

Data Integration and Cloud Services

Course Project on

Retailer Database Management System

Bachelor of Engineering

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted By

Team No: 09

Priyanshu Kumar	01FE21BCS244	438
Harsh Kashyap	01FE21BCS238	332
Bhavya Shivanand	01FE21BCS351	448
Mita Avadhani	01FE21BCS011	102

Faculty In charge

Prof. Neha Tarannum

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

HUBLI-580031 (India).

Academic year 2023-24

Table of Content	
Chapters	Page No
1. Introduction	3-4
1.1 Preamble	3
1.2 Problem Definition	4
1.3 Objectives	4
2. ER diagram	5
3. Data set description	6-7
4. Transformations (screenshots) with explanation	8-12
5. Conclusion (insights from data – Business Intelligence)	13

1. Introduction

1.1 Preamble

A Retailer Database Management System (DBMS) is a central place where all the important data related to a retail business is stored and managed. This includes information about products, customer details, and sales transactions. By having all this data in one place, retailers can keep track of their operations more easily and ensure everything runs smoothly.

The system also offers powerful tools for analyzing this data and generating reports. Retailers can use these tools to look at sales trends, understand customer behavior, and evaluate their inventory performance. This helps them make informed decisions to improve their business strategies, enhance customer experiences, and increase profits.

Informatica PowerCenter is a robust platform that helps with complex data integration tasks. It connects different data sources, transforms the data into usable formats, and provides insights through detailed analysis. This is especially useful for retail businesses that gather data from various places like store sales, and online transactions.

PowerCenter automates the process of integrating data from different sources into the Retailer DBMS, ensuring that the data is consistent and accurate. It also cleans and standardizes the data, making it more useful for analysis. With its advanced reporting tools, PowerCenter allows retailers to generate reports that highlight trends, identify issues, and suggest improvements.

1.2 Problem Definition

Retail businesses face significant challenges in managing and utilizing the vast amounts of data generated from various sources such as product catalogs, customer interactions, and sales transactions. The primary problem is the lack of a centralized, efficient, and reliable system to store, integrate, and analyze this data to support informed decision-making and optimize business operations.

1.2 Problem Description

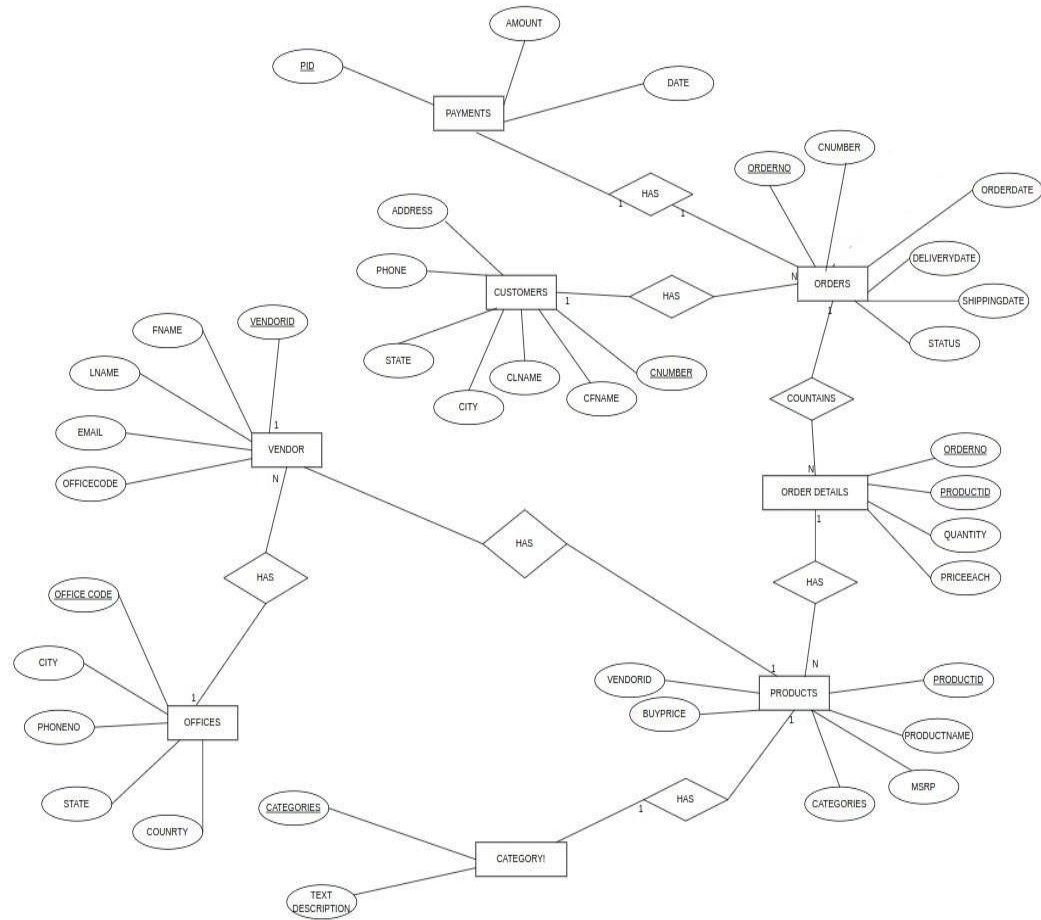
Retailers collect data from many different sources, like product lists, customer information, and sales records. This data is often stored separately, making it hard to get a complete picture of the business. This disorganized data storage prevents effective analysis of sales trends, inventory, and customer behavior, leading to poor decision-making.

