

Case Study

ETL and Reporting Solution in Retail (Supermarket) Industry

Problem Statement

A supermarket chain faced critical challenges that impeded their ability to make timely and informed business decisions:

1. **Slowness of Reports:**
 - Reports took hours to generate, delaying operational and strategic insights.
 - High query complexity on large datasets led to performance bottlenecks.
 2. **Fragmented Data Sources:**
 - Data was spread across multiple systems: SQL Server, Teradata, Salesforce, and CSV files.
 - Manual consolidation resulted in inconsistencies and errors.
 3. **Outdated Reporting Structure:**
 - Reports were poorly designed, lacking actionable insights and a unified structure.
 - Key performance indicators (KPIs) were scattered across multiple, unrelated reports.
-

Proposed Solution

To address these challenges, the supermarket chain implemented a robust ETL pipeline using **SQL Server Integration Services (SSIS)**, consolidated data into a **SQL Server** data warehouse, and restructured reporting with **Power BI**.

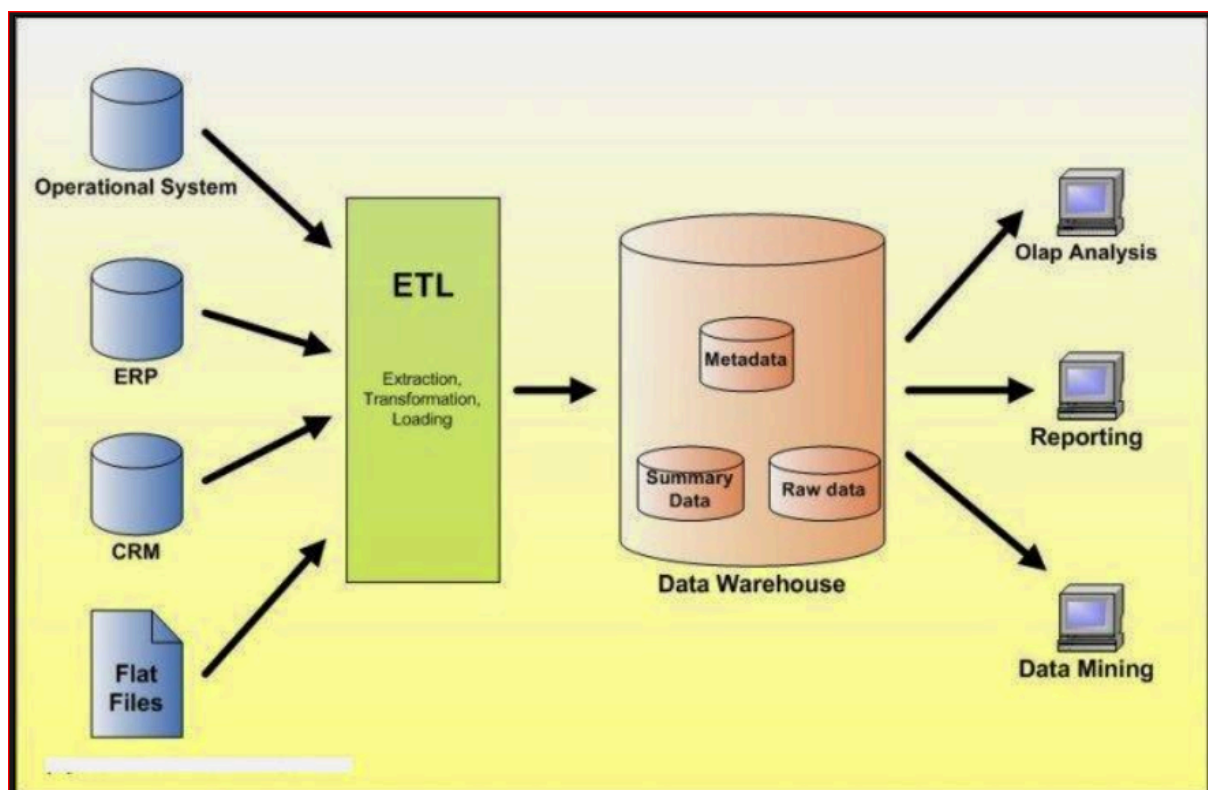
Scenario Details

Data Sources

1. **SQL Server:**
 - Stores transactional sales data from POS systems.
