1. Overview

This milestone delivers a fully functional real-time traffic monitoring and anomaly detection system that processes live traffic telemetry every 5 minutes, enriches it with static reference data, detects abnormal traffic patterns, and persists **only anomalous events** for alerting and historical analysis.

The system is built on Azure (IoT Hub $\rightarrow$ Stream Analytics $\rightarrow$ Azure SQL Database) and currently monitors **7 strategic locations** in Greater Cairo.

2. Core Components

Component	Technology	Purpose
TrafficInput (Event Hub / IoT Hub)	Streaming input	Raw traffic events (Timestamp, LocationID, AverageSpeedKMH, CongestionPercentage, TrafficIncident, WeatherCondition, etc.)
Azure Stream Analytics Job	ASA Query (shown below)	5-minute tumbling window aggregation + rule-based anomaly detection
Reference Tables (SQL DB)	Azure SQL Database	Static lookup tables for locations, incidents and weather impact
Anomalies Table (SQL DB)	Azure SQL Database	Stores only anomalous 5-minute windows (normal traffic is filtered out)

3. Database Schema & Reference Data

3.1. dbo.Locations – Monitored Traffic Hotspots in Cairo

Column	Type	Description	Example
LocationID	NVARCHAR(10) PK	Unique identifier	LOC001
LocationName	NVARCHAR(100)	Human readable name	Tahrir Square
Latitude	FLOAT	GPS latitude	30.0444
Longitude	FLOAT	GPS longitude	31.2357
MaxCapacity	INT	Approximate max vehicles before saturation	120

7 locations currently configured: Tahrir, Ramses, 6th October Bridge, Nasr City, Heliopolis, Maadi Corniche, Ahmed Orabi.

3.2. dbo.IncidentReference – Incident Impact on Traffic Speed

IncidentType	SpeedImpactFactor	Effect on average speed
Major Accident	0.450	55% speed reduction
Minor Accident	0.750	25% reduction
None	1.000	No impact
Police Checkpoint	0.850	15% reduction
Road Construction	0.550	45% reduction
Vehicle Breakdown	0.650	35% reduction

Used downstream for predictive "expected speed" calculations (future phase).

3.3. dbo.WeatherReference – Weather Impact on Traffic Speed

WeatherCondition	SpeedImpactFactor	Typical effect
Clear	1.000	Normal speed
Cloudy	0.980	–2%
Light Rain	0.920	–8%
Heavy Rain	0.820	–18%
Foggy	0.800	–20%
Sandstorm	0.550	–45%

Ready for future enrichment of incoming events with weather-adjusted thresholds.

3.4. dbo.Anomalies – Persistent Anomaly Events

Column	Type	Description
EventID	NVARCHAR(100) PK	Unique ID → EVT_$<$LocationID$>$_$<$HHMMSS$>$ (e.g., EVT_LOC001_103000)
LocationID	NVARCHAR(10)	FK → Locations
LocationName	NVARCHAR(100)	Denormalized for fast reads
Latitude/Longitude	FLOAT	Denormalized GPS coordinates
AnomalyType	NVARCHAR(50)	One of: critical_incident, severe_congestion, high_congestion, high_speed, volatile_traffic, minor_incident
Severity	NVARCHAR(20)	critical / high / medium / low
Value	FLOAT NULL	The metric value that triggered the rule
Incident	NVARCHAR(50) NULL	FK → IncidentReference (if incident-driven)

Column	Type	Description
AvgSpeed, MaxSpeed, MinSpeed, MaxCongestion	FLOAT	5-minute aggregated metrics
DetectedAt	DATETIME2	End timestamp of the 5-minute tumbling window

Indexes created for fast dashboards:
IX_Anomalies_DetectedAt (DESC), IX_Anomalies_Severity

