Real-Time Capstone Project: Retail Sales & Customer Insights Dashboard-Cloud That

Project Overview

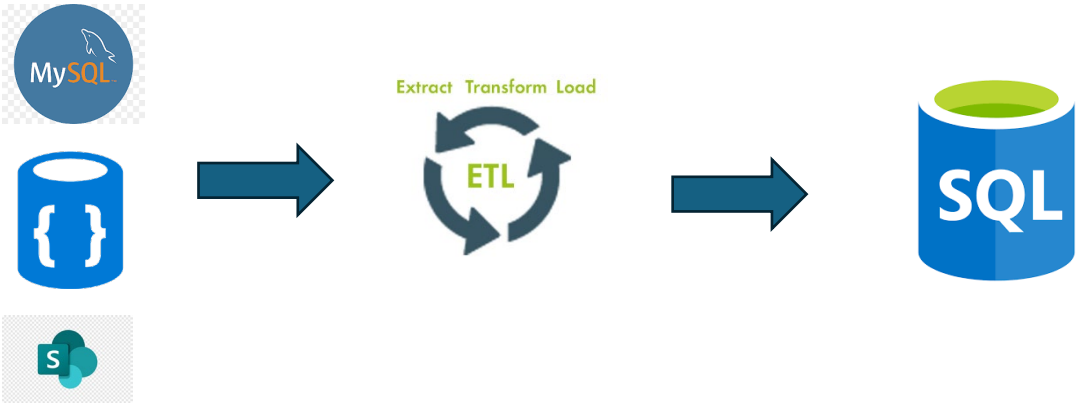
This project aims to address the challenges faced by a retail chain with multiple stores, such as fluctuating sales performance and difficulty understanding customer behaviour. It integrates various data sources into a data warehouse and provides insights through data analytics and Power BI. Sales tracking can be complex without clear insights into best-selling products, seasonal trends, and underperforming items. Understanding customer demographics and purchasing patterns is crucial for personalized marketing and retention but can be difficult without proper segmentation. The project focuses on improving decision-making, understanding complex purchasing patterns, visualizing regional sales trends, and analyzing customer feedback to enhance product offerings and service quality.

Business Problem Statement

A retail chain with multiple stores is experiencing several challenges due to its extensive and varied customer base, wide range of products, and regional distribution. These challenges include:

1. **Identifying Top-Performing Products:**
2. **Understanding Complex Purchasing Patterns:**
    - Seasonality
    - Promotions
    - Customer Demographics
3. **Visualizing Regional Sales Trends:**
    - Inventory Optimization
    - Marketing Strategies

Task Activity:



Phase 1: Project Initiation & Requirement Analysis

Data Understanding: Review data sources (Sales Data, Customer Data, Product Data).

Phase 2: Data Warehouse Design & Development

Schema Design: Create a Star Schema with Fact and Dimension tables.

- **Data Source Identification:**
  - Sales Data (MySQL)
  - Customer Data (NoSQL)
  - Product Data (SharePoint)
- **ETL Pipeline Implementation:**
  - Extract: Collect data from MySQL, NoSQL, and CSV.
  - Transform: Data cleansing, normalization, and formatting.
  - Load: Store transformed data into the Data Warehouse.
- **Staging & Storage:**
  - Staging area for raw data.
  - Fact Tables (Sales Transactions).
  - Dimension Tables (Customer, Product, Region, etc.).

Phase 3: Data Analysis & SQL Implementation

- **Query Development:**
  - Identify best-selling products.
  - Segment customers based on purchase patterns.
  - Analyze regional sales trends.
- **KPI Calculation:**
  - Total Sales Revenue.
  - Sales Growth Rate.
  - Customer Lifetime Value (CLV).
  - Product Return Rate.

Phase 4: Power BI Dashboard Development

- **Data Preparation:**
  - Import data from SQL, NoSQL, and CSV sources.
  - Clean and transform data using Power Query.
- **Data Modeling:**
  - Define relationships between tables.
  - Use **DAX (Data Analysis Expressions)** to create calculated measures.
- **Visualization & Dashboard Creation:**
  - Build visual elements (bar charts, line graphs, tables).
  - Implement slicers and filters for dynamic user interaction.
  - Create drill-through pages for detailed insights.

Data Warehouse:

A **Data Warehouse (DW)** is a centralized repository designed for storing, managing, and analyzing large volumes of structured data from multiple sources. In your **Retail Sales & Customer Insights Dashboard** project, the data warehouse will integrate sales transactions, customer demographics, and

product data to provide insights into sales trends, customer behaviour, and product performance.