 - *Example Data:* Transaction ID, store ID, product ID, quantity sold, sales amount.
 - *Daily Volume:* 5 million rows.
2. **Teradata:**
 - Contains historical and aggregated sales data for trend analysis.
 - *Example Data:* Monthly sales summaries, customer loyalty data, and product trends.
 - *Daily Volume:* 3 million rows.
3. **Salesforce:**

- Maintains customer data, including loyalty profiles and purchase history.
 - *Example Data:* Customer ID, loyalty tier, lifetime value, and recent interactions.
 - *Daily Volume:* 1 million rows.
4. **CSV Files:**
- Regional campaign data and inventory restocking schedules provided by regional managers.
 - *Example Data:* Campaign ID, region, forecasted sales, and restocking dates.
 - *Daily Volume:* 1 million rows.

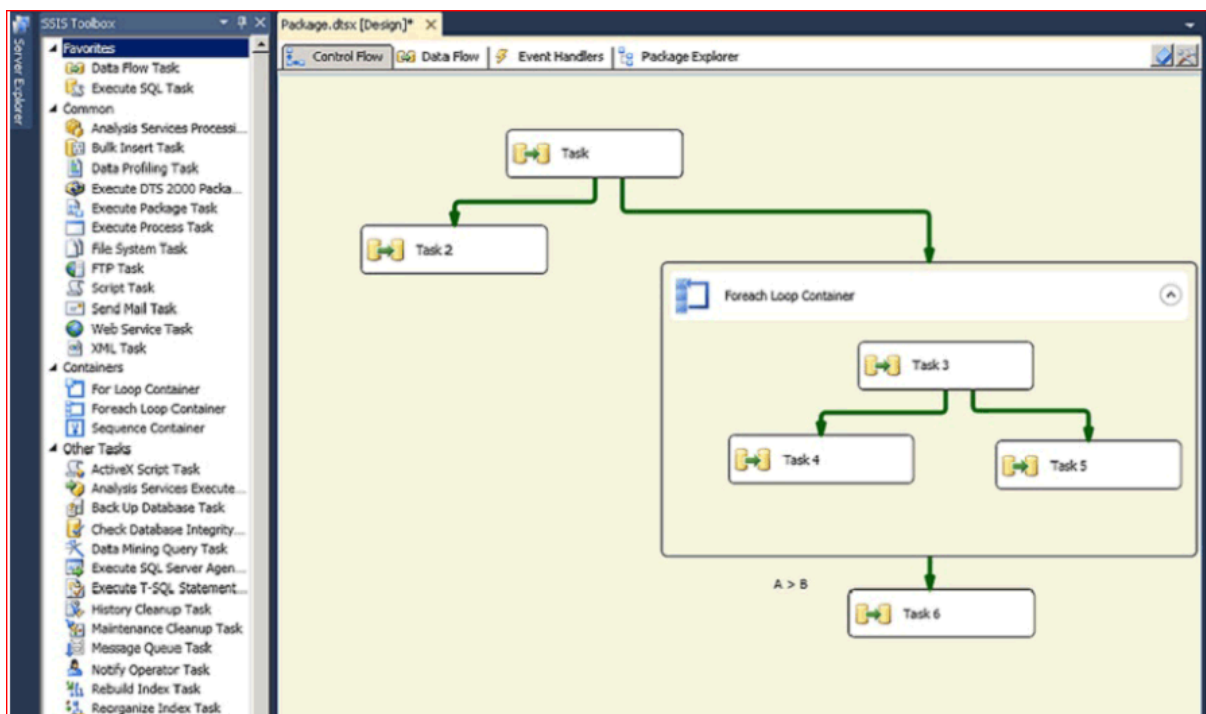
Solution Architecture



1. ETL Tool: SQL Server Integration Services (SSIS)

- **Data Extraction:**
 - SSIS connects to SQL Server, Teradata, Salesforce, and ingests CSV files via FTP or direct upload.
 - Incremental loading ensures only new and updated records are processed.
- **Data Transformation:**
 - **Data Cleansing:**
 - Remove duplicates, resolve null values, and validate data types.
 - **Standardization:**

