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

4. **Prerequisite:** Introduction to Databases, ER modeling

5. **Requirements:** Personal Computer, Windows XP operating system, Internet Connection, Microsoft Word

6. **Theory:**
   Include details about:
   - Why there is a need of DW for this system?
     - Airlines collect a vast amount of data, but it remains scattered without a centralized system.
     - A Data Warehouse helps organize and integrate this data for easy analysis.
     - It enables airlines to identify the main causes of delays and cancellations.
     - Improves decision-making by analyzing past trends and patterns.
     - Enhances customer satisfaction by reducing delays and improving efficiency.
     - Supports real-time monitoring and better resource allocation.
     - Provides actionable insights through data visualization tools like Power BI.
   - What criteria/facts need to be measured?
     - Flight Details: Flight ID, flight number, departure and arrival airports.
     - Delays: Reasons for delays (e.g., weather, technical issues) and delay duration.

- ■ Cancellations: Reasons flights get canceled (e.g., insufficient demand, security issues).
- ■ Timing Factors: Scheduled vs. actual departure/arrival times, day of the week, month, and year.
- ■ Airline Performance: Number of delayed and canceled flights, average delay time.
  - ○ What goals need to be set?
    - ■ Reduce flight delays by identifying the most common causes.
    - ■ Improve operational efficiency by analyzing performance data.
    - ■ Enhance customer satisfaction by minimizing disruptions.
    - ■ Optimize resource allocation to reduce unnecessary costs.
    - ■ Use data-driven insights to improve flight scheduling and management.
  - ○ What problems need to be solved?
    - ■ Lack of a centralized system for analyzing flight delays and cancellations.
    - ■ Difficulty in spotting patterns behind frequent delays.
    - ■ Inefficient airline operations leading to increased costs.
    - ■ Poor customer experience due to unreliable schedules.
    - ■ Limited use of historical data for making informed decisions.

7. **Laboratory Exercise:**
   Based on the Case study –
- ● Explain the system in your own words (problem statement)
  Flight delays and cancellations will always be such a challenge because they lead to customer dissatisfaction and sometimes increased operational cost. Despite gathering a lot of data on airline operations, flights are still difficult to identify to pinpoint the actual causes of flights that are easily delayed or even canceled. If airlines do not have insight into problems, they cannot strategically work on some of these flaws to improve performance.
  To address this issue, the aim is to study a comprehensive dataset that includes delay reasons, cancellation reasons, and flight performance metrics. Using Power BI for data visualization, we want to identify the patterns of disruptions in flights at different airlines and airports. In the end, we seek actionable recommendations to help airlines make improvements in operational efficiency, reduce delay times, and enhance customer satisfaction.
- ● Design / Architecture
  - ○ Dimension table with all dimensions
    1. **Dim_Airline**
       - Airline ID (Primary Key)
       - Airline Name
       - Contact Information

    2. **Dim_Flight**
       - Flight_ID (Primary Key)

- Flight Number
- Departure Airport
- Arrival Airport
- Scheduled Departure Time
- Scheduled Arrival Time

