IoT-Based Smart Traffic Analytics Pipeline

Transforming urban mobility with real-time insights

Team Leader: *Ahmed Elsayed*

Omar Awad – Co-Architect
Reem Elsayed
Reham Yasser
Amira Elsokary
Abdelrahman Wael

Project Overview

End-to-End Analytics

Simulating, processing, and visualizing real-time traffic data to understand urban flow.

Robust Data Pipeline

Utilizing Kafka for streaming, Airflow for ETL, and Hive/SQLite for storage.

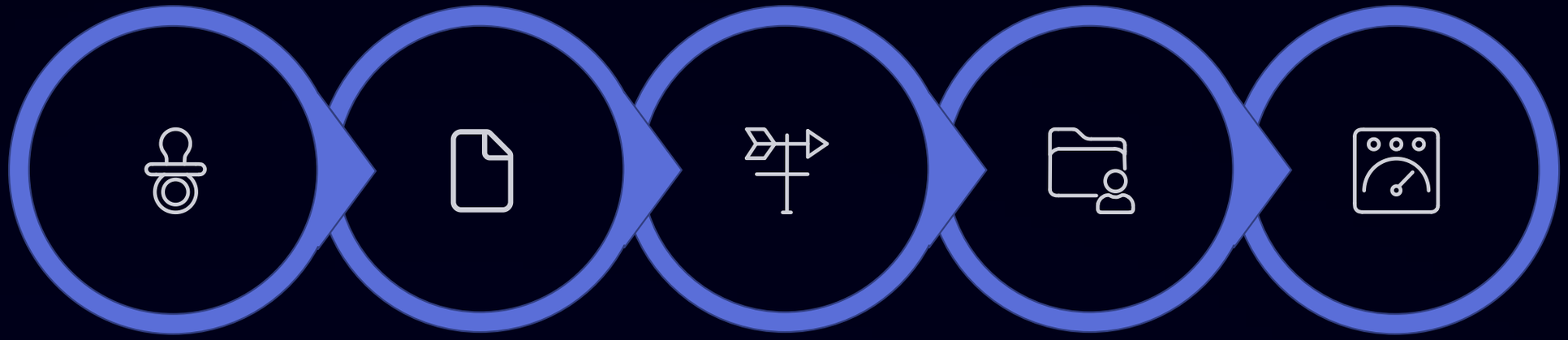
Dynamic Visualization

An interactive Streamlit dashboard for actionable insights and data exploration.

