



PRESENTER 1 — REEYA

Slide 2 – Project Overview

“Hello everyone, I’ll begin the presentation by giving a brief overview of the project. Our goal was to build a complete end-to-end **Data Warehouse and BI analytics solution** using the Global Superstore dataset. We designed a Snowflake-based ETL pipeline that converts raw transactional data into a clean, optimized analytical model. The warehouse supports detailed reporting on sales, products, customers, and shipping. Finally, we visualized these insights using Power BI dashboards.”

Slide 3 – Business Problem

“The organization originally faced several challenges:

- Large volumes of sales data stored in disconnected files
- No centralized warehouse or single source of truth
- Slow and repetitive manual reporting
- No unified view of customers, regions, or product categories
- And no automated KPI monitoring.

Our goal was to **solve these problems** by building a scalable warehouse and a BI solution that supports fast, data-driven decision-making.”

Slide 4 – Objectives

“We defined five main objectives:

1. Transform raw transactional data into a structured analytical format
 2. Build a Snowflake warehouse with conformed dimensions
 3. Automate the ETL pipeline from Extract → Transform → Load
 4. Develop enterprise-quality dashboards for four business processes
 5. Enable trend analysis, regional performance, customer segmentation, and delivery KPI measurement.”
-

Slide 5 – Data Source

“Our dataset is the well-known **Global Superstore dataset**, containing around **51,000 rows and 24 attributes**.

It includes detailed information about orders, customers, products, geography, discounts, and shipping. The dataset is structured and came in CSV format, which we uploaded into Snowflake’s **RAW_STAGE** schema as the starting point for our ETL pipeline.”



PRESENTER 2 — Sakshi

Slide 6 – System Architecture

“Our architecture follows a **three-layer warehouse design**, which separates ingestion, transformation, and presentation.

- The RAW_STAGE layer holds the raw CSV data.
 - The TRANSFORM layer applies cleaning, standardization, surrogate keys, and business rules.
 - The DW layer stores the final dimension and fact tables, ready for BI tools like Power BI.
- This modular architecture ensures scalability and maintainability.”
-

Slide 7 – ETL Methodology

“Our ETL pipeline is as follows:

- **Extract:** We loaded the CSV file into Snowflake’s internal stage and created the staging table STG_GLOBAL_SUPERSTORE.
 - **Transform:** This is the most important part — here we cleaned data, standardized fields, generated MD5 surrogate keys, and built all fact and dimension tables.
 - **Load:** Finally, we moved fully validated tables into the DW schema and performed row count checks to ensure data integrity.”
-

Slide 8 – Key Transformations

“We implemented several key transformations:

- MD5 surrogate keys for Geography and Shipping to maintain uniqueness
 - A role-playing Date dimension that supports both Order Date and Ship Date
 - Monthly rollups for customer and product performance
 - Days-to-ship calculation using Snowflake’s DATEDIFF
 - Profit margin, average order value, and repeat purchase metrics
 - Standardized segmentation for regions and markets
- These transformations allow powerful, flexible analytics.”
-

Slide 9 – Star Schema Model

“Our final model follows a **Star Schema**.

We created five conformed dimensions:

Date, Customer, Product, Geography, and Shipping.

And four fact tables:

Sales, Shipping, Monthly Customer Performance, and Monthly Product Performance.

This schema supports fast queries, easy drill-downs, and consistent reporting across all dashboards.”

PRESENTER 3 — SHIVANI

Slide 10 – Data Validation Checks

“To ensure accuracy, we performed extensive validation.

This included:

- Row count reconciliation between Stage, Transform, and DW
- Duplicate key checks
- Null key detection
- Date continuity checks
- Profit, discount, and quantity sanity checks
- Validation of the late/on-time shipping logic

These checks guaranteed clean, reliable data in the warehouse.”

Slide 11 – BI Layer Overview

“Our BI solution is organized into four dashboards:

1. Sales Overview
2. Product Performance
3. Customer Insights
4. Shipping & Delivery Performance

Each dashboard is designed for a different set of business users — such as executives, category managers, the logistics team, and customer relationship teams.

Together, they give a complete 360-degree view of the business.”

Slide 12 – Sales Overview Dashboard

“The Sales dashboard presents key KPIs: Total Sales, Profit, Orders, and Profit Margin %

It includes yearly sales trends, regional performance, category balance, and top 10 customers.

From this dashboard we observed steady year-over-year sales growth, strong contribution from the Technology category, and the East region as the most profitable market.”

Slide 13 – Product Performance Dashboard

“This dashboard focuses on category-level and sub-category-level performance.

It includes a monthly trend line, a category/sub-category treemap, a profitability matrix, and the top 15 products by profit.

Key insights include: Q4 peak sales every year, Phones and Chairs as top-performing products, and Bookcases and Furnishings showing weaker margins.”



PRESENTER 4 — Amulya

Slide 14 – Customer Insights Dashboard

“The Customer Insights dashboard highlights segment-wise contributions, Avg order value, repeat ordering trends, and geographic distribution.

Our analysis showed that the Consumer segment contributes almost half of the revenue, Corporate customers have the highest Avg order value, and repeat orders increase significantly over the months.

This helps identify retention opportunities and high-value customer groups.”

Slide 15 – Shipping & Delivery Performance

“This dashboard tracks delivery efficiency. We analyzed on-time vs late deliveries, days-to-ship distribution, shipping cost by mode, and regional risk zones.

Around 17% of deliveries were late, Standard and Second-Class shipments dominate volume, and regions like APAC and EMEA show the highest delays.

These insights help identify logistics improvements and SLA risks.”

Slide 16 – Business Impact

“This system helps multiple stakeholders:

- Executives get a real-time view of profitability
- Category managers optimize product mix
- Customer teams target high-value segments
- Logistics teams track delays and cost efficiency

Overall, the warehouse replaces manual reporting with automated, accurate dashboards.”

Slide 17 – Conclusion

“In conclusion, our project demonstrates the full lifecycle of a Data Warehouse solution: We designed ETL pipelines, implemented a scalable warehouse in Snowflake, validated data integrity, and developed rich BI dashboards.

This system is accurate, fast, and provides actionable insights across sales, customers, products, and logistics.

Thank you.”