3. **Dim_Date**
   - Date_ID (Primary Key)
   - Date
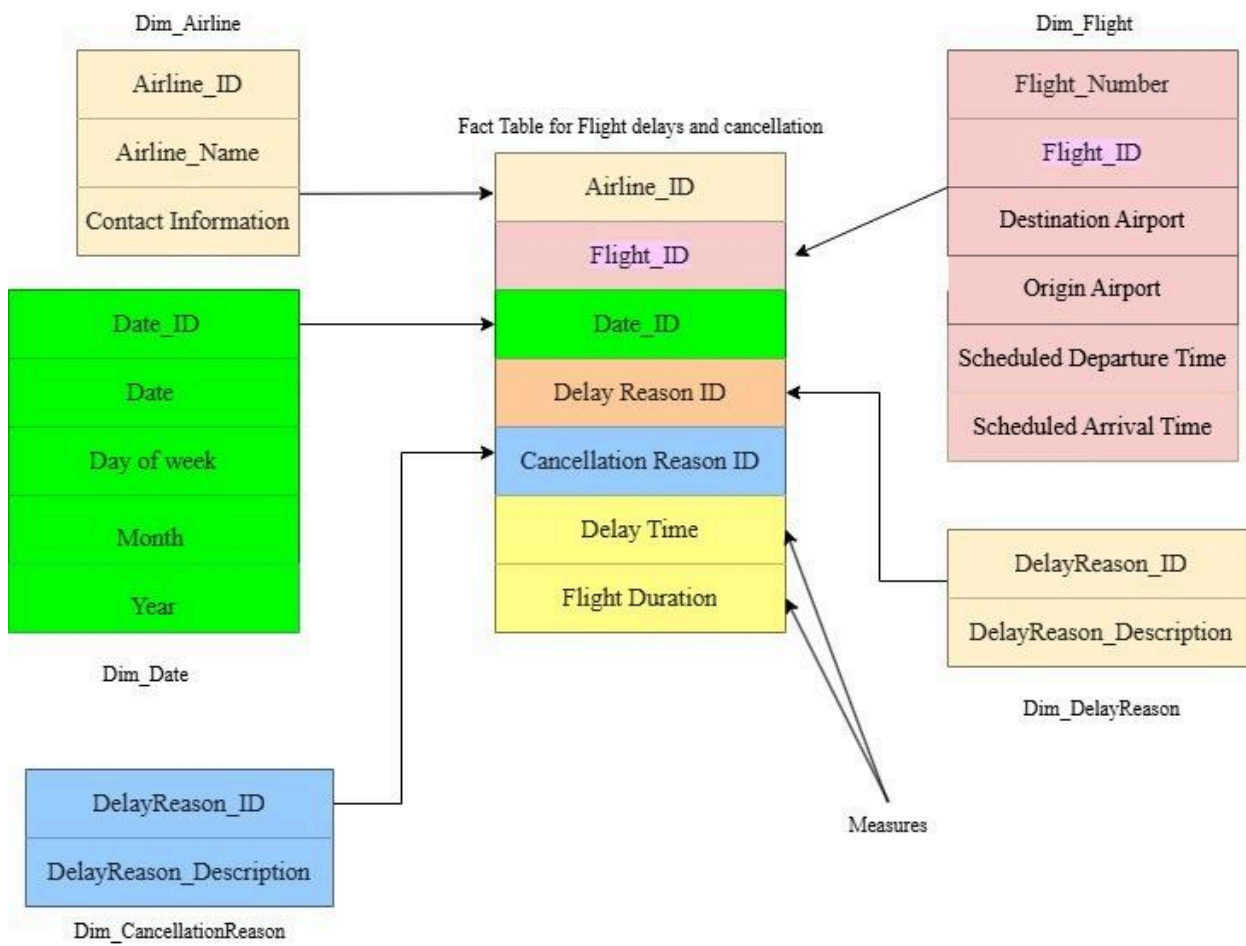   - Day of Week
   - Month
   - Year

4. **Dim_DelayReason**
   - Delay Reason ID (Primary Key)
   - Delay Reason Description (e.g., Weather, Mechanical Issues)

