Power BI** for visualization. The warehouse will support business users in analyzing sales, shipping performance, customer behavior, and product performance.

---

2. Business Processes & Key Questions

The data warehouse is built to answer critical business questions in the following areas:

1. **Sales Analysis**
    - What are total sales, profit, and profit margin by region, product, and customer segment?
    - Which categories drive the most revenue?

2. **Shipping Analysis**
    - What is the average shipping lead time?
    - What percentage of orders breach SLA (service level agreements)?

3. **Customer Behavior**
    - What is the monthly revenue contribution of each customer?
    - Which customers generate the highest profit?

4. **Product Performance**

- Which products sell the most each month?
- Which subcategories contribute to profit growth?

---

3. Business Matrix (Bus Matrix)

Fact \ Dim	DimDate	DimMonth	DimCustomer	DimProduct	DimGeography	DimShipping
FactSales	✓(Order)		✓	✓	✓	✓
FactShipping	✓(Order/Ship)				✓	✓
FactCustomerMonthly		✓	✓		✓	
FactProductMonthly		✓		✓	✓	

---

4. Star Schema Designs

4.1 FactSales Star Schema

Business Process: Sales Transactions

Grain: One row per *order line item*

Keys: DateKey, CustomerKey, ProductKey, GeographyKey, ShipKey

Measures: Sales, Profit, Quantity, Discount, ShippingCost, ProfitMargin

FactSales Table

Keys: OrderID (degenerate), OrderDateKey, CustomerKey, ProductKey, GeographyKey, ShipKey

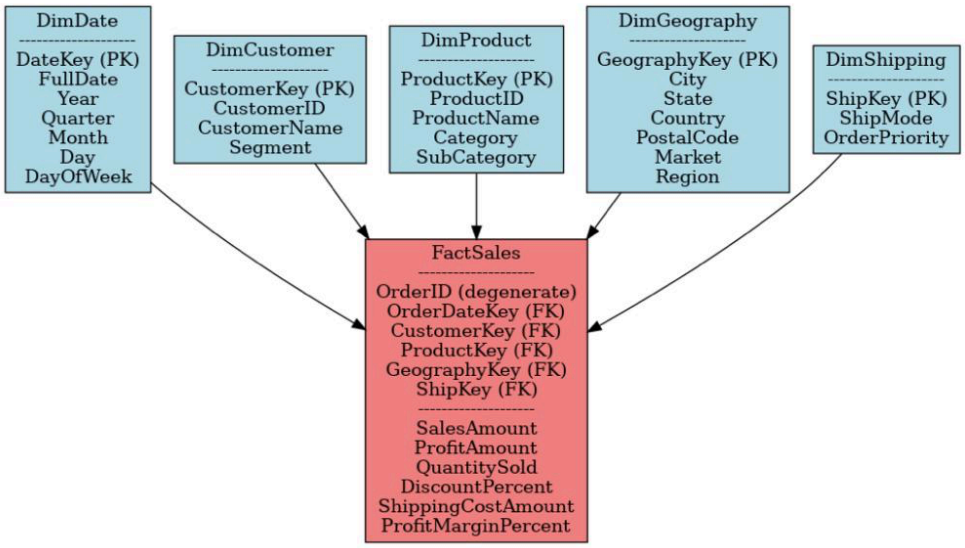
Measures: SalesAmount, ProfitAmount, QuantitySold, DiscountPercent, DiscountAmount, ShippingCostAmount, ProfitMarginPercent

Dimensions

- *DimDate:* DateKey, FullDate, Year, Quarter, Month, DayOfWeek
- *DimCustomer:* CustomerKey, CustomerID, CustomerName, Segment
- *DimProduct:* ProductKey, ProductID, ProductName, Category, SubCategory
- *DimGeography:* GeographyKey, City, State, Country, Region, Market, PostalCode
- *DimShipping:* ShipKey, ShipMode, OrderPriority

Notes

- **Degenerate Dimension:** OrderID is treated as a degenerate dimension, stored directly in the fact table since it is used for reporting but does not require descriptive attributes.
- **Conformed Dimensions:** DimDate, DimCustomer, DimProduct, and DimGeography are shared with other fact tables, ensuring consistency across the warehouse.
- **Analytical Use:** This schema supports queries on sales and profitability across customers, products, time, shipping methods, and geographic regions.