4. Azure Stream Analytics Query – Real-Time Anomaly Detection Logic

```
WITH TrafficStats AS (
    SELECT
        System.Timestamp() AS AnalysisTime,
        LocationID,
        LocationName,
        Latitude,
        Longitude,
        AVG(AverageSpeedKMH) AS AvgSpeed,
        MAX(AverageSpeedKMH) AS MaxSpeed,
        MIN(AverageSpeedKMH) AS MinSpeed,
        MAX(CongestionPercentage) AS MaxCongestion,
        MAX(CAST(TrafficIncident AS NVARCHAR(MAX))) AS Incident
    FROM [TrafficInput]
    TIMESTAMP BY CAST([Timestamp] AS datetime)
    GROUP BY
        LocationID, LocationName, Latitude, Longitude,
        TumblingWindow(minute, 5)
),
AnomalyData AS (
    SELECT
        CAST(CONCAT('EVT_', LocationID, '_',
                REPLACE(SUBSTRING(CAST(AnalysisTime AS NVARCHAR(MAX)), 12, 8), ':', ''))
            AS NVARCHAR(MAX)) AS EventID,
        CAST(LocationID AS NVARCHAR(MAX)) AS LocationID,
        CAST(LocationName AS NVARCHAR(MAX)) AS LocationName,
        CAST(Latitude AS FLOAT) AS Latitude,
        CAST(Longitude AS FLOAT) AS Longitude,
        CAST(CASE
            WHEN Incident IN ('Major_Accident', 'Road_Closure') THEN 'critical_incident'
            WHEN AvgSpeed < 15 THEN 'severe_congestion'
            WHEN MaxCongestion > 100 THEN 'high_congestion'
            WHEN MaxSpeed > 90 THEN 'high_speed'
            WHEN (MaxSpeed - MinSpeed) > 30 THEN 'volatile_traffic'
            WHEN Incident <> 'None' THEN 'minor_incident'
            ELSE 'normal'
        END AS NVARCHAR(MAX)) AS AnomalyType,
        CAST(CASE
            WHEN Incident IN ('Major_Accident', 'Road_Closure') THEN 'critical'
            WHEN AvgSpeed < 15 OR MaxCongestion > 100 THEN 'high'
            WHEN MaxSpeed > 90 OR (MaxSpeed - MinSpeed) > 30 THEN 'medium'
            WHEN Incident <> 'None' THEN 'low'
            ELSE 'normal'
        END AS NVARCHAR(MAX)) AS Severity,
        CAST(CASE
```

```
            WHEN Incident IN ('Major_Accident', 'Road_Closure') THEN 100
            WHEN AvgSpeed < 15 THEN AvgSpeed
            WHEN MaxCongestion > 100 THEN MaxCongestion
            WHEN MaxSpeed > 90 THEN MaxSpeed
            WHEN (MaxSpeed - MinSpeed) > 30 THEN (MaxSpeed - MinSpeed)
            ELSE NULL
        END AS FLOAT) AS Value,
        CAST(Incident AS NVARCHAR(MAX)) AS Incident,
        CAST(AvgSpeed AS FLOAT) AS AvgSpeed,
        CAST(MaxSpeed AS FLOAT) AS MaxSpeed,
        CAST(MinSpeed AS FLOAT) AS MinSpeed,
        CAST(MaxCongestion AS FLOAT) AS MaxCongestion,
        CAST(AnalysisTime AS DATETIME) AS DetectedAt
    FROM TrafficStats
)
SELECT
    EventID, LocationID, LocationName, Latitude, Longitude,
    AnomalyType, Severity, Value, Incident,
    AvgSpeed, MaxSpeed, MinSpeed, MaxCongestion, DetectedAt
INTO [AnomaliesOutput]
FROM AnomalyData
WHERE AnomalyType <> 'normal';
```

Key Design Decisions

- Tumbling 5-minute windows → predictable EventID and easy historical alignment.

- Priority-based CASE statements → most severe condition wins.

- All columns explicitly CAST to satisfy Stream Analytics limitations.

- Normal traffic discarded → Anomalies table stays small and actionable.

5. Current Status & Achievements (Milestone Complete)

Item	Status	Notes
7 Cairo locations defined	Completed	Real GPS coordinates
Incident & Weather reference tables	Completed	Ready for future ML-based expected-speed model
Anomalies table with FKs & indexes	Completed	Optimized for Power BI / alerting
Stream Analytics job deployed & tested	Completed	Successfully detects synthetic anomalies
Only anomalous events persisted	Completed	90–95% data reduction in normal conditions

Prepared by: Ali Abdulaziz / Ebrahim Essam