

+

**D.K.T.E. Society's Textile and Engineering
Institute, Ichalkaranji.**

**(An Empowered Autonomous Institute,
Affiliated to Shivaji University, Kolhapur)**

Department of Artificial Intelligence and Data Science

2024-25



**Project Synopsis On
Superstore Sales Data Warehouse**

Under The Guidance Of

PROF.SATISH P PISE

Submitted By:

- | | |
|-------------------|----------|
| 1. Sanket Bhagat | 22UAD301 |
| 2. Rushikesh Kate | 22UAD302 |
| 3. Umer Mujawar | 22UAD304 |

PROF.SATISH P PISE
Project Guide

Prof. (Dr.) T.I.Bagban
Head of Department

Index

CONTENTS	PAGE NO.
1. Introduction	1
2. Technologies	2
3. SQL Queries	3-6
4. Dimension Schema	7
5. Graphical Representation	8
6. Conclusion	9

1. Introduction

In today's competitive retail environment, businesses need efficient data solutions to understand trends, monitor performance, and make strategic decisions. This project, **Superstore Sales Data Warehouse**, aims to create a consolidated data platform to analyze transactional data from multiple perspectives—customers, products, orders, sellers, and sales metrics.

Using **Snowflake**, we build a data warehouse with a **star schema architecture**. The central fact table, **SalesFact**, captures transactional data such as sales, quantity, discounts, and profits. It connects to relevant dimension tables (Customer, Product, Order, and Seller) for enriched insights.

This project enables interactive reporting using **Power BI**, providing critical insights into sales trends, customer preferences, product performance, and seller contribution to enhance business operations.

Key Tables:

- **CustomerDimension**: Stores customer-specific details like name, location, and segment.
- **ProductDimension**: Contains product information such as category, sub-category, and product name.
- **OrderDimension**: Tracks order details, shipping dates, and shipping modes.
- **SellerDimension**: Captures seller details such as seller name and contact information.
- **SalesFact**: Stores transactional data, including sales, quantity, discounts, and profits.

2. Technologies

- **Snowflake:**

A cloud-based data warehousing platform providing scalability, elasticity, and efficient data management for structured and semi-structured data.

- **SnowSQL:**

A command-line tool for managing Snowflake objects, performing data manipulation, and automating processes such as data loading and querying.

- **PowerBI:**

A business intelligence tool used to visualize data insights and create interactive dashboards for better decision-making.

- **VSCode:**

A versatile code editor used for writing SQL scripts and integrating with Snowflake efficiently.

3. SQL Queries

File Format Definition

```
CREATE OR REPLACE FILE FORMAT my_csv_format  
TYPE = 'CSV'  
FIELD_OPTIONALLY_ENCLOSED_BY = ''''  
SKIP_HEADER = 1  
DATE_FORMAT = 'DD-MM-YYYY';
```

Superstore Data Table

```
CREATE OR REPLACE TABLE SuperstoreData (  
  Row_ID INT,  
  Order_ID STRING,  
  Order_Date DATE,  
  Ship_Date DATE,  
  Ship_Mode STRING,  
  Customer_ID STRING,  
  Customer_Name STRING,  
  Segment STRING,  
  Country STRING,  
  City STRING,  
  State STRING,  
  Postal_Code STRING,  
  Region STRING,  
  Product_ID STRING,  
  Category STRING,  
  Sub_Category STRING,  
  Product_Name STRING,  
  Seller_ID STRING,  
  Sales FLOAT,  
  Quantity INT,  
  Discount FLOAT,  
  Profit FLOAT  
);
```

Data Loading from Stage

```
COPY INTO SuperstoreData
FROM @my_stage/Superstore.csv.gz
FILE_FORMAT = (FORMAT_NAME = 'my_csv_format')
ON_ERROR = 'CONTINUE';
```

Dimension Tables

Customer Dimension

```
CREATE OR REPLACE TABLE CustomerDimension AS
SELECT DISTINCT
    Customer_ID,
    Customer_Name,
    Segment,
    Country,
    City,
    State,
    Postal_Code,
    Region
FROM SuperstoreData;
```

Product Dimension

```
CREATE OR REPLACE TABLE ProductDimension AS
SELECT DISTINCT
    Product_ID,
    Category,
    Sub_Category,
    Product_Name
FROM SuperstoreData;
```

Order Dimension

```
CREATE OR REPLACE TABLE OrderDimension AS
SELECT DISTINCT
    Order_ID,
    Order_Date,
    Ship_Date,
    Ship_Mode
FROM SuperstoreData;
```

Seller Dimension

```
CREATE OR REPLACE TABLE SellerDimension AS
SELECT DISTINCT
    Seller_ID,
    Seller_Name,
    Seller_Phone
FROM SuperstoreData;
```