Integrating data from these sources is complicated and time-consuming. Different formats and quality issues make it even harder to combine the data reliably. Without a single system to keep data consistent and accurate, it's challenging to generate useful insights.

1.3 Objectives

- Use Informatica PowerCenter to integrate data from various sources, ensuring accuracy and consistency across the database.
- Apply data transformations such as aggregation, router, and expression to standardize and enrich the data, ensuring reliable information for real-time analysis and decision making.
- Consolidate all retail data, such as products, customers, vendors, orders, and payments, into a single, unified database.

2. ER Diagram



3. Data set description

- **Customers:**

Attributes:

- **CNUMBER:** Unique identifier for each customer.
- **CLNAME:** Last name of the customer.
- **CFNAME:** First name of the customer.
- **ADDRESS:** Physical address of the customer.
- **PHONE:** Contact phone number for the customer.
- **STATE:** State in which the customer resides.
- **CITY:** City in which the customer resides. .

Description: Stores details of customers who place orders, including their contact details and location.

- **Vendors:**

Attributes:

- **VENDORID:** Unique identifier for each vendor.
- **FNAME:** First name of the vendor representative.
- **LNAME:** Last name of the vendor representative.
- **EMAIL:** Email address of the vendor representative.
- **OFFICECODE:** Code indicating which office the vendor is associated.

Description: Manages information about vendors from whom products are sourced, including how to contact them and where they are based.

- **Offices:**

Attributes:

- **OFFICECODE :** Unique code for each office.
- **CITY:** City where the office is located.
- **PHONENO:** Contact phone number for the office.
- **STATE:** State where the office is located.
- **COUNTRY:** Country where the office is located.

Description: Contains information about different office locations, which might be relevant for shipping, contact purposes, or vendor management.

- **Orders:**

Attributes:

- **ORDERNO:** Unique identifier for each order.
- **ORDERDATE:** Date when the order was placed.
- **DELIVERYDATE:** Date when the order is scheduled to be delivered.
- **SHIPPINGDATE:** Date when the order was shipped.
- **STATUS:** Current status of the order (e.g., pending, shipped, delivered).

- **CNUMBER:** Reference to the customer who placed the order.

Description: Represents customer orders. Each order is tracked with its date details and status, and it links back to the customer who placed it

Order Details:

Attributes:

- **ORDERNO:** Reference to the order containing the product.
- **PRODUCTID:** Reference to the product being ordered.
- **QUANTITY:** Number of units of the product ordered.
- **PRICEEACH:** Price per unit of the product at the time of the order.

Description: Captures specifics of each individual product within an order, including the product's identifier, quantity ordered, and price per unit.

• Products:

Attributes:

- **PRODUCTID:** Unique identifier for each product.
- **PRODUCTNAME:** Name of the product.
- **BUYPRICE:** Buying price of the product.
- **MSRP:** Manufacturer's Suggested Retail Price.
- **VENDORID:** Reference to the vendor supplying the product.
- **CATEGORY:** Reference to the product category.

