St. Francis Institute of Technology, Mumbai-400 103

**Department Of Information Technology**

A.Y. 2024-2025

Class: TE-ITA/B, Semester: VI

Subject: **Business Intelligence Lab**

**Experiment 1**

**Problem Definition for a Data Warehouse, and Construction of Star/ Snowflake Schema**

1. **Aim: To construct/formulate problem Definition for a Data Warehouse, and Construction of Star Schema/ Snowflake schema**
2. **Objectives:** After study of this experiment, the student will be able to
   * Understand different types of Dimensional modelling concepts
3. **Outcomes:** After study of this experiment, the student will be able to

**CO1: Demonstrate an understanding of the importance of data warehousing and data mining and the principles of business intelligence**

1. **Prerequisite:** Introduction to Databases, ER modeling
2. **Requirements:** Personal Computer, Windows XP operating system, Internet Connection,

Microsoft Word

1. **Theory:**

Include details about:

1. **Why there is a need of DW for any system?**

**ANS:** A Data Warehouse (DW) is needed in this system because it allows for the consolidation of large amounts of diverse data from multiple sources into a single, organized repository. In the case of an e-commerce platform, the data comes from various areas such as customer behavior, product sales, and demographic information. A DW allows this data to be stored in a structured and efficient way, enabling quick access and analysis. This is crucial for identifying trends, patterns, and making informed business decisions, such as optimizing inventory, personalizing marketing efforts, and predicting customer behavior.

**2. What criteria/facts need to be measured?**

**ANS:** Several key metrics and data points need to be measured to gain meaningful insights from the e-commerce platform:

1. Customer Demographics: Age, gender, location, income level, etc., to understand the composition of the customer base.
2. Purchasing Patterns: Frequency of purchases, purchase value, time between purchases, etc., to determine customer loyalty and behavior.
3. Product Preferences: Which product categories or specific products are most popular among different customer segments.
4. Sales Performance: Revenue, profit margin, sales growth, conversion rate, etc., to gauge overall business health.
5. Customer Behavior: Interaction with the platform, abandoned carts, browsing patterns, and time spent on the site.
6. Seasonal Sales Trends: Identifying high-demand seasons, holidays, and sales events to optimize inventory and marketing efforts.
7. Churn Rate: Percentage of customers who stop buying, which is crucial for retention strategies.
8. Customer Lifetime Value (CLV): Total value a customer brings over their entire engagement with the platform.

**3. What goals need to be set?**

**ANS**: The goals of implementing a Data Warehouse for this system should include:

1. Maximizing Revenue: Targeting profitable customer segments and optimizing pricing and promotions to boost sales.
2. Improving Customer Retention: Identifying ways to keep customers coming back, such as personalized offers or loyalty programs.
3. Optimizing Inventory: Ensuring the right products are available at the right time to meet customer demand without overstocking.
4. Enhancing Customer Experience: Personalizing the shopping experience, from product recommendations to tailored content, improving user satisfaction.
5. Identifying Lucrative Customer Segments: Using data to find high-value customers and offering them tailored marketing campaigns.
6. Predicting Future Trends: Using predictive analytics to forecast customer demand, trends, and behavior, helping in decision-making.

**4. What problems need to be solved?**

**ANS:**Several challenges exist that need to be addressed:

1. Data Silos and Inaccessibility: Data may be scattered across different systems (e.g., sales, customer service, website analytics), making it difficult to analyze effectively. A DW solves this by centralizing all the data in one place.
2. Inaccurate Demand Forecasting: Without the right tools, forecasting future demand for products can be unreliable, leading to either stockouts or excess inventory. A DW allows for better demand forecasting through historical data analysis.
3. Personalization: E-commerce platforms need to provide personalized experiences for customers, which can be challenging without integrated customer behavior data. A DW provides a clear view of customer preferences and purchasing history, enabling personalization.
4. Inefficient Marketing Campaigns: Without clear insights into customer segments and behaviors, marketing efforts may be less effective. A DW helps identify which customers are most likely to respond to specific offers or promotions.
5. Customer Churn: Predicting which customers may stop purchasing can be difficult. A DW supports predictive modeling to identify potential churn risks and take preemptive actions to retain these customers
6. **Laboratory Exercise:**