- Harmonize date formats, product codes, and region identifiers across all sources.
- **Enrichment:**
 - Join Salesforce loyalty data with POS sales for enhanced customer insights.
 - Integrate campaign data with historical sales from Teradata.
- **Aggregation:**
 - Calculate metrics like revenue, profit margins, customer lifetime value, and sales per region.
- **Data Loading:**
 - Transformed data is loaded into **SQL Server** as the centralized data warehouse.
 - Partitioned tables by date, region, and product category optimize query performance.



2. Target Data Warehouse: SQL Server

- **Schema Design:**
 - A star schema with fact tables for sales, inventory, and campaign performance.
 - Dimension tables for products, customers, regions, and time.
- **Performance Enhancements:**
 - Clustered indexing on transaction dates and product IDs.
 - Compression and partitioning minimize storage overhead and improve query speed.

3. Power BI for Reporting

- **Restructured Dashboards:**
 - Unified dashboards for sales, inventory, and marketing campaigns.

- KPIs such as daily revenue, stock levels, and campaign ROI are visually highlighted.
 - **Optimizations:**
 - Power BI leverages SQL Server's aggregated data for faster query execution.
 - DirectQuery mode ensures real-time insights for frequently accessed data.
 - Scheduled refresh every 6 hours ensures dashboards remain current.
-

Reporting Capabilities

1. **Sales Performance Dashboard:**
 - Track daily and monthly revenue trends by store, region, and product category.
 - Identify top-selling products and customer segments driving sales growth.
 2. **Customer Insights Dashboard:**
 - Analyze customer loyalty tiers, purchase patterns, and lifetime value.
 - Segment customers by region and preferences to tailor promotions.
 3. **Inventory and Campaign Dashboard:**
 - Monitor stock availability and flag potential stockouts or overstocks.
 - Evaluate the ROI and sales uplift from marketing campaigns.
-