Description: Stores information about the products available for purchase, including their names, prices, and categories they belong to.

• Categories:

Attributes:

- **CATEGORY:** Unique name or identifier for the category.
- **TEXT DESCRIPTION:** Description of the category.

Description: Organizes products into different categories for easier catalog management and navigation.

• Payments:

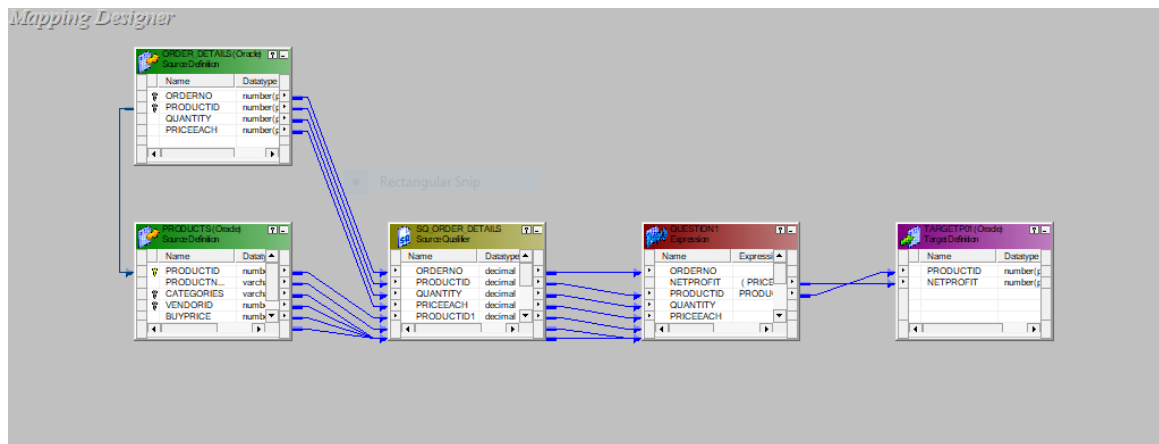
Attributes:

- **PID:** Unique identifier for each payment.
- **AMOUNT:** Amount of the payment.
- **DATE:** Date of the payment.
- **ORDERNO:** Reference to the order the payment is for.

Description: Records details of payment transactions related to customer orders, including the amount paid and the date of the transaction.

