Architecture for Real-Time Air Quality Monitoring

and Personalized Health Recommendations in Bogotá

Jose Alejandro Cortazar López, Johan Esteban Castaño Martínez, Stivel Pinilla Puerta
Systems Engineering Program - Francisco José de Caldas District University

Air pollution is a critical public health challenge. This work presents a production-ready architecture for **real-time air quality monitoring in Bogotá** that integrates heterogeneous data sources (AQICN, Google Air Quality API, IQAir) into a unified 3NF PostgreSQL schema with sub-100 millisecond query latencies and personalized health recommendations.

Primary Goal: Design and validate a production-ready architecture integrating multi-source air quality data with normalized schema and personalized health recommendations.

Specific Objectives:

1. Integrate AQICN, Google, & IQAir with 216 readings/hour via Python APScheduler
2. Implement 3NF PostgreSQL schema (8 entities) with temporal partitioning
3. Achieve sub-100ms query latencies across 85,000+ readings
4. Support 50–100 concurrent users with documented concurrency mitigation
5. Deliver rule-based health recommendations per EPA/WHO AQI standards

Database Schema: 8 normalized entities with referential integrity:

- Station, Pollutant, Provider, AirQualityReading
- AppUser, Alert, Recommendation, ProductRecommendation

Query Optimization:

- Composite B-tree indexes on (Station, Timestamp, Pollutant)
- Temporal partitioning with constraint exclusion (30.2% improvement)
- Materialized views for monthly aggregates (35× row reduction)
- Redis caching with 5–10 minute TTL

Data Pipeline:

- JSON validation via Pydantic models
- Deduplication and idempotent upserts
- MVCC isolation prevents read blocking
- Zero-downtime ingestion compatible with analytics

Personalization Engine: Rule-based health guidance mapping AQI to recommendations by age, respiratory conditions, and activity level (EPA/WHO standards).