eyouth X DEPI Project Proposal

Team: Data Forge

Team Members:

- Abdallah Ramadan Abdallah Ali
- Ahmed Yehya Saad Nafea
- Rania Ossama Hassan Abd Elhaleem
- Merehan Ibrahim El Moatasem Hassan
- Hagar Mohamed Mabrouk Mohamed
- Youssef Wael Omar Abu Dief

Project Description

Project Title: Real-time Transit and Traffic Management System

This project aims to address the persistent challenges of urban traffic congestion and unreliable public transit schedules. By leveraging real-time data from various transportation sources, we will design and implement a **fully automated and scalable data engineering pipeline**. The solution will ingest, process, and analyze traffic flow data. The ultimate goal is to provide a solid foundation for advanced analytics and empowering users to derive meaningful insights and improve decision-making within the transportation domain.

Group Members & Roles

To ensure a balanced distribution of work and clear ownership, the following roles have been assigned to the team members:

- **Abdallah Ramadan Abdallah Ali: Team Leader & Project Manager** Oversees the project timeline, coordinates team efforts, and serves as the main point of contact.
- **Ahmed Yehya Saad Nafea: Cloud & Infrastructure Engineer** Responsible for setting up, configuring, and managing the Azure cloud resources.
- **Rania Ossama Hassan Abd Elhaleem: ETL Developer** Focuses on developing the data extraction and transformation logic within Azure Data Factory and Databricks.
- **Merehan Ibrahim El Moatasem Hassan: Data Quality & Testing Analyst** Ensures data integrity, validates the pipeline outputs, and manages the quality assurance process.
- **Hagar Mohamed Mabrouk Mohamed: Data Warehouse & BI Specialist** Designs the data warehouse schema in Azure Synapse and develops the final visualizations in Power BI.
- **Youssef Wael Omar Abu Dief: ETL Developer** Collaborates on building and optimizing the data ingestion and processing pipelines.

Objectives

The primary objective is to build an end-to-end data engineering solution that handles all ETL (Extract, Transform, Load) processes. Key goals include:

- Build a fully automated, end-to-end ETL (Extract, Transform, Load) data pipeline.

- Ingest diverse and high-velocity data streams from various transportation sources.
- Establish a scalable and secure cloud infrastructure using Microsoft Azure services.
- Transform raw data into a clean, structured, and analytics-ready format.
- Serve processed data via an interactive dashboard to enable data-driven decision-making.

Tools & Technologies

The project will be built entirely on the Microsoft Azure cloud platform, leveraging a modern data stack to ensure scalability, reliability, and security. The following services will be used:

- **Infrastructure Management:**

  - **Azure Resource Manager:** For deploying and managing all project resources in a consistent and repeatable manner.

- **Data Storage:**

  - **Azure Blob Storage:** To serve as the primary data lake for storing raw, semi-structured, and unstructured data.
  - **Azure SQL Database:** For storing relational data, metadata, or application state.

- **Compute & Transformation:**

  - **Azure Servers:** To host specific applications or services as needed.
  - **Azure Databricks:** For large-scale data processing and transformation using Apache Spark.

- **ETL & Orchestration:**

  - **Azure Data Factory (ADF):** To build, schedule, and orchestrate the end-to-end ETL/ELT pipelines.

- **Analytics & Warehousing:**

  - **Azure Synapse Analytics:** To serve as the central cloud data warehouse, integrating data for large-scale analytics.

- **Visualization & Reporting:**

  - **Power BI:** To create interactive dashboards and reports for users to explore the data and gain insights.

Milestones & Deadlines

The project is broken down into a series of detailed milestones to ensure steady progress and clear deliverables at each stage. The timeline is structured to move logically from foundational setup to final delivery.

Phase	Milestone	Key Activities	Deadline
1. Foundation & Ingestion	M1: Planning & Setup	• Finalize project scope and technical requirements. • Provision and configure all Azure resources (Blob Storage, ADF, Synapse).	Oct 24, 2025
	M2: Initial Ingestion	• Develop initial data ingestion pipelines using Azure Data Factory. • Establish connections to source systems. • Land raw data successfully into the Azure Blob Storage data lake.	Oct 31, 2025
2. Processing & Warehousing	M3: Data Transformation	• Develop data cleaning, validation, and transformation scripts in Azure Databricks. • Handle data quality issues like missing values and inconsistencies.	Nov 7, 2025
	M4: Warehouse Modeling	• Design the star/snowflake schema for the data warehouse. • Implement tables and	Nov 14, 2025

Phase	Milestone	Key Activities	Deadline
		relationships in Azure Synapse Analytics.	
	M5: Full Pipeline Load	• Integrate the transformation step into the main ADF pipeline. • Perform the first full load of cleaned, structured data into Azure Synapse.	Nov 21, 2025
3. Delivery & Finalization	M6: BI Development	• Connect Power BI to Azure Synapse Analytics. • Build the data model within Power BI and create necessary DAX measures. • Develop the primary interactive dashboard and reports.	Nov 28, 2025
	M7: Testing & QA	• Conduct end-to-end pipeline testing for reliability and performance. • Validate data accuracy from source to dashboard. • Gather feedback from stakeholders (User Acceptance Testing).	Dec 3, 2025
	M8: Final Handover	• Finalize and deploy all project components. • Complete technical documentation and user guides. • Prepare and deliver the final project presentation.	Dec 5, 2025