4.2 FactShipping Star Schema

Business Process: Shipping & Delivery Performance

Grain: One row per shipment

FactShipping Table

Keys: OrderID (degenerate), OrderDateKey, ShipDateKey, CustomerKey, GeographyKey, ShipKey

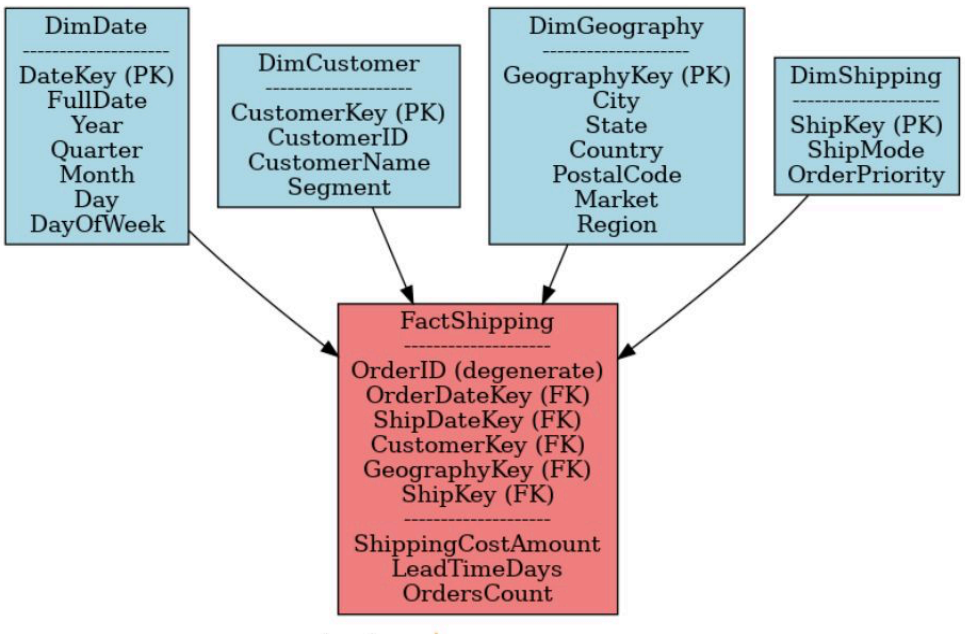
Measures: ShippingCostAmount, LeadTimeDays, OrdersCount

Dimensions

- DimDate: DateKey, FullDate, Year, Quarter, Month, DayOfWeek
- DimCustomer: CustomerKey, CustomerID, CustomerName, Segment
- DimGeography: GeographyKey, City, State, Country, PostalCode, Market, Region
- DimShipping: ShipKey, ShipMode, OrderPriority

Notes

- OrderID is modeled as a degenerate dimension because it is useful for reporting but does not require a separate table.
- DimDate, DimCustomer, and DimGeography are conformed dimensions shared across other facts (FactSales, FactCustomerMonthly, FactProductMonthly) to ensure consistency.
- This schema supports analysis of shipping cost, delivery lead times, SLA compliance, and customer-level shipping behavior across time and geography.



4.3 FactCustomerMonthly Star Schema

Business Process: Monthly Customer Performance

Grain: One row per Customer × Month × Geography

FactCustomerMonthly Table

Keys: CustomerKey, DateKey, GeographyKey

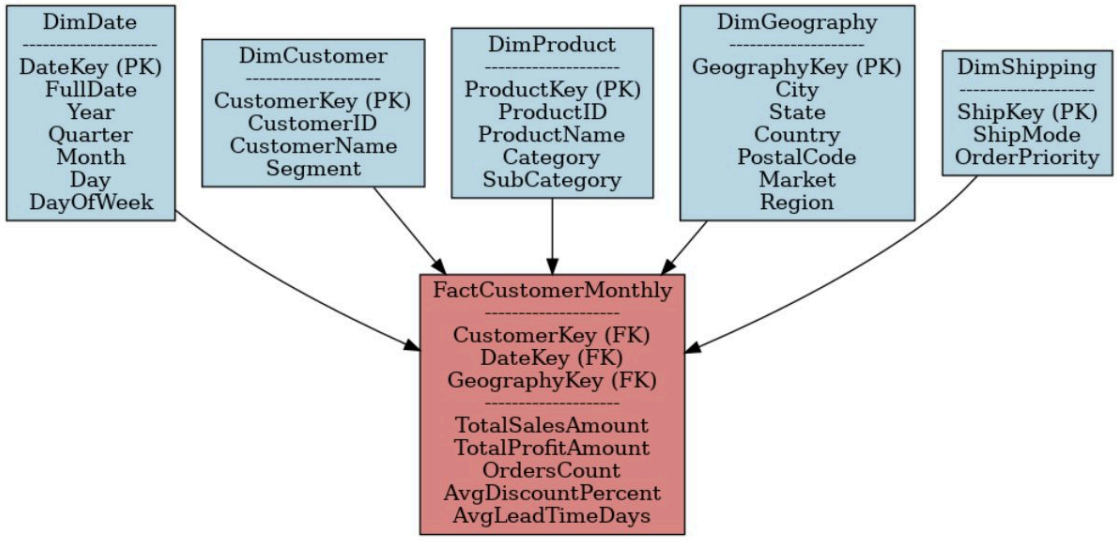
Measures: TotalSalesAmount, TotalProfitAmount, OrdersCount, AvgDiscountPercent, AvgLeadTimeDays