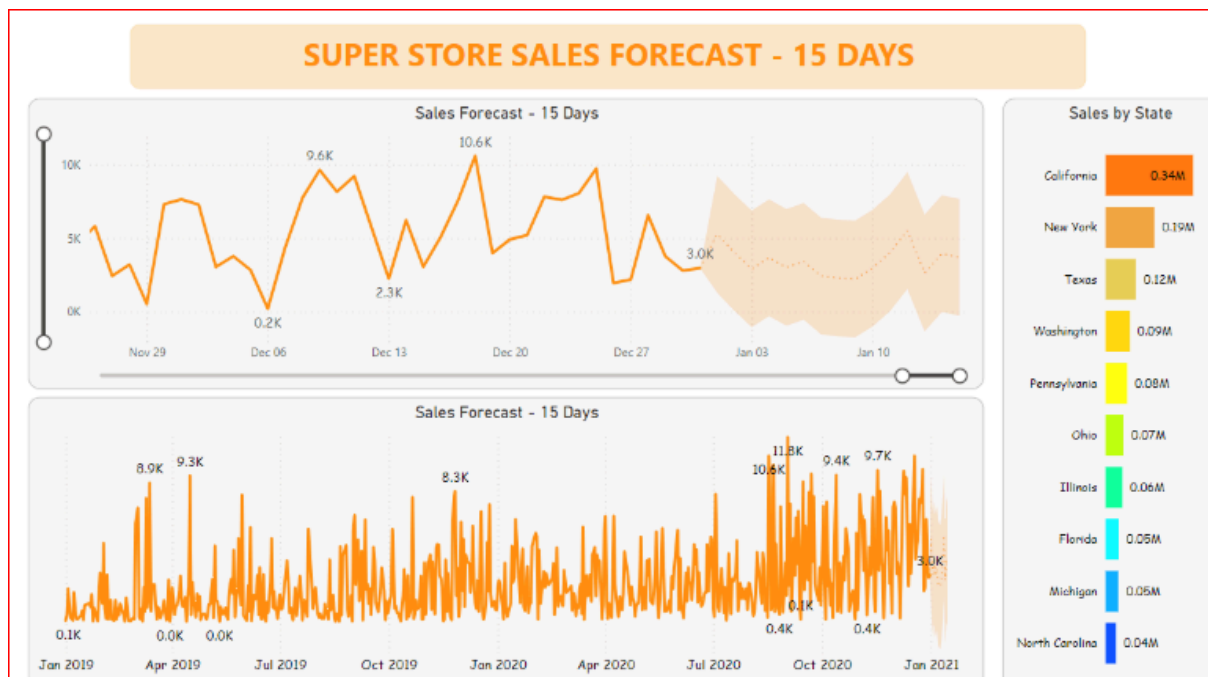
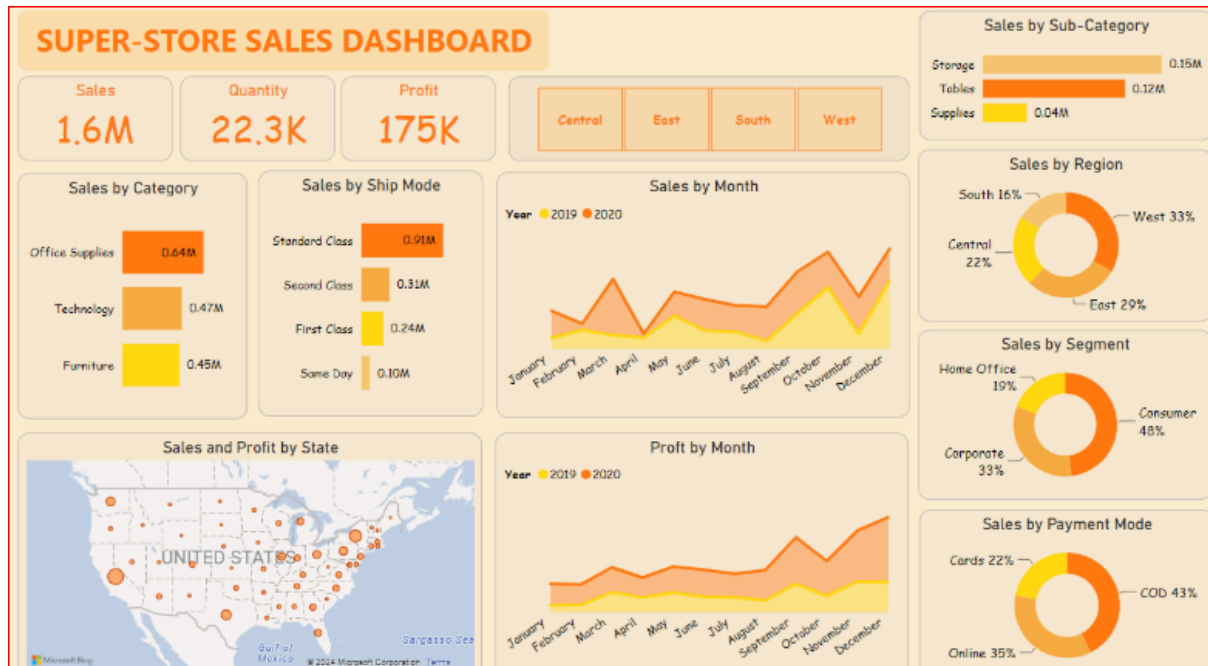
Scalability and Performance Enhancements