1. Data Sources

Schema: Create a star schema for your data warehouse. Here is the Schema for the project

Data Sources

Source Files:

1. **Sales Data (MYSQL)**

Column Name	Data Type	Constraints	Description
SaleID	INT	PRIMARY KEY	Unique identifier for each sale
ProductID	INT	FOREIGN KEY → Product(ProductID)	References the product sold
CustomerID	INT	FOREIGN KEY → Customer(CustomerID)	References the customer who made the purchase
SalesAmount	DECIMAL(10,2)	NOT NULL	Total sales amount for the transaction
Quantity	INT	NOT NULL	Number of units sold
Timestamp	DATETIME	NOT NULL	Date and time of the sale

2. **Customer Data (NoSQL)**

Column Name	Data Type	Constraints	Description
CustomerID	INT	PRIMARY KEY	Unique identifier for each customer
FirstName	VARCHAR(50)	NOT NULL	Customer's first name
LastName	VARCHAR(50)	NOT NULL	Customer's last name
Gender	VARCHAR(10)	NULLABLE	Customer's gender
Region	VARCHAR(50)	NULLABLE	Customer's geographical region
SSN	VARCHAR(11)	UNIQUE, NOT NULL	Customer's Social Security Number

3. **Product Data (SharePoint)**

Column Name	Data Type	Constraints	Description
ProductID	INT	PRIMARY KEY	Unique identifier for each product
ProductName	VARCHAR(100)	NOT NULL	Name of the product
Category	VARCHAR(50)	NOT NULL	Category of the product

2. Data Integration and ETL (Extract, Transform, Load) Process

- **Data Extraction**: Collect data from various sources.
- **Data Transformation**: Cleanse, normalize, and transform data to ensure consistency and accuracy.
- **Data Loading**: Load transformed data into the data warehouse.

3. Data Storage

- **Staging Area**: Temporary storage for raw data before transformation.
- **Data Warehouse**: Centralized repository for structured and transformed data.
    - **Fact Tables**: Store quantitative data for analysis.
    - **Dimension Tables**: Store descriptive attributes related to facts (e.g., customer demographics, product categories, store locations).

SQL:

Leverage SQL to analyze sales data for a retail chain, to identify best-selling products, segment customers based on their purchasing behaviour, and visualizing regional sales trends. By utilizing SQL, we will gain actionable insights to drive better decision-making, optimize marketing strategies, and improve overall business performance.

Key Metrics and KPIs

1. Total Sales Revenue
2. Sales Growth Rate
3. Top-Selling Products
4. Customer Lifetime Value (CLV)
5. Regional Sales Performance

Python:

Do initial data cleaning, transformation, and data quality check.

Power BI:

Developing a Power BI dashboard involves several key steps to ensure it effectively presents data insights and meets the business's requirements.

Data Preparation

- **Data Integration:** Import data from various sources like SQL databases, and cloud services into Power BI.
- **Data Cleaning:** Ensure data quality by removing duplicates, handling missing values, and correcting errors.
- **Data Transformation:** Use Power Query Editor to reshape and transform data into a suitable format for analysis.

Create a Data Model

- **Design Data Model:** Establish relationships between tables and create a logical data model.
- **Define Measures:** Use DAX (Data Analysis Expressions) to create calculated measures and columns for advanced analysis.
- **Optimize Performance:** Ensure the data model is optimized for performance by using appropriate indexing and avoiding complex calculations.

Build Visualizations

- **Select Visualizations:** Choose appropriate visualizations (e.g., bar charts, line charts, pie charts, tables) to represent different metrics and KPIs.
- **Add Visuals:** Drag and drop visuals onto the report canvas and configure their properties.
- **Use Slicers and Filters:** Add slicers and filters to enable users to interact with the data and view specific segments.
- **Create Drill-Through Pages:** Allow users to drill down into more detailed data by creating drill-through pages.

Create KPIs:

Preparing different KPIs based on stakeholders' requirement and here are some KPIs that frequently used in Retail Sales & Customer Insights

Sales Performance

- Total Sales Revenue
- Sales Growth Rate- Monthly, quarterly, yearly.
- Average Transaction Value (ATV)
- Sales by Product Category
- Sales by Region.

Customer Insights

- Customer Lifetime Value (CLV)
- Customer Demographics Analysis.

Product Performance

- Top-Selling Products
- Product Return Rate

Operational Efficiency

- Cost of Goods Sold (COGS)