Based on the Case study –

* Explain the system in your own words (problem statement)
* Design / Architecture

§ Dimension table with all dimensions

§ Fact table with Facts/ measures

* Construct a star schema
* Construct snowflake schema

Attach Printout of case study along with output

**8. Post Experiment Exercise:**

1. **Exercise (Handwritten)**
   * What is fact less fact table?
   * Give differences between star and snowflake schema
2. **Conclusion:**

1. Summary of Experiment

2. Importance of Experiment

3. Application of Experiment

1. **Reference:** Data Mining: Concept & Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei, Elsevie

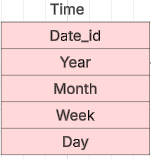
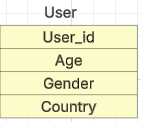
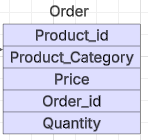
Explain the system in your own words (problem statement)

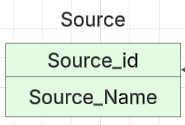
**Problem Statement:**

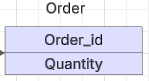
With the rise of e-commerce platforms, understanding customer behavior and optimizing sales strategies are critical to success. This dataset provides synthetic but realistic data about customer demographics, purchasing patterns, and product preferences on an eCommerce clothing platform. The task is to analyze this data to identify key trends, such as the most lucrative customer segments, high-performing product categories, and seasonal sales patterns. Such insights can help design targeted marketing campaigns, improve inventory management, and enhance customer experience.

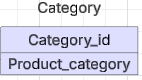
Additionally, the data offers an opportunity to predict future customer behavior based on historical patterns. By building predictive models, the platform can recommend products, anticipate inventory needs, and identify potential churn risks. The ultimate goal is to leverage these insights to improve operational efficiency, maximize revenue, and create a personalized shopping experience for customers.

**Design / Architecture:**

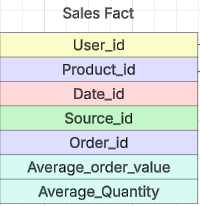
1. Dimension table with all dimensions

****

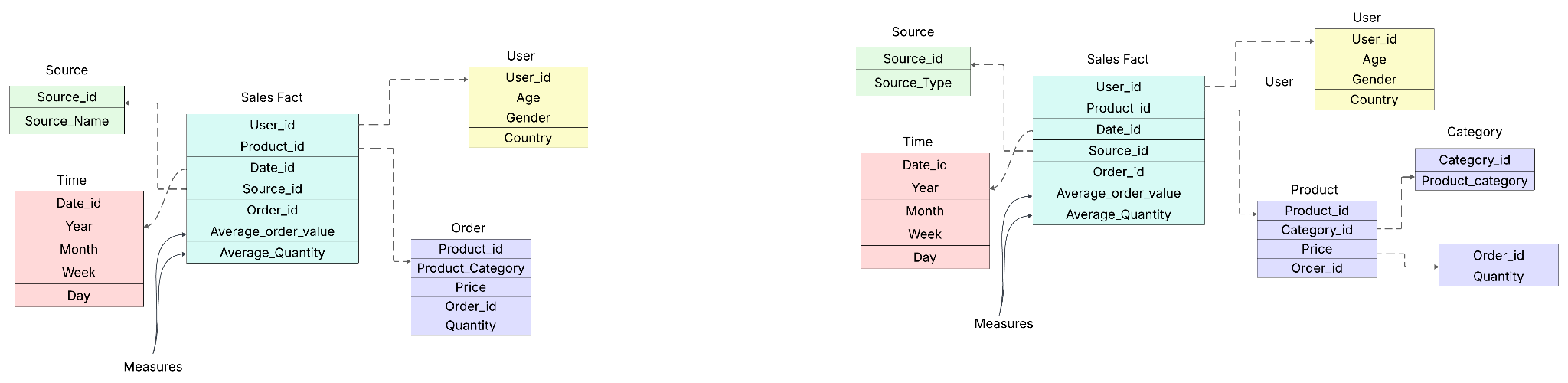


****

1. Fact table with Facts/ measures



**Construct a star schema:**



**Construct snowflake schema:**