Workshop 1 - Project

Air Quality Analysis Platform

Members:

Stivel Pinilla Puerta – Cod.: 20191020024

Johan Castaño Martinez - Cod.: 20191020029

Teacher:

Carlos Andrés Sierra

Universidad Distrital Francisco José De Caldas

Faculty Of Engineering

Systems Engineering

Databases II

May 2025, Bogotá D.C.

1. Business Model:

Air Quality Analysis Platform such as AQICN

The application integrates spatial databases, big data and business intelligence to provide real-time information.

Key Partners

- Governmental environmental agencies: IDEAM, EPA, and others providing official data and scientific validation.
- Air quality sensor companies: PurpleAir, BreezoMeter, as providers of hardware and real-time data.
- Weather data providers: NOAA, OpenWeather, needed to correlate environmental data.
- Tourism companies: Hotels, airlines, booking platforms that can benefit from the added value of environmental quality data.
- Universities and research centers: To validate prediction models, collaborate in innovation and obtain scientific support.
- Media and environmental NGOs: For dissemination, legitimization and possible social impact alliances.

Key Activities

- Collection, cleaning and validation of air quality and meteorological data.
- Real-time data processing using big data and AI algorithms.
- Development and maintenance of web platform and mobile app.
- Generation of customized visualizations and reports.
- Integration of APIs with third party systems (tourism, health, mobility).
- Creation of customized recommendation algorithms.
- Research and improvement of predictive models.
- User acquisition strategies and educational marketing.

Key Resources

- Technological infrastructure: servers, storage, spatial databases.
- Human talent: developers, data scientists, DevOps, UX/UI designers, BI specialists.
- Agreements/licenses with meteorological and environmental data sources.
- Analysis, visualization and recommendation algorithms.
- Accessible and scalable graphical interface for end users.
- Integration capability through open APIs.

Value Proposition

- Real-time environmental information contextualized by location.
- Personalized recommendations based on health, tourism and mobility.
- Visualizations and reports in clear language for citizens, authorities and companies.
- BI platform for analysis of historical trends and behaviors.
- Easy integration via API with other platforms and dashboards.
- Intelligent alerts to minimize exposure risks.
- Contribution to compliance with environmental regulations and policies.

Customer Relationships

- Personalized online assistance via chat, email and community forums.
- Educational content (blogs, videos, infographics) on health and environment.
- Proactive air quality notifications based on location or interests.
- Access to premium reports under subscription or one-time payment.
- Loyalty programs (gamification, challenges, rewards).
- Specialized attention for governments and companies through technical consulting.

Channels

- Web platform and mobile application as main service channels.
- Email marketing with customized alerts and newsletters.
- Social networks to generate awareness and brand positioning.
- Alliances with media to amplify reach.
- Participation in events, technology and environmental fairs.
- Integrations with health, transportation and tourism apps through APIs.

Customer Segments

- Citizens aware of their environmental health: athletes, parents, seniors.
- Companies with physical locations interested in improving indoor air quality and sustainable reputation.
- Local and national governments formulating policies and seeking monitoring and warning systems.
- NGOs and foundations interested in research and promotion of environmental awareness.
- Tourists and travel agencies seeking low pollution destinations.
- Educational institutions and research centers in need of reliable data.

Cost Structure

- Fixed infrastructure costs (servers, big data storage).
- Development, testing and maintenance of the platform (web/app).
- Technical and administrative staff salaries.
- Licensing of meteorological data and third-party APIs.
- Research and development (AI models, BI, data visualization).
- Digital marketing, partnerships, outreach and training costs.
- Technical support and customer service.

Revenue Streams

- Premium subscription for users who access detailed reports and advanced functions.
- Customized consulting for governments and companies on environmental strategies.
- Advertising and sponsorship of healthy and sustainable brands.
- API sales and access to real-time data for developers and companies.
- Licensing of dashboards to educational, municipal or tourism entities.
- Indirect monetization through partnerships with apps that integrate the platform.

