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:

* + Why is there a need for DW for any system?

A Data Warehouse (DW) is needed to:

* **Centralize Data**: Combine data from multiple sources into one place.
* **Enable Better Analysis**: Organize data for efficient querying and reporting.
* **Retain Historical Data**: Store past data for trend analysis and forecasting.
* **Improve Decision-Making**: Provide cleaned, consistent data for insights.
* **Optimize Performance**: Separate analytics from operational systems.
* **Support BI Tools**: Feed structured data to reporting and visualization tools.
* **Ensure Scalability**: Handle growing data needs.
  + What criteria/facts need to be measured?

To measure a system's effectiveness, focus on:

* **Performance**: Response time, speed, uptime, scalability.
* **Data Quality**: Accuracy, consistency, completeness, timeliness.
* **Business Impact**: Revenue growth, ROI, cost efficiency, trends.
* **User Behavior**: Interaction patterns, popular features.
  + What goals need to be set?

Goals to be set for a system include:

* **Performance Goals**: Ensure fast response time, high uptime, and scalability.
* **Data Goals**: Maintain accuracy, consistency, and timeliness of data.
* **User Goals**: Enhance user satisfaction, engagement, and ease of use.
* **Business Goals**: Increase revenue, reduce costs, and achieve ROI targets.
* **Security Goals**: Protect data with robust authentication and compliance measures.
* **Scalability Goals**: Support growing user and data demands efficiently.
* **Analytical Goals**: Enable actionable insights through advanced reporting tools.
  + What problems need to be solved?

The key problems that need to be solved for an effective system include:

1. **Data Management**: Handle large volumes of data, eliminate redundancy, and ensure consistency.
2. **Performance Issues**: Reduce response time, optimize queries, and improve system reliability.
3. **User Experience**: Simplify navigation, improve usability, and meet user expectations.
4. **Integration Challenges**: Seamlessly combine data from multiple sources or systems.
5. **Security Risks**: Protect against unauthorized access, data breaches, and compliance violations.
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 factless 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, Elsevier.

**Laboratory Exercise:**

Based on the Case study –

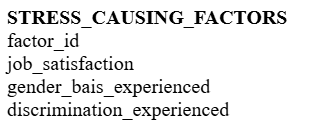
* Explain the system in your own words (problem statement)

Corporate stress is an escalating concern in today’s high-pressure work environments, where relentless deadlines and overwhelming expectations leave employees grappling with exhaustion and burnout. This chronic stress doesn't just erode individual well-being—it saps productivity, dampens creativity, and undermines organizational success. Yet, many organizations remain ill-equipped to recognize and address these invisible struggles. The availability of a corporate stress dataset offers a glimmer of hope, a chance to uncover the hidden patterns behind workplace stress and empower organizations to foster healthier, more compassionate work environments.

By harnessing this dataset with advanced analytics and machine learning, we can illuminate the connections between stress and factors like excessive work hours, insufficient managerial support, and inequitable policies. These insights are not just numbers—they are a call to action. They can guide organizations in crafting tailored strategies, from personalized stress management programs to equitable workload distribution. The ultimate goal is to replace burnout with balance, turning workplaces into spaces where individuals thrive and collective success is inevitable.

* Design / Architecture

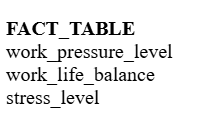
1. Dimension table with all dimensions and Fact table with all measures.

**TIME** 

time\_key

year

quarter

month

**EMPLOYEE**

employee\_id

age

gender

martial\_status

job\_role

sleeping\_hours

health\_issues

**DEPARTMENT**

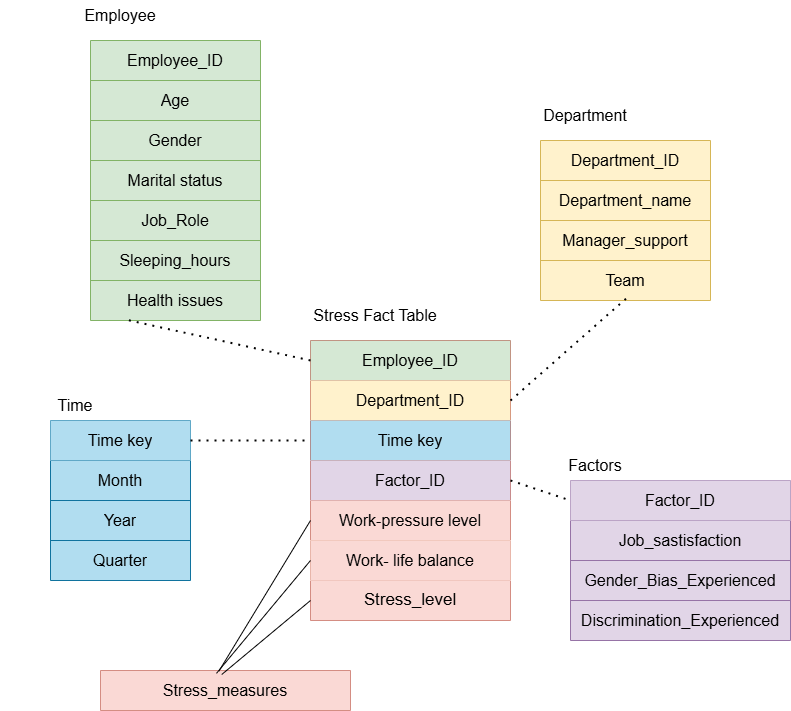
department\_id

department\_name

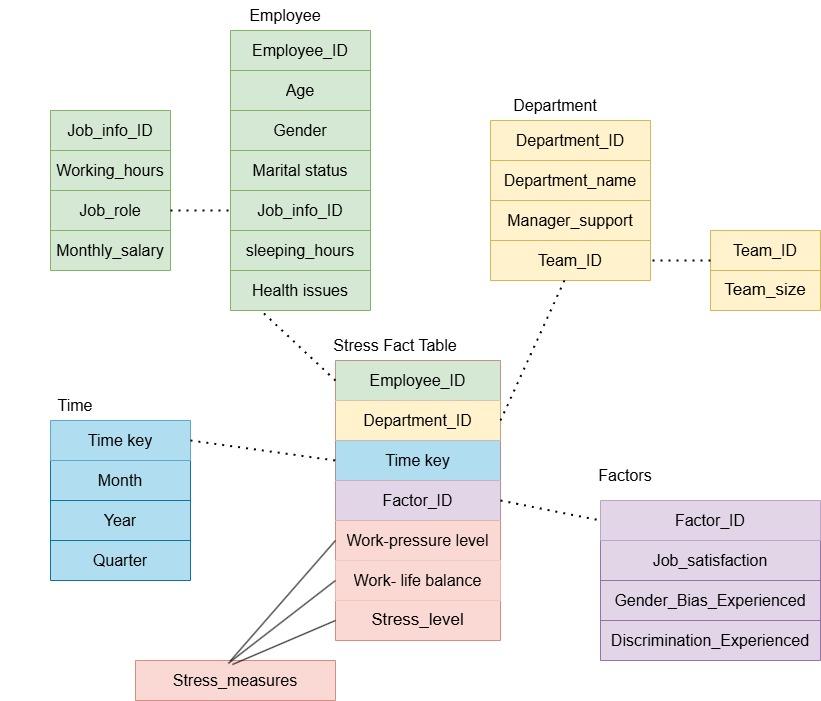
manager\_support

team

**Star Schema**

****

**SnowFlake Schema**

****