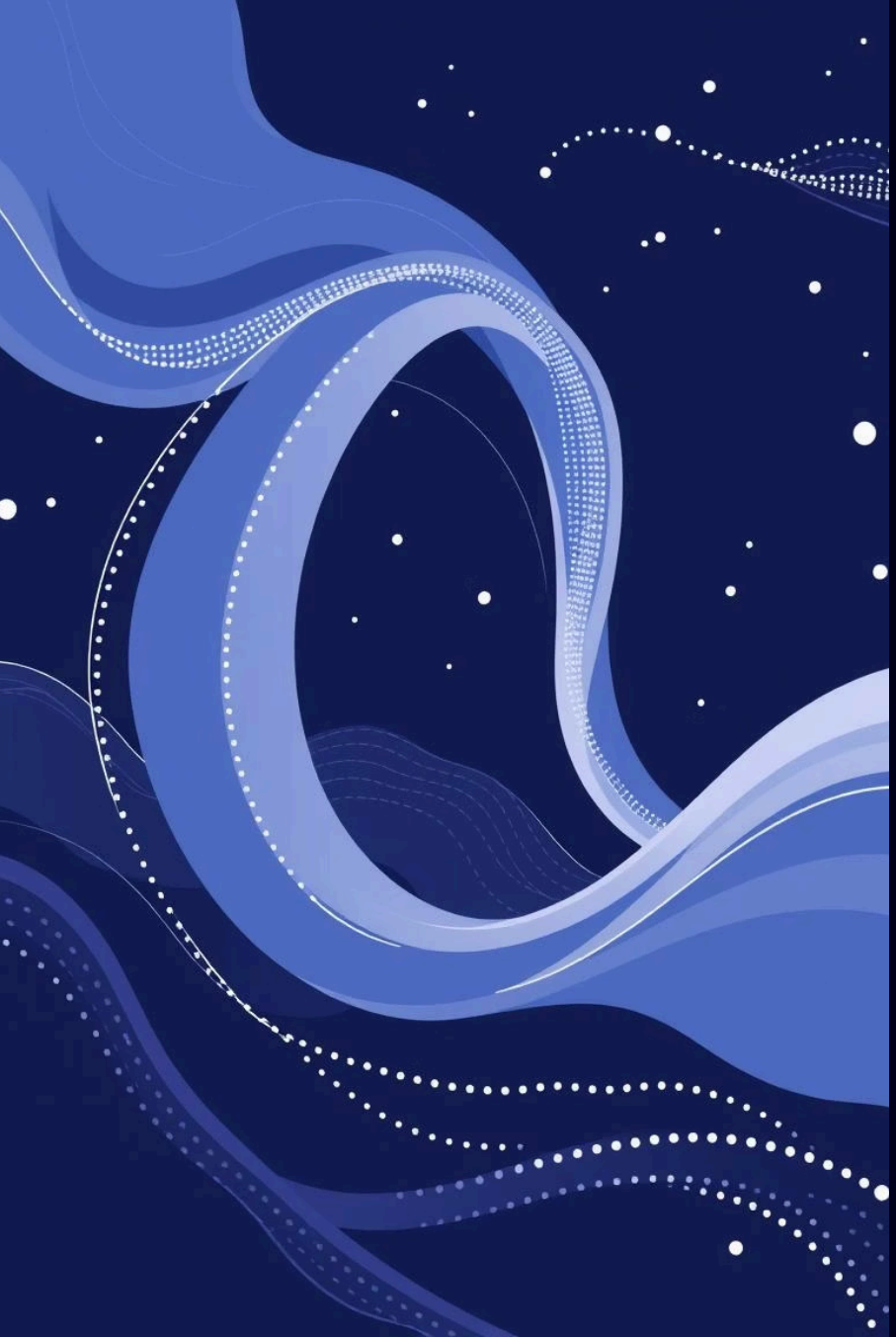
Deep Dive Analytics

Leveraging DuckDB for in-depth inspection and ad-hoc data analysis.

System Architecture: From Simulation to Insights



This architecture ensures a seamless flow of traffic data, enabling real-time processing and comprehensive analysis.



ETL & Data Flow: Processing the Urban Pulse

Extract

Traffic data simulated and collected, ready for ingestion.

Transform

Data is cleaned, validated, and enriched using Airflow workflows.

Load

Processed data is stored efficiently in Hive (for big data) and SQLite (for local access).

Our ETL process ensures data quality and readiness for analytical consumption, providing reliable insights.

Data Model: Understanding Traffic Dynamics

Key fields captured and processed to provide a comprehensive view of traffic conditions.

Field Name	Data Type	Description
timestamp	Datetime	Time of data recording.
vehicle_id	String	Unique identifier for each vehicle.
latitude	Float	Geographic latitude of the vehicle.
longitude	Float	Geographic longitude of the vehicle.
speed_kmh	Integer	Vehicle speed in kilometers per hour.
road_segment	String	Identifies the road segment or street.
traffic_jam	Boolean	Indicates if a traffic jam is detected.

Dashboard Wireframe: Your Traffic Control Center

An intuitive Streamlit interface designed for quick access to critical traffic information.

Sidebar Navigation

- Real-time Metrics
- Historical Trends
- Map View
- Data Export

Key Features

- KPIs for immediate status
- Interactive charts
- Geospatial visualization
- CSV data download

The wireframe highlights a user-friendly design to streamline traffic management and analysis.

Visualizations: Actionable Traffic Insights

Key Performance Indicators (KPIs)