2. Requirements Gathering:

ID	Type	Requirement	Associated User Stories
FR1	Functional	The system must collect real-time air quality data from multiple external sources (e.g., APIs, stations) and store it for further processing.	US1, US13, US14
FR2	Functional	The system must allow users to query and visualize historical air quality data filtered by date, location, and pollutant type.	US2, US7
FR3	Functional	The system must process, clean, and store high-volume data with transformation mechanisms to ensure quality.	US1, US3
FR4	Functional	The system must display real-time dashboards with key performance indicators (KPIs) on air quality.	US4
FR5	Functional	The system must generate customized reports with filters for date, location, and indicators, available for download or email.	US5
FR6	Functional	The system must present interactive graphs showing air quality evolution over time.	US6
FR7	Functional	The system must allow users to export historical data in standard formats like CSV and JSON.	US7
FR8	Functional	The system must provide personalized recommendations based on user location and air quality conditions.	US8
FR9	Functional	The system must send customizable air quality alerts when thresholds are exceeded.	US9
FR10	Functional	The system must suggest certified protective products during high pollution periods.	US10
FR11	Functional	The system must display maps highlighting areas with better air quality for navigation.	US11
FR12	Functional	The system must support geographic search for air quality data by country, city, or region.	US15
FR13	Functional	The system must be responsive and compatible with mobile, tablet, and desktop devices.	US16
FR14	Functional	The system must allow users to share air quality data and reports on social media with pre-generated links and previews.	US17
NFR1	Non-Functional	Queries on large datasets (≥1 million records) must execute in under 2 seconds 95% of the time.	US3, US12
NFR2	Non-Functional	The system must support continuous streaming data ingestion 24/7 without manual intervention.	US1, US14
NFR3	Non-Functional	Data storage must be distributed and optimized for big data processing.	US1, US3

NFR4	Non-Functional	Customized reports must be generated in under 10 seconds.	US5
NFR5	Non-Functional	The recommendation engine must update in real time as air quality data changes.	US8, US10
NFR6	Non-Functional	Air quality data and visualizations must load in less than 2 seconds for 95% of user requests.	US12
NFR7	Non-Functional	The system must implement load balancing to handle high concurrent traffic.	US13
NFR8	Non-Functional	The system architecture must include fault tolerance and geographic redundancy.	US14
NFR9	Non-Functional	The system must scale horizontally to support growth in data volume and users.	US13, US14
NFR10	Non-Functional	The system UI must function correctly on all major browsers and operating systems without errors.	US16
NFR11	Non-Functional	Data consistency and user personalization must be preserved across all user devices.	US16

3. Enhanced User Stories:

User Story		
Theme 1: Data Ingest and Data Processing (Big Data)		
User Story 1	As a Technical Administrator, I want to collect real-time data from multiple stations and APIs so that I can provide accurate and up-to-date information on air quality.	The system receives updates at least every 10 minutes from configured sources., Processes at least 1000 data points per minute without loss., Logs successful and failed ingestions with error codes.
User Story 2	As a Researcher/Analyst, I want to access historical data by date and location so that I can perform longitudinal analysis and scientific research.	User can filter by city, date, and pollutant type., Data export available in CSV and JSON., Historical records available for up to 3 years.
User Story 3	As a Technical Administrator, I want to run queries over large volumes of data quickly so that I can avoid delays when searching for information.	Queries over 1 million records return in under 3 seconds., Indexes and partitions are used for performance.
Theme 2: Business Intelligence (BI) Module		
User Story 4	As a Public Policy Manager, I want to access dashboards with real-time analysis so that I can issue alerts or recommendations to the public.	Dashboards include real-time maps, graphs, and alerts., Automatic refresh without manual reload., Critical thresholds can be configured for alerts.
User Story 5	As a Public Policy Manager, I want to generate customized reports on	User can choose indicators, date ranges, and export formats., Reports are downloadable in PDF or sent via email.,