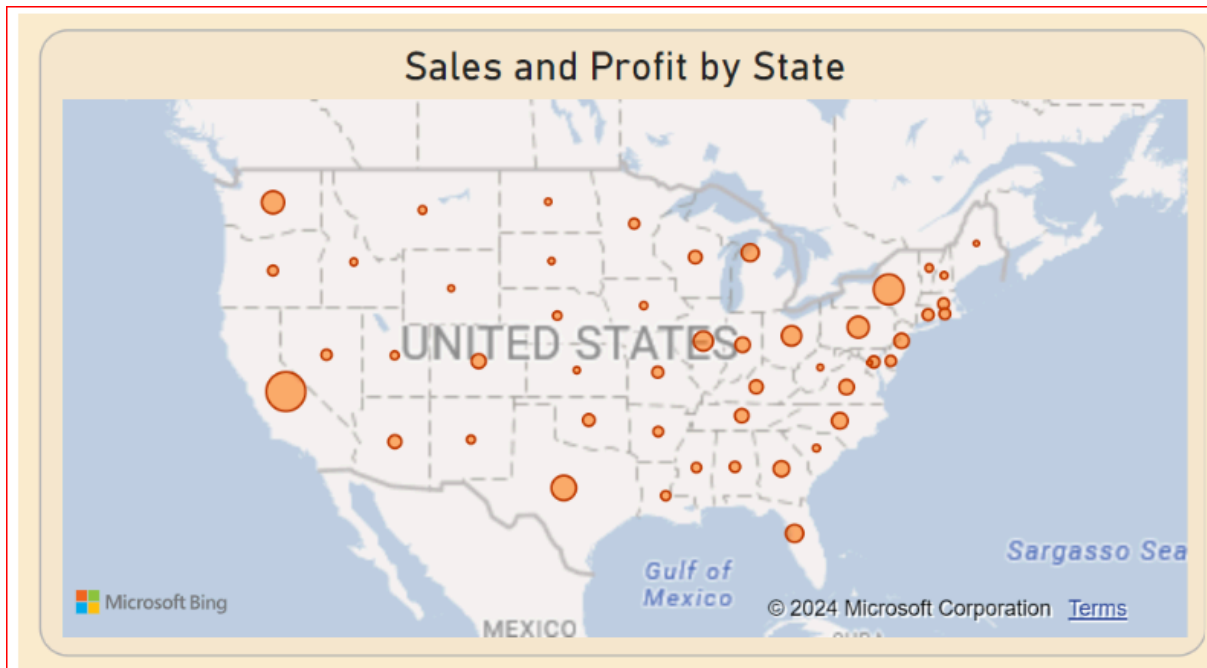
1. **Data Volume Management:**
 - SQL Server handles 10 million daily rows with partitioning and indexing.
 - SSIS pipelines are designed for parallel execution, reducing ETL runtime.
 2. **Improved Reporting Speed:**
 - Pre-aggregated data in SQL Server minimizes query complexity in Power BI.
 - Reports load within 3-5 seconds for 90% of queries, compared to hours before.
 3. **Resilience:**
 - SSIS error-handling mechanisms include retries and logging for failed ETL jobs.
 - SQL Server backups ensure data integrity and disaster recovery capabilities.
-

KPIs Monitored Post-Implementation

1. **Sales Performance:** Achieved 15% monthly revenue growth.
2. **Customer Engagement:** Improved customer retention by 8% through loyalty analysis.
3. **Inventory Efficiency:** Reduced stockouts to less than 1.5% of SKUs.

4. **Campaign ROI:** Increased ROI to 5x by targeting high-value customer segments.
5. **Report Performance:** Reduced report loading time to under 5 seconds for 90% of queries.





Outcomes and Benefits

1. **Streamlined ETL Process:** SSIS efficiently handles 10 million daily rows across SQL Server, Teradata, Salesforce, and CSV files.
2. **Centralized Data Warehouse:** SQL Server consolidates data from disparate sources, ensuring consistency and accessibility.
3. **Optimized Reports:** Power BI provides faster, actionable, and visually appealing insights.
4. **Enhanced Decision-Making:** Unified reporting empowers teams to identify trends, optimize inventory, and improve customer engagement.

This solution resolved the supermarket chain's data integration and reporting issues, delivering a scalable and high-performance analytics platform.