VIJAYBHOOMI UNIVERSITY

INDIA'S FIRST LIBERAL PROFESSIONAL UNIVERSITY

Name of the School: VSST

MID-TERM EXAMINATIONS NOV 2025

Course Name: Hacthon-3

Course Code: STSE203 **Max. Marks: 20**

Mid-Term Hackathon (20 Marks)

🎯 Title: *Data Pipeline Design & Implementation (ETL Stage)*

🧩 Problem Statement:

The city government wants to analyze public transport usage patterns and traffic congestion trends. As a data engineering analyst, your task is to:

1. Collect and clean data from **two sources**:
    - CSV: *Public Transport Usage Data* (e.g., bus routes, ridership, timings)
    - JSON/API: *Traffic Sensor Data* (e.g., average speed, congestion index)
2. Build a **data pipeline** that:
    - Extracts, transforms, and loads (ETL) data into a unified table in a database (e.g., MySQL/PostgreSQL).
    - Includes **data validation** (missing values, data types, timestamp checks).
    - Creates a **summary CSV** with average ridership and congestion index per route.

⚙️ Technical Requirements:

- Use Python or shell scripting for ETL.
- Schedule pipeline using **Cron or Airflow** (if available).
- Store processed data in a relational DB.
- Use awk, sed, or grep for quick data checks before ingestion.

🧾 Deliverables:

1. Folder structure of the project (/raw, /processed, /scripts, /reports).
2. Clean and processed dataset.
3. A **readme/report** explaining:
    - Pipeline design

- Tools used
- Transformation steps

---------- **END OF TEST** ----------