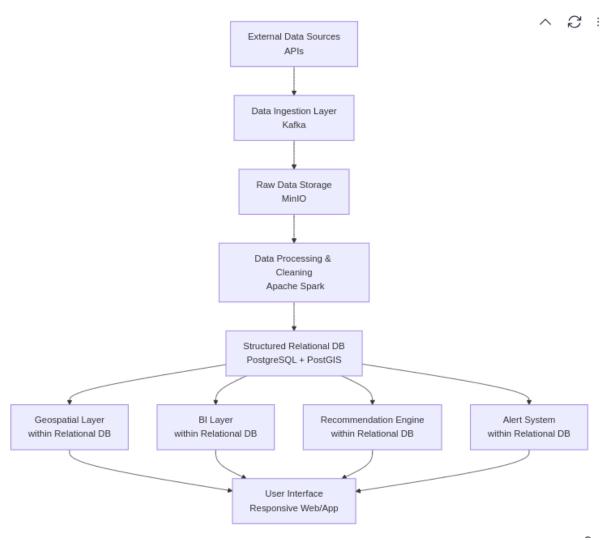|	pollution trends so that I can design evidence-based public policies.	Graphs are auto-generated from selected data.
User Story 6	As a Citizen, I want to view interactive graphs about air quality evolution so that I can understand changes and make informed decisions.	Filters for location, date, and pollutant available., Charts update dynamically with parameter changes., Visualizations load in under 2 seconds.
User Story 7	As a Researcher/Analyst, I want to export historical data in standard formats so that I can analyze it with my own statistical tools.	Interface allows selection of location, date, and format., Files download successfully without errors., Download limits prevent system overload.
Theme 3: Personalized Recommendation System		
User Story 8	As a Citizen, I want to receive personalized recommendations based on air quality so that I can know if it is safe to do outdoor activities.	Location and user preferences are considered., Alert color coding (green, yellow, red) is used., Suggested actions are clearly displayed.
User Story 9	As a Citizen, I want to receive early warnings when air quality is harmful so that I can take precautions in time.	User can set alerts by pollutant and critical threshold., Notifications sent via email or mobile app., Alert triggers when AQI surpasses configured limits.
User Story 10	As a Citizen, I want to receive suggestions for protective products so that I can protect my health during high pollution levels.	Suggestions appear only during high pollution periods., Products are certified and include links., User can disable this feature in preferences.
User Story 11	As a Mobile Citizen, I want to see areas with better air quality so that I can avoid the most polluted zones when moving.	System displays a map with AQI levels per area., Users can compare different city routes., Interface is mobile-friendly and responsive.
Theme 4: High Availability and Scalability		
User Story 12	As a Citizen, I want to load air quality data quickly so that I can access information without delay.	Air quality data loads in under 2 seconds normally., System handles 100 concurrent users without slowdown., Cache is used for frequently accessed queries.
User Story 13	As a Technical Administrator, I want to handle traffic peaks without failure so that I can maintain user experience during high demand.	Stress test simulates 1000 concurrent users., System remains responsive under load., Latency remains below 5 seconds per request.
User Story 14	As a Technical Administrator, I want to keep the platform available 24/7 so that I can ensure constant access to information.	Platform includes geographic redundancy., Uptime monitoring with automated alerts is enabled., Monthly availability is $\geq$ 99.9%.
Theme 5: Multiregion and Multidevice Access		

User Story 15	As a Citizen, I want to check air quality in different regions so that I can plan trips and activities.	Users can switch between countries/regions easily., Information is shown in local language where possible.
User Story 16	As a Citizen, I want to access the platform from various devices so that I can always access data regardless of the device.	Responsive design across mobile, tablet, and desktop., Core functions work on all devices.
User Story 17	As a Citizen, I want to share air quality data on social media so that I can inform and raise awareness among others.	Sharing available on X, Facebook, Instagram, WhatsApp., Preview includes summary and image., Links and QR codes are auto-generated for sharing.