1200

Active Vehicles

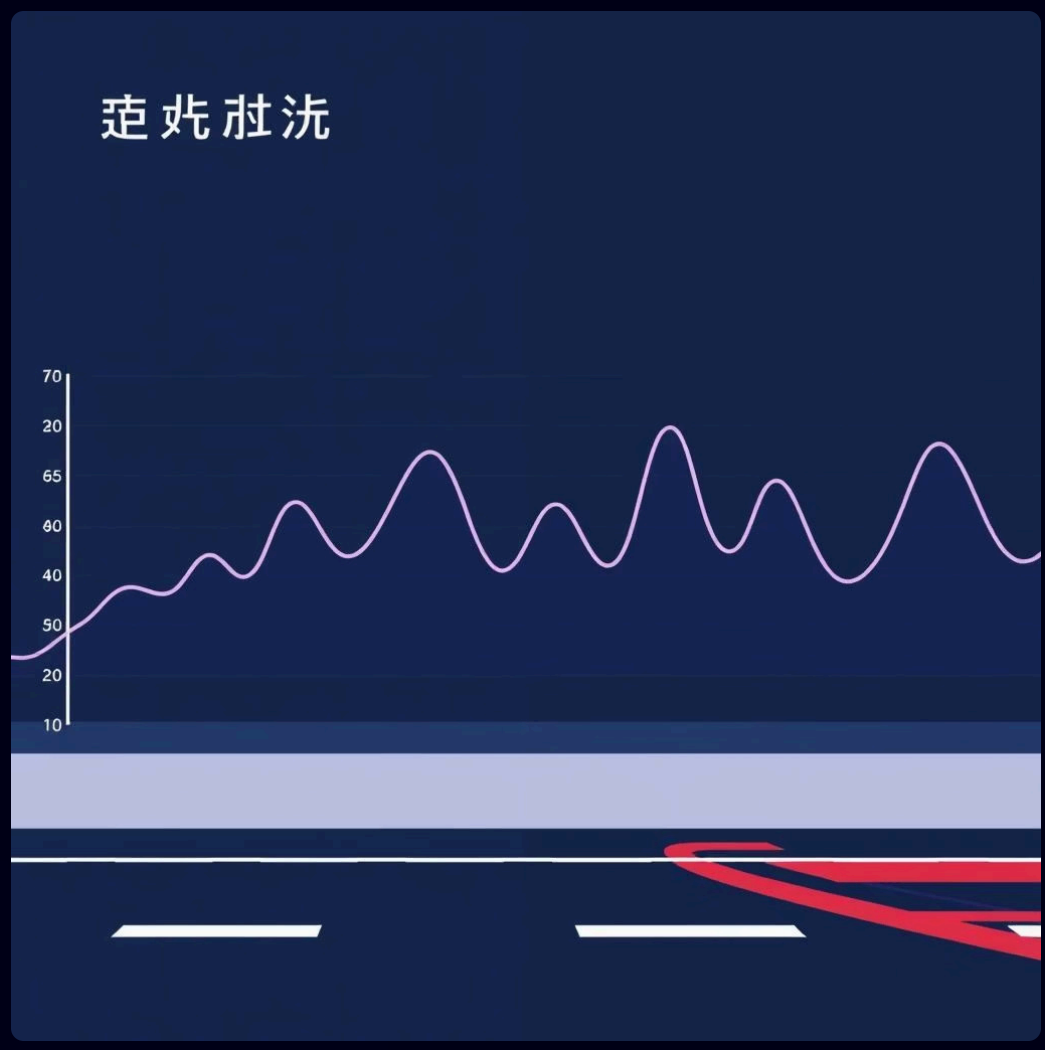
75

Overspeed Alerts

15

Current Traffic Jams

Time-Series & Map View



From real-time counts to geospatial analysis, our dashboard provides diverse views for comprehensive understanding.

Tools & Technologies: Powering the Pipeline

Simulation

Custom Python simulator generates realistic traffic data.

Streaming

Apache Kafka for high-throughput, fault-tolerant data ingestion.

ETL & Orchestration

Apache Airflow manages and automates data processing workflows.

Storage

Hive for scalable data warehousing; SQLite for lightweight local storage.

Visualization

Streamlit for interactive and responsive dashboards.