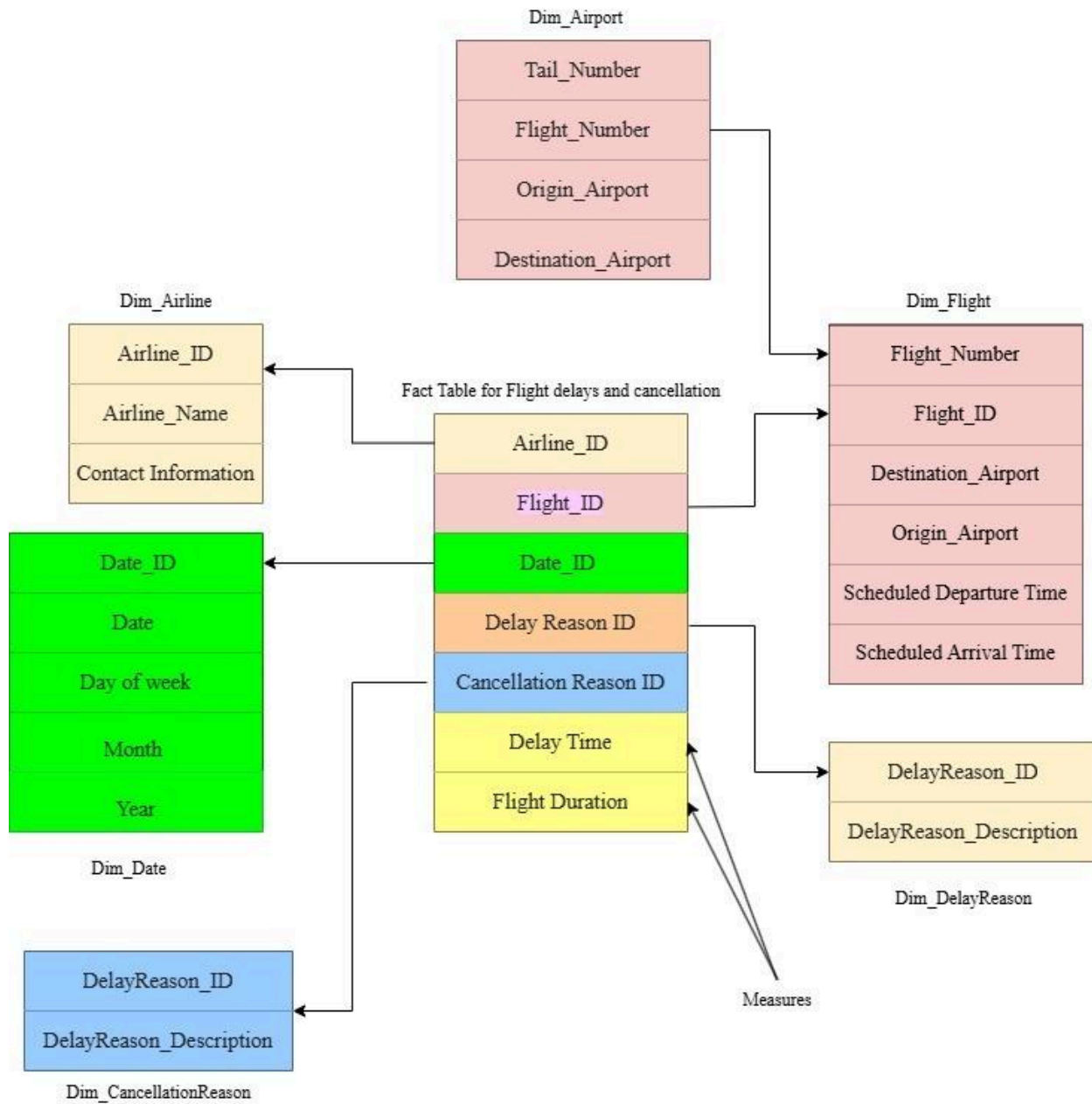5. **Dim_CancellationReason**
   - Cancellation Reason ID (Primary Key)
   - Cancellation Reason Description (e.g., Weather, Insufficient Demand)

○ Fact table with Facts/ measures
   ■ Airline ID (Foreign Key)
   ■ Flight ID (Foreign Key)
   ■ Date ID (Primary Key)
   ■ Delay Reason ID (Primary Key)
   ■ Cancellation Reason ID (Primary Key)
   ■ Delay Time
   ■ Flight Duration

● Construct a star schema

● Construct snowflake schema

Dim_Airport

Tail_Number
Flight_Number
Origin_Airport
Destination_Airport

Dim_Airline

Airline_ID
Airline_Name
Contact Information

Dim_Date

Date_ID
Date
Day of week
Month
Year

Fact Table for Flight delays and cancellation

Airline_ID
Flight_ID
Date_ID
Delay Reason ID
Cancellation Reason ID
Delay Time
Flight Duration

Dim_Flight

Flight_Number
Flight_ID
Destination_Airport
Origin_Airport
Scheduled Departure Time
Scheduled Arrival Time

Dim_DelayReason

DelayReason_ID
DelayReason_Description

Dim_CancellationReason

DelayReason_ID
DelayReason_Description

Measures

8. Post Experiment Exercise:

   A) Exercise (Handwritten)

- What is fact less fact table?
- Give differences between star and snowflake schema

   B) Conclusion:

1. Summary of Experiment
2. Importance of Experiment
3. Application of Experiment

   C) **Reference:** Data Mining: Concept & Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei, Elsevier.