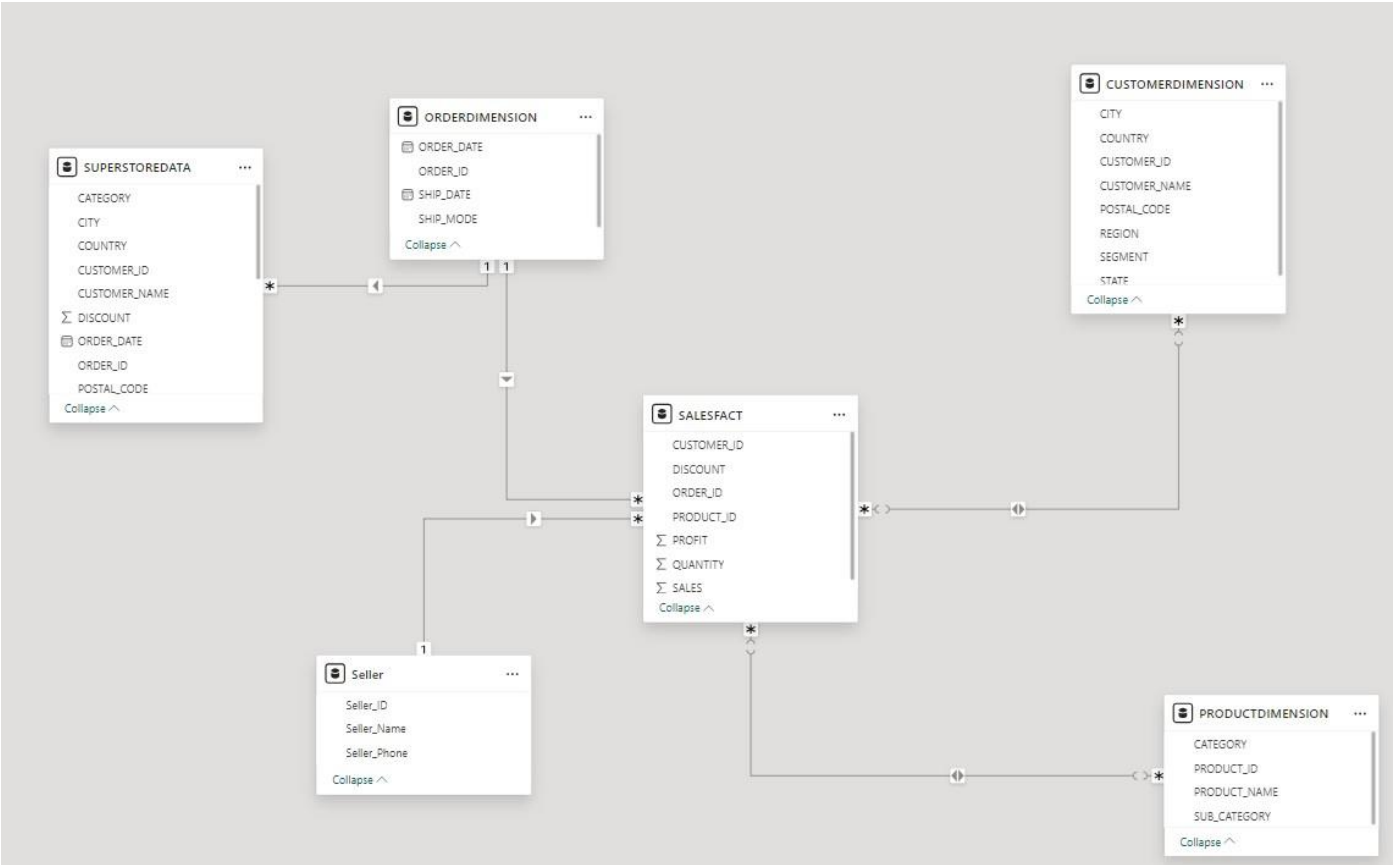
Fact Table: Sales Fact

```
CREATE OR REPLACE TABLE SalesFact AS
SELECT
    Order_ID,
    Customer_ID,
    Product_ID,
    Seller_ID,
    Sales,
    Quantity,
    Discount,
    Profit
FROM SuperstoreData;
```

Query for Analyzing Top Products, Sellers, and Profits

```
SELECT
    c.Customer_Name,
    p.Product_Name,
    s.Seller_Name,
    f.Sales,
    f.Profit
FROM SalesFact f
JOIN CustomerDimension c ON f.Customer_ID = c.Customer_ID
JOIN ProductDimension p ON f.Product_ID = p.Product_ID
JOIN SellerDimension s ON f.Seller_ID = s.Seller_ID
ORDER BY f.Profit DESC;
```


4. Dimension Schema



5. Graphical Representation

1. Scatter Plot



2. Bar Chart



3. Line Chart



6. Conclusion

The **Superstore Sales Data Warehouse** project demonstrates the integration of modern data tools to build a robust and scalable data platform. **Snowflake** enabled efficient data management, and **SnowSQL** facilitated smooth data loading and querying.

This data warehouse, with its star schema, supports advanced reporting through **Power BI**, offering insights into sales performance, customer behavior, product trends, and seller contributions. These insights allow businesses to make data-driven decisions, enhance operational efficiency, and optimize customer experiences.

The addition of the **SellerDimension** ensures deeper analysis of seller performance, fostering better collaborations and improving sales strategies. This project showcases how cloud-based technologies like Snowflake can transform retail data into actionable insights, driving growth and innovation.