# Smart City Traffic Data Platform - Project Overview

## 1. Introduction

This document provides a comprehensive overview of the Smart City Traffic Data Pipeline developed using Microsoft Fabric. It covers architecture, ingestion strategy, data modeling, and visualization designed to monitor congestion, incidents, and road conditions across a smart city network.

## 2. Architecture Overview

### 2.1 Architecture Diagram

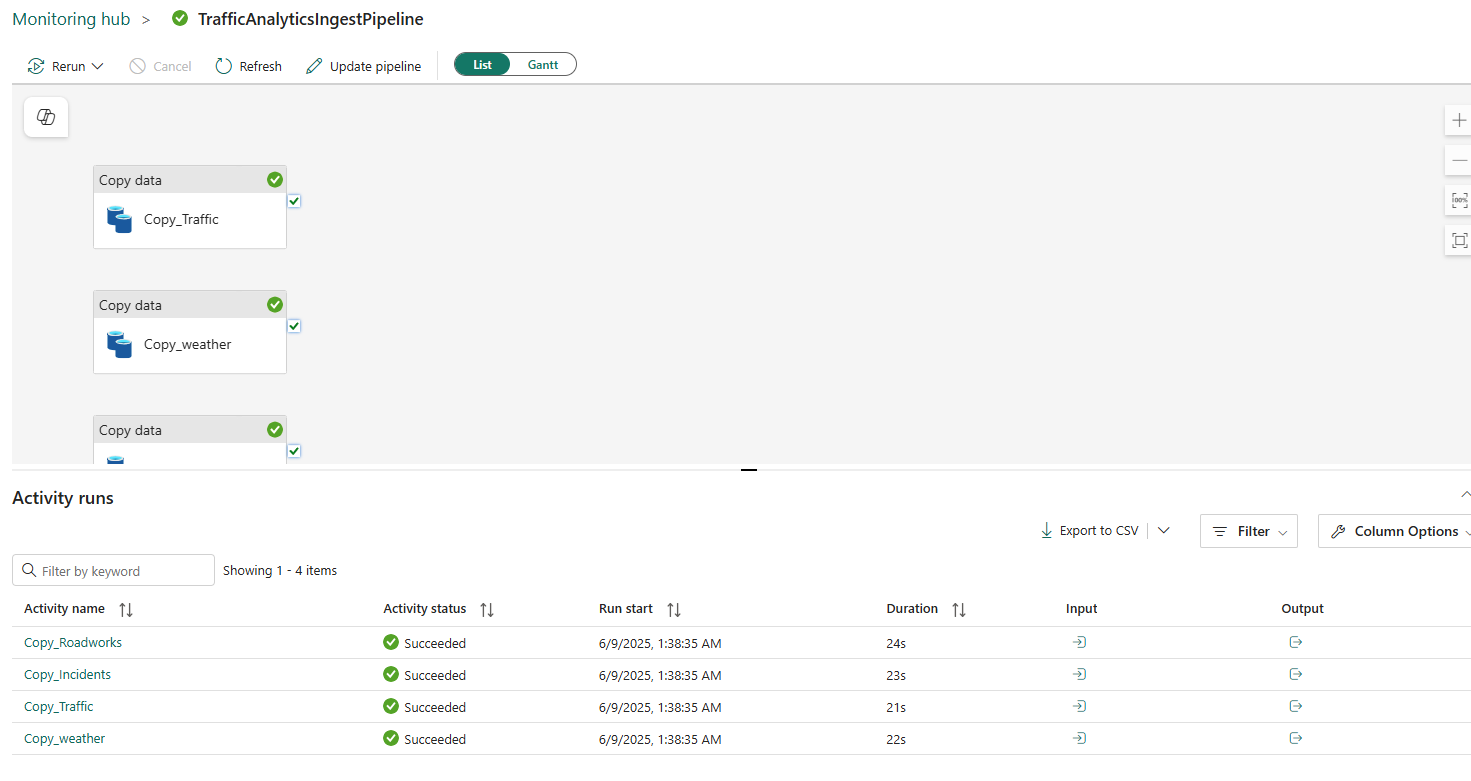
### 2.2 Key Components

* **Traffic Sensors**: Real-time data like vehicle count, speed, and intersection ID.
* **Weather System (JSON API)**: Daily summaries including precipitation, wind speed, and weather conditions.
* **Incident Reports (CSV)**: Includes severity, duration, and location data.
* **Roadworks (Excel)**: Information about road maintenance schedules.
* **Fabric Data Pipelines**: Used for ingestion into Lakehouse staging tables.
* **Spark Notebooks**: Applied for feature engineering and multi-source enrichment.
* **Delta Lake Tables**: Store enriched, optimized datasets for analytics.
* **Power BI**: Final consumer layer with visual dashboards.

## 3. Data Ingestion Strategy

### 3.1 Lakehouse Ingestion Paths

* /Files/raw/traffic/ → stg\_traffic
* /Files/raw/weather/ → stg\_weather
* /Files/raw/incidents/ → stg\_incidents
* /Files/raw/roadworks/ → stg\_roadworks

Ingestion via Data Factory (Copy Activity) into Fabric Lakehouse tables.  
  


## 4. Transformation Highlights

* **Casting and Normalization**: Standardize types (e.g., timestamp, double, int).
* **Derived Metrics**:
  + congestion\_index = vehicle\_count \* occupancy\_rate / avg\_speed
* **Time Features**:
  + Extract hour\_of\_day, is\_peak\_hour, and date.
* **Weather Enrichment**:
  + Join on date, derive is\_severe\_weather.
* **Incident Join (±15 minutes)**:
  + Join on intersection\_id + windowed timestamp.
  + Add incident\_impact = severity \* duration\_minutes.
* **Roadworks Join (range between dates)**:
  + Join on intersection\_id + start\_date ≤ date ≤ end\_date.
  + Add expected\_delay\_minutes.
* **Window Aggregations**:
  + 3-hour rolling metrics per intersection\_id.

A screenshot of a computer

AI-generated content may be incorrect.

## 5. Power BI Visualization

### Dashboard Features:

* **Congestion Heatmaps** per time-of-day and intersection
* **Peak Hour Alerts**
* **Incident Impact Overlays**
* **Weather-Congestion Correlation**
* **Roadwork Forecast Timeline**

## 6. Data Validation

* Row count validation between Bronze and Silver layers
* Schema checks and type coercion audit
* Visual inspection via Power BI

## 7. Deliverables

* ETL script (ETL\_Pipeline.py)
* README.md and this Project\_Overview
* Architecture Diagram
* Power BI Report (published to Fabric)

## 8. Limitations & Future Enhancements

* **Limitation**: Weather data only updated daily.
* **Improvement**: Integrate real-time feeds via Event Streams or IoT Hub.
* **Future Add-on**: ML models for congestion forecasting and incident risk scoring.