Stage 2 – Complete System Design Documentation

A. Data Model Finalization

Entities:

- Stations
- AQI_Readings
- Weather_Readings
- Traffic_Readings
- Industrial_Zones
- Users
- Citizen_Reports
- Pollution_Events
- Event_Classifications

Key Relationships:

- One Station → Many AQI_Readings
- One Station → Many Weather_Readings
- One Station → Many Traffic_Readings
- One Station → Many Pollution_Events
- One Pollution_Event → One Event_Classification
- One User → Many Citizen_Reports

Index Strategy:

- Composite index on (station_id, timestamp) for AQI_Readings
- Composite index on (station_id, timestamp) for Weather_Readings
- Composite index on (station_id, timestamp) for Traffic_Readings
- Index on timestamp for Citizen_Reports
- Index on event_id for Event_Classifications

Timestamp Handling:

- All timestamps stored in UTC
- ISO 8601 format
- Fixed polling interval: 5 minutes

- Rolling average window: 1 hour (12 readings)

Data Retention Policy:

- Raw readings retained for 1 year
- Aggregated summaries retained indefinitely
- Citizen media stored securely in cloud storage

ER Diagram (Logical Representation):

```
Stations (1) ■■■< AQI_Readings
Stations (1) ■■■< Weather_Readings
Stations (1) ■■■< Traffic_Readings
Stations (1) ■■■< Pollution_Events ■■■ (1) Event_Classifications
Users (1) ■■■< Citizen_Reports
Industrial_Zones (spatial entity)
```

B. Classification Engine Specification

Spike Detection Logic:

- Rolling 1-hour AQI average (12 readings at 5-min interval)
- If current AQI > 1.3 × rolling average → Trigger event
- OR if current AQI > 200 → Trigger event
- Create Pollution_Event record

Classification Thresholds:

- If highest score < 30 → Uncertain
- If difference < 10 → Mixed Contribution
- Else assign highest category

Wind Modifier:

If windSpeed < 1 m/s apply stagnation adjustment.

Confidence Formula:

Confidence (%) = ((TopScore – SecondScore) / TopScore) × 100

C. API Integration Plan

OpenAQ:

- Authentication: Public API access
- Rate Limit: Approx. 60 requests/min
- Response: JSON with pollutant values & station metadata
- Error Handling: 3 retries with exponential backoff
- Fallback: Use cached AQI snapshot

Open-Meteo:

- Authentication: No key required
- Rate Limit: Fair use
- Response: JSON wind speed & direction
- Error Handling: Validate ranges before storing
- Fallback: Retain last valid wind data

Google Traffic API:

- Authentication: API key in environment variables
- Rate Limit: Based on billing tier
- Response: Traffic congestion mapped to 0–1
- Error Handling: Log failure and continue
- Fallback: Use previous congestion value

Data Ingestion Blueprint:

- Fetch AQI every 5 minutes
- Fetch weather & traffic data
- Normalize responses
- Store in DB (UTC timestamps)
- Run spike detection
- If spike → Run classification
- Store Event_Classification