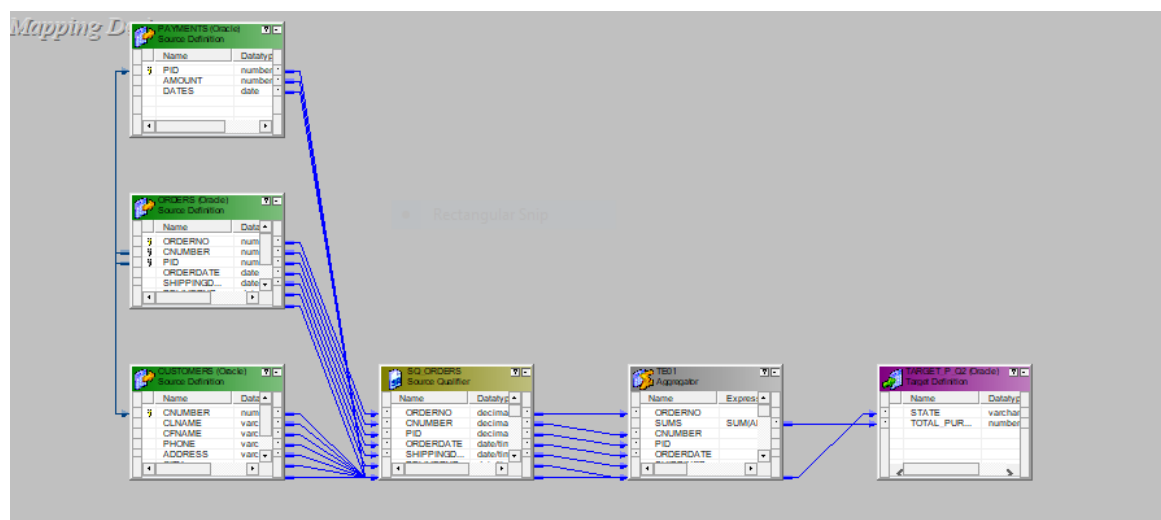
3. Transformations

1. You are a sales manager responsible for analyzing sales performance. You need to determine the profit of each product

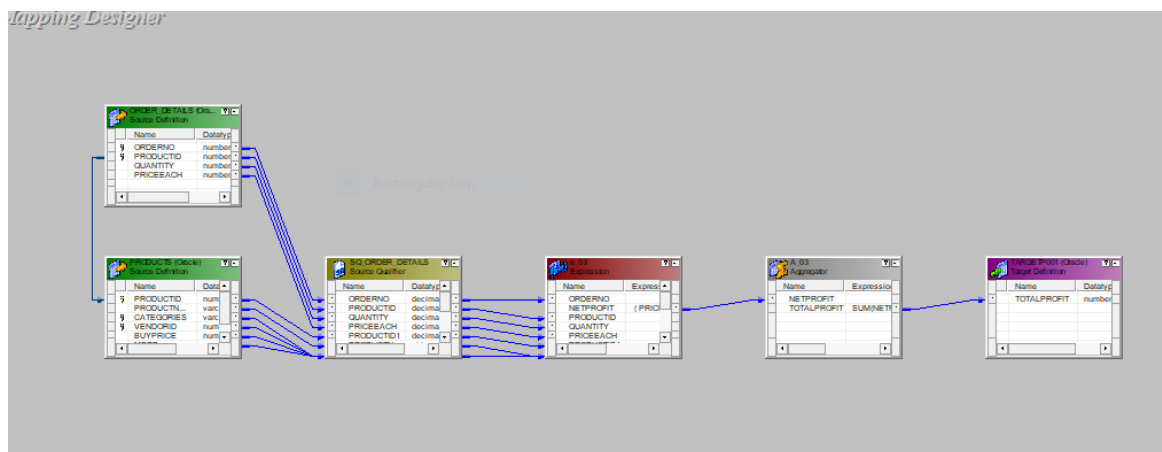


- **Source Qualifier Transformation:** Joins the orderDetails table with Products table to bring together information about each product's orders and their corresponding details.
- **Expression Transformation:** Calculates profit or loss for each product by subtracting PriceEach (selling price) from BuyPrice (cost price) for each order detail.
- **Aggregator Transformation:** Aggregates the revenue generated by each product across all orders to determine the net profit earned from sales of each product.

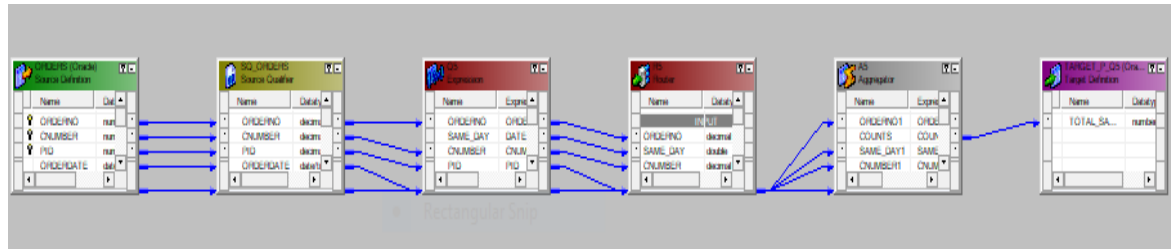
2. You are part of a marketing team aiming to promote various customer segments. Your task involves segmenting customers according to their respective regions.



- **Source Qualifier Transformation:** Connects to Customers, Orders, and Payments tables to gather relevant data about customer details, their orders, and payment transactions.
- **Aggregator Transformation:** Aggregates order data to compute the total purchase amounts for each customer. Groups the total purchase amounts by state to analyze and segment customers based on their geographical regions.