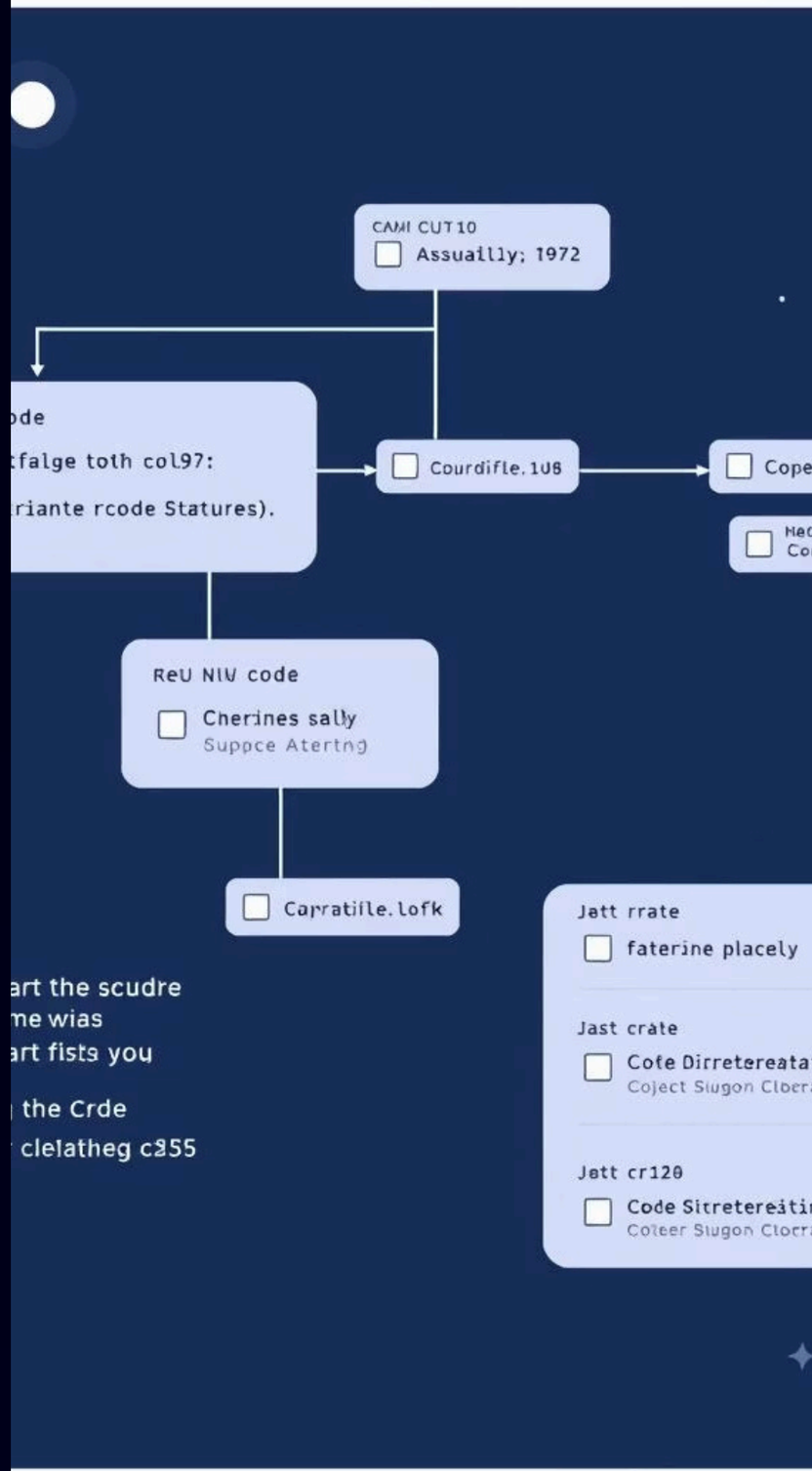
Containerization

Docker ensures consistent environments and easy deployment.

Project Folder Structure: Organized for Clarity

A clean and logical directory structure for easy navigation and development.

```
.
├── dags/
│   └── traffic_etl_dag.py
├── Data/
│   └── raw/
│   └── traffic_data_out.jsonl
├── src/
│   ├── analytics/
│   │   ├── view_data.py
│   │   └── traffic_dashboard.py
│   ├── etl/
│   │   ├── __pycache__/
│   │   │   └── traffic_etl.py
│   │   └── traffic_etl.py
│   └── simulation/
│   └── traffic_simulator.py
├── .gitignore
├── docker-compose.kafka.yml
├── README.md
└── requirements.txt
```



Future Enhancements & Live Demo

Suggested Improvements

- **Real IoT Integration:** Connect to actual traffic sensors.
- **ML Prediction:** Implement models for traffic forecasting.
- **Cloud Database:** Migrate to a scalable cloud solution like Snowflake or BigQuery.
- **Alerting System:** Automated notifications for critical events.

Live Dashboard

Experience the traffic analytics dashboard in action:

[Access Dashboard]