Mahmoud Alsayed Gaballah(dagabz234@gmail.com)

1. Project Description

Design and implement a Real-Time IoT Data Pipeline to simulate and process sensor data (temperature and humidity) using both batch and streaming techniques. The system will detect anomalies and display metrics in a real-time dashboard.

2. Group Members & Roles

Member Name	Role
Mahmoud Elsayed Gaballah	Team Leader / Data Engineer
Malak Hosam Soliman	IoT Data Simulation Developer
Mennatullah Waleed Hassanen	Batch ETL Engineer
Salma Abd Elaziz Daby	Streaming & Alerting Engineer
Ganna Salah Ahmady	Database & Storage Specialist
Mennatullah Mohammed Naam	Visualization & Reporting Analyst

3. Team Leader

Mahmoud Elsayed Gaballah

4. Objectives

- Build an IoT system that simulates real-time sensor data flow.
- Implement batch and streaming processing for IoT data.
- Detect anomalies and generate alerts.
- Visualize key metrics in a live dashboard.

5. Tools & Technologies

Phase	Tools / Technologies
Data Simulation	Python / Kafka Streaming
Batch ETL	PySpark / Azure Data Factory
Streaming	Kafka / Spark Structured Streaming
Storage	Azure Data Lake / SQL Database
Dashboard	Power BI / Streamlit

Reporting	PowerPoint / PDF

6. Milestones & Deadlines

Milestone	Description	Deadline
Milestone 1: Data Simulation	Python script + sensor data	31/9/2025
Milestone 2: Batch ETL	ETL pipeline + processed dataset	11/10/2025
Milestone 3: Streaming Analytics	Real-time alerts pipeline	1/11/2025
Milestone 4: Dashboard & Report	Visualization + final report	22/11/2025

7. KPIs (Key Performance Indicators)

1. Data Preprocessing
    - 100% of missing/duplicate data correctly handled.
    - Script efficiency: execution time within expected threshold.
2. SQL Integration
    - Query accuracy: ≥95%.
    - Query performance: average execution time under 3 seconds.
3. Visualization
    - Dashboard load time: <3 seconds.
    - ≥90% of required KPIs visualized.
4. Presentation
    - Report completeness: 100%.
    - Stakeholder clarity/feedback score: ≥4/5.