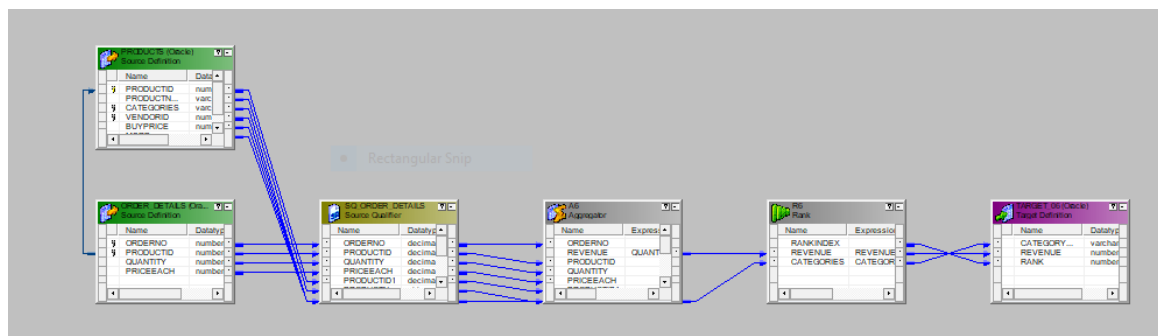


4. The management wants to determine the logistic efficiency by analyzing the orders that are shipped and delivered on the same day.



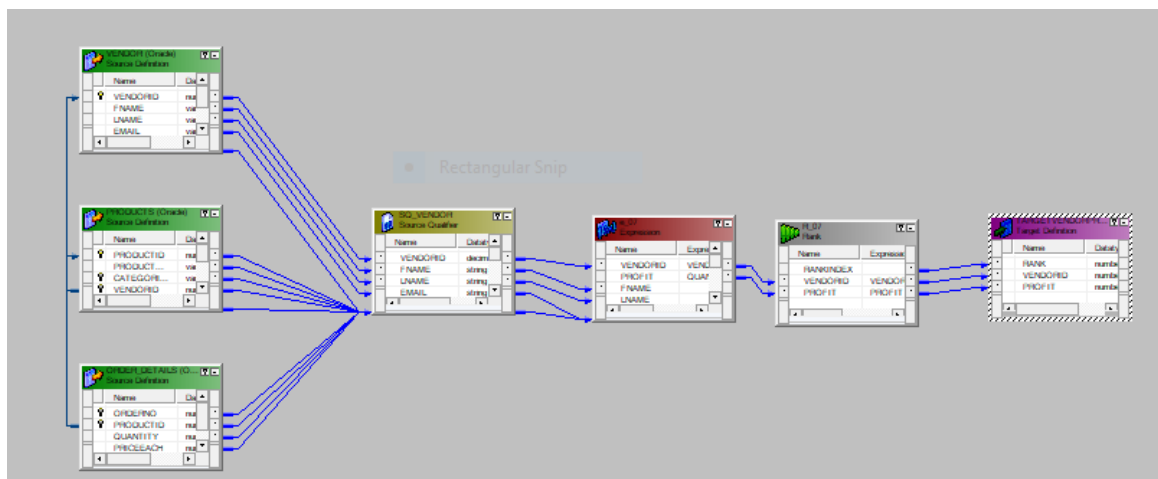
- **Expression Transformation:** Calculates the difference between deliveryDate and requiredDate to determine the time elapsed between when an order was required to be delivered and when it actually was delivered.
- **Router Transformation:** Filters out rows where the calculated difference between deliveryDate and requiredDate equals 0. This identifies orders that were both shipped and delivered on the same day, providing insight into logistic efficiency.

5. You are a product manager who needs to understand the sales distribution across different product:



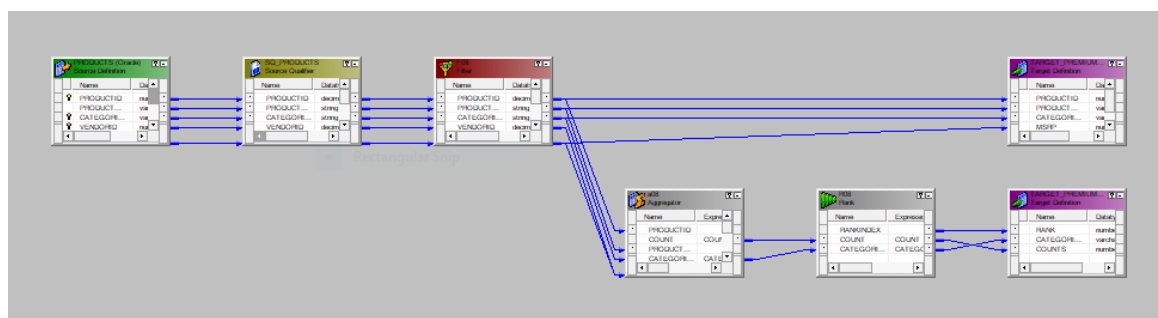
- **Source Definitions (Products and Order_Details):** Define and extract data from the source tables Products and Order_Details.
- The **Aggregator** transformation is used to perform aggregate calculations such as sums or averages. Here, it appears to aggregate the ORDERNO, PRODUCTID, QUANTITY, PRICEEACH, and compute an aggregated quantity and price (possibly total revenue).
- The **Rank** transformation ranks the data based on specified criteria. In this example, it ranks the aggregated data on the basis of revenue within each category.