Dimensions

- DimDate: DateKey, FullDate, Year, Quarter, Month, Day, DayOfWeek
- DimCustomer: CustomerKey, CustomerID, CustomerName, Segment
- DimProduct: ProductKey, ProductID, ProductName, Category, SubCategory
- DimGeography: GeographyKey, City, State, Country, PostalCode, Market, Region
- DimShipping: ShipKey, ShipMode, OrderPriority

Notes

- The schema is designed for monthly aggregation of customer metrics across geographies.
- DimDate and DimGeography are conformed dimensions, ensuring consistent reporting across facts.
- This fact supports KPIs such as customer profitability, order volume, discount trends, and average shipping lead time by customer segment and region.



4.4 FactProductMonthly Star Schema

Business Process: Monthly Product Performance

Grain: One row per Product × Month × Geography

FactProductMonthly Table

Keys: ProductKey, DateKey, GeographyKey

Measures: TotalSalesAmount, TotalProfitAmount, TotalQuantitySold, OrdersCount, AvgDiscountPercent

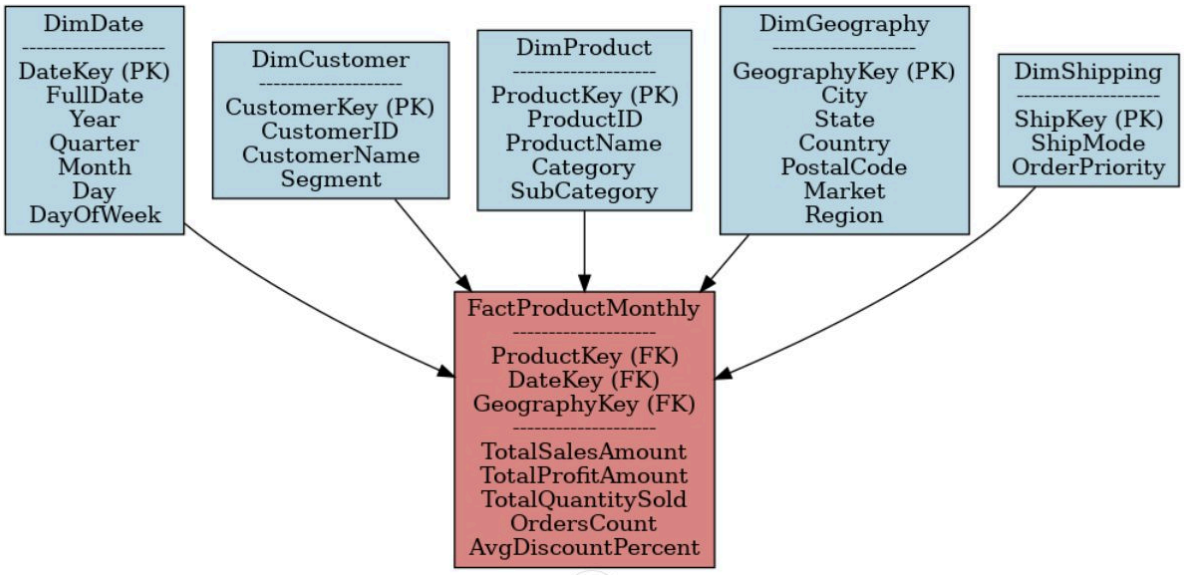
Dimensions

- DimDate: DateKey, FullDate, Year, Quarter, Month, DayOfWeek
- DimCustomer: CustomerKey, CustomerID, CustomerName, Segment
- DimProduct: ProductKey, ProductID, ProductName, Category, SubCategory
- DimGeography: GeographyKey, City, State, Country, PostalCode, Market, Region
- DimShipping: ShipKey, ShipMode, OrderPriority

Notes

- This schema provides monthly aggregation of product-level performance across regions.
- DimProduct and DimGeography are conformed dimensions, enabling consistent analysis across sales and customer facts.

- Supports KPIs such as top-selling products, category profitability, discount impact, and product performance by region and time.



---

5. Next Steps (Milestone 2 & 3)

- **Milestone 2**: Implement ETL in Snowflake (load CSV → staging → dimensions/facts). Build initial Tableau KPIs.

- **Milestone 3**: Finalize ETL automation, design Tableau dashboards (overview, sales, shipping, customer, product), and prepare demo video.

---

6. References

- Kimball, R. & Ross, M. *The Data Warehouse Toolkit.*
