Part 1 – Project Planning

Project Title: Smart Radar Traffic Monitoring System

Project Overview

The *Smart Radar Traffic Monitoring System* is designed to simulate radar sensors deployed across a city to monitor and analyze vehicle traffic. Each radar generates live data such as vehicle license plate, speed, color, location, seat belt usage, and phone usage.
The project leverages data engineering tools and cloud technologies to process this information in both **batch** and **streaming** modes for real-time insights and historical analysis.

Objectives

1. Develop a realistic vehicle data simulator using Python for real-time data ingestion.
2. Build a scalable **batch processing pipeline** with Apache Spark and Azure Data Lake for historical data analysis.
3. Implement a **streaming pipeline** for real-time violation detection and alerts.
4. Create an **interactive Power BI dashboard** for visualization and decision-making.
5. Deliver complete system documentation detailing architecture, data flow, and implementation.

Team Members and Roles

Member Name	Assigned Role	Responsibilities
Hala Farouk	**& Team Leader Lead Data Engineer**	Manage project progress, oversee pipeline design, ensure integration between modules
Marev Wasim	**Data Engineer & Ingestion) (Simulation**	Responsible for Milestone 1: Building and running a data simulation script and connecting it to Kafka
Mohamed Farrag	**Data Engineer (Batch Processing)**	Responsible for Milestone 2: Building Batch Pipeline using PySpark and managing Data Lake

Member Name	Assigned Role	Responsibilities
Mera Sameh	**Data Engineer Streaming) (Processing**	Responsible for Milestone 3: Building Streaming Pipeline and developing instant alerts system
Yossef Ahmed	**Cloud & DevOps Engineer**	Responsible for setting up and managing cloud resources (Azure), Docker setup, and database management
David Bahaa	**& BI Developer Data Analyst**	Responsible for Milestone 4: Power BI Dashboard Development and Data Analysis

Team Leader: Hala Farouk Mohamed Mutawa

Tools & Technologies

Category	Technology
Data Ingestion & Streaming	Python (Faker), Azure Event Hubs
Data Processing	Apache Spark (PySpark)
Data Storage	Azure Blob Storage (Data Lake -JSON), (Data Warehouse)
Orchestration	Azure Data Factory
Visualization & Dashboard	Power BI
Containerization	Docker

Milestones & Deadlines

Milestone	Key Deliverables	**Estimated Deadline (New)**
M1: Data Simulation & Ingestion	Python generator script	**October 19, 2025** (No Change)
M2: Batch ELT Pipeline	PySpark script, Data Lake (JSON), SQL DWH schema.	**November 2, 2025** (No Change)
	--- University Exam Period (No Deliverables) ---	**November 7 - November 14, 2025**
M3: Streaming Pipeline & Alerts	Real-time processing code, alert system integration (Slack/Email).	**November 23, 2025** (Postponed)
M4: Dashboard & Final Report	Interactive Power BI dashboard, final PDF report with diagrams.	**December 3, 2025** (Postponed)
Final Project Submission	All code, documentation, and final presentation slides.	**December 5, 2025** (Postponed)

Key Performance Indicators (KPIs)

1. Data Processing & Pipeline Performance

Data Quality: 100% of records with null plate_number or negative speed are correctly filtered or cleaned during processing. (Target: 100%)

Batch Job Efficiency: The daily batch ELT job (PySpark) completes processing for 1 million records within a 15-minute threshold.

Stream Latency: End-to-end latency from data generation to alert trigger for critical events is under 5 seconds.

2. SQL & Data Warehouse Integration

Query Accuracy: Aggregation queries for the dashboard (e.g., total violations per day, top speeding locations) match manually verified control values. (Target: 100%)

Query Performance: The average execution time for queries powering the dashboard visuals is under 2 seconds.

3. Visualization

Dashboard Load Time: The initial load time for the main dashboard page is less than 5 seconds.

(Target: < 5 sec)

Data Representation: 100% of required project metrics (Violation Trends, Real-time Map, Heatmap) are successfully and accurately visualized on the dashboard.

(Target: 100%)

4. Presentation & Documentation

Report Completeness: The final report includes all required sections, featuring a detailed System Architecture Diagram and clear explanations of the data flow.

(Target: 100%)

Stakeholder Clarity/Feedback Score: Achieve a clarity and satisfaction score of ≥ 4/5 from the project evaluators based on the final presentation.