4. Database Architecture:
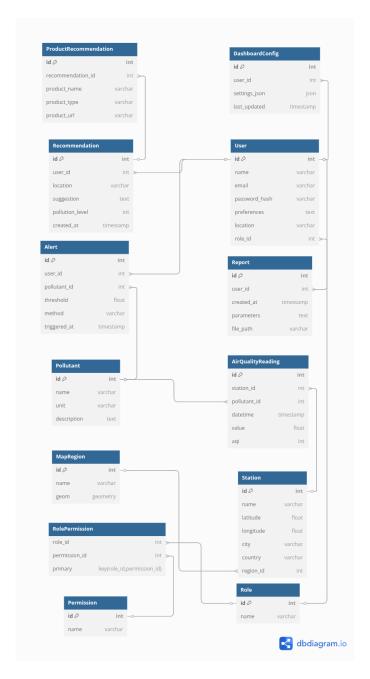
High-Level Database Architecture:

Component	Description	Suggested Tool
External Data Sources	Sources of air quality data including public APIs and monitoring stations.	Custom integration using REST APIs
Data Ingestion Layer	Captures and streams incoming data in real-time from external sources.	Apache Kafka
Raw Data Storage (Data Lake)	Stores unstructured raw data for historical reference and reprocessing.	MinIO
Data Processing & Cleaning Layer	Cleans, validates, and transforms raw data before storage in the relational database.	Apache Spark

Structured Relational Database	Houses cleaned, structured data with support for geospatial and time-series queries.	PostgreSQL + PostGIS
Geospatial Layer (within Relational DB)	Supports geospatial queries and mapping functionalities.	
BI Query Layer (within Relational DB)	Indexes, views and partitions for optimized analytics and reporting.	
Recommendation Data Layer (within Relational DB)	Holds user preferences, rules, and outputs for suggestions.	
Alert Management Layer (within Relational DB)	Stores thresholds and triggered alert events.	
User Interface (Web/App)	Enables users to interact with data through maps, charts, and reports.	React + Leaflet.js

ER Diagram

Component	Entity Name	Description	Attributes
Processed Relational Database	Station	Monitoring station metadata and location.	id (PK), name, latitude, longitude, city, country
Processed Relational Database	AirQualityReading	Air quality measurements from stations.	id (PK), station_id (FK), datetime, pollutant_type, value, aqi
Processed Relational Database	Pollutant	Catalog of pollutants and their units.	id (PK), name, unit, description

			id (PK), name, email,
Processed Relational Database	User	Platform user and profile settings.	id (PK), name, email, password_hash, preferences, location
Alert Management Layer	Alert	Alert configurations and triggered events.	id (PK), user_id (FK), pollutant_type, threshold, method, triggered_at
Recommendation Data Layer	Recommendation	User-specific recommendations based on AQI.	id (PK), user_id (FK), location, suggestion, pollution_level, created_at
BI Query Layer	Report	Generated analytical reports.	id (PK), user_id (FK), created_at, parameters, file_path
Geospatial Layer	MapRegion	Regions used for geospatial filtering and navigation.	id (PK), name, geom (geometry)
BI Query Layer	DashboardConfig	User preferences for dashboard visualizations.	id (PK), user_id (FK), settings_json, last_updated
Recommendation Data Layer	ProductRecommendation	Suggested certified products during high pollution levels.	id (PK), recommendation_id (FK), product_name, product_type, product_url
Processed Relational Database	Role	Defines user roles in the platform (e.g., Admin, Citizen, Analyst).	id (PK), name (unique)
Processed Relational Database	Permission	Specifies distinct permissions that can be granted to roles.	id (PK), name (unique)
Processed Relational Database	RolePermission	Associates roles with permissions in a many-to-many relationship.	role_id (FK), permission_id (FK), PRIMARY KEY (role_id, permission_id)

5. References

Dbdiagram.io. Holistics Software. (2024). Design and visualize your database. Available in: https://dbdiagram.io/

Apache Kafka. Apache Software Foundation. (2024). *Apache Kafka: A Distributed Streaming Platform*. Available in: https://kafka.apache.org/

PostGIS. Refractions Research. (2024). *PostGIS - Spatial and Geographic Objects for PostgreSQL*. Available in: https://postgis.net/

AQICN API Documentation. World Air Quality Index Project. (2024). *Real-time Air Quality Index API*. Available in: https://aqicn.org/api/