6. As a financial analyst, you want to determine the profitability of products supplied by each vendor by analyzing your database.



- **Source Qualifier Transformation:** Joins the Vendors, Products, and OrderDetails tables to bring together information about each product's orders and their corresponding details.
- **Expression Transformation:** Calculates profit for each product by subtracting BuyPrice (cost price) from PriceEach (selling price) for each order detail.
- **Aggregator Transformation (Product Level):** Aggregates the revenue generated by each product across all orders to determine the net profit earned from sales of each product.
- **Aggregator Transformation (Vendor Level):** Summarizes the total profit by vendor.

7. As a Marketing manager you need to determine the details of the premium products that have been sold and rank them categorically:

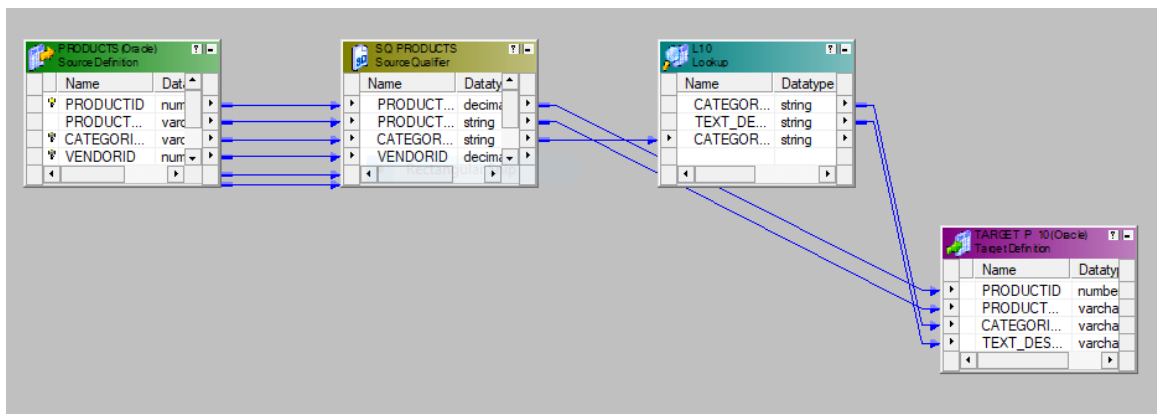


- **Source Qualifier Transformation:** Fetch ProductID, ProductName, Category, and

MSRP from the Products table.

- **Filter Transformation:** Filter out products with MSRP less than 15000.
- **Aggregator Transformation:** Aggregate the count of premium products by category.
- **Rank Transformation:** Rank the categories based on the number of premium products.

8. As a vendor you need to know what text description of different categories is being shown to the customers:



- **Source Qualifier Transformation:** Fetch ProductID, ProductName, and CategoryID from the Products table.
- **Lookup Transformation:** Fetch the CategoryID and CategoryDescription from the Categories table.
- **Join Transformation:** Combine the data from the Products and Categories tables to get the product details along with their category description.

3. Conclusion

Using Informatica PowerCenter in a retailer database offers a robust solution for data integration, ensuring high data quality and real-time processing. This enhances operational efficiency through automation, reducing manual efforts and minimizing errors. PowerCenter scales effectively with growing data volumes, supporting the retailer's expansion needs. It also provides comprehensive analytics, enabling informed decision-making by offering valuable insights into sales, inventory, and customer behavior.

Ultimately, Informatica PowerCenter drives business growth and customer satisfaction by ensuring that data is accurate, accessible, and actionable. It empowers retailers to respond quickly to market changes, optimize supply chain operations, and enhance the overall customer experience, thereby fostering loyalty and boosting revenue.