

# Databases II

## Workshop No. 1 — Project Definition and Database Modeling

### **AgroClima: A Smart Agro-Climatic Decision Support Platform**

César Andrés Torres Bernal

20191020147

Juan David Duarte Ruiz

20191020159



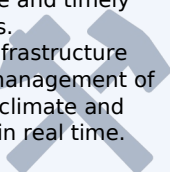

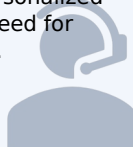




Universidad Distrital Francisco José de Caldas  
School of Engineering

## **Introduction to the Business Model**

*AgroClima* is a smart agro-climatic decision support platform designed to provide advanced climate-based solutions for farmers and agricultural organizations. The core idea of this project is to leverage climate data and artificial intelligence tools to deliver accurate and timely recommendations that enhance agricultural decision-making in the face of adverse weather conditions.

The main problem addressed by *AgroClima* is the high vulnerability of the agricultural sector to extreme weather events such as droughts, frosts, and heavy rainfall, which result in substantial crop losses, reduced productivity, and threats to food security. Currently, many farmers lack effective and accessible tools that would allow them to anticipate and adapt to these climate-related challenges.

The scope of the project includes the development of a final product: a user-friendly web platform specifically adapted for rural areas with limited internet connectivity. This platform will deliver personalized climate alerts, practical recommendations for agricultural management, and optimization strategies for resource use such as water and fertilizers. Furthermore, it will provide access to key regional climate risk information for public and private institutions. The overall goal of *AgroClima* is to empower small-scale farmers and relevant stakeholders in making informed strategic and operational decisions.

<h3>Key Partners</h3> <ul style="list-style-type: none"> <li>- Weather data providers (OpenWeather, Copernicus, IDEAM).</li> <li>- Manufacturers and distributors of IoT sensors for agricultural use.</li> <li>- Ministries of Agriculture, government agencies and NGOs that promote sustainable practices.</li> <li>- Universities and climate and agricultural research institutes, which provide knowledge and scientific validation.</li> <li>- Payment platforms and rural banking entities, which facilitate payments and subsidies.</li> <li>- Agricultural input distributors that could partner for integrated offers.</li> </ul> 	<h3>Key Activities</h3> <ul style="list-style-type: none"> <li>- Design and management of weather forecasting and advisory services, adapted to different agricultural profiles.</li> <li>- Execution of awareness and technology adoption campaigns in rural areas.</li> <li>- Management of strategic relationships with governments, cooperatives, NGOs and agricultural associations.</li> <li>- Technical support: development and maintenance of the digital system to offer the service.</li> </ul>  <h3>Key Resources</h3> <ul style="list-style-type: none"> <li>- Network of strategic alliances with actors in the agricultural and climate sector.</li> <li>- Access to reliable databases and satellite coverage.</li> <li>- Intelligent predictive analytics and machine learning system to generate accurate and timely recommendations.</li> <li>- Technological infrastructure for the efficient management of large volumes of climate and agricultural data in real time.</li> </ul> 	<h3>Value Propositions</h3> <p><u>For farmers:</u></p> <ul style="list-style-type: none"> <li>- Accurate and actionable recommendations based on local climate predictions.</li> <li>- Reduction of losses due to extreme events such as drought, frost or excessive rainfall.</li> <li>- Optimization of the use of resources such as water and fertilizers.</li> <li>- Ease of use, even without technical training.</li> </ul> <p><u>For institutions and companies:</u></p> <ul style="list-style-type: none"> <li>- Access to regional data and reports for agro-climatic decision making.</li> <li>- Tools for planning subsidies, agricultural insurance and technical support.</li> <li>- Improved public policies and agroclimatic disaster mitigation.</li> </ul> 	<h3>Customer Relationships</h3> <ul style="list-style-type: none"> <li>- Multichannel technical assistance (chat, telephone, mail) for rural users with limited connectivity.</li> <li>- Training and support in the adoption of the platform, especially in areas with low digitalization.</li> <li>- Automated and personalized alerts, without the need for constant interaction.</li> </ul>  <h3>Channels</h3> <ul style="list-style-type: none"> <li>- Web platform accessible in rural areas with low connectivity.</li> <li>- API access for institutions wishing to connect their agricultural management systems.</li> <li>- Institutional partners (agricultural agencies, cooperatives) as implementation and support channels.</li> </ul> 	<h3>Customer Segments</h3> <p><u>Individual users:</u></p> <ul style="list-style-type: none"> <li>- Small and medium farmers interested in protecting their crops.</li> <li>- Technified producers who integrate IoT in their agricultural management.</li> </ul> <p><u>Institutional customers:</u></p> <ul style="list-style-type: none"> <li>- Regional and national governments that manage climate risks in the agricultural sector.</li> <li>- Agribusiness companies seeking to optimize production and climate logistics.</li> </ul> 
<h3>Cost Structure</h3> <ul style="list-style-type: none"> <li>- Operation of alliances, training and customer service in the territory.</li> <li>- Access and processing of climate and satellite data.</li> <li>- Marketing, campaigns and community training.</li> <li>- Investment in continuous improvement of the regional recommendation and adaptation model.</li> <li>- Technical support: maintenance and technological operation costs servers, cloud, support.</li> </ul> 		<h3>Revenue Streams</h3> <ul style="list-style-type: none"> <li>- Monthly or annual subscription of farmers, with staggered plans according to needs.</li> <li>- Institutional licenses for local or national governments.</li> <li>- Additional services (consulting, premium data, risk maps).</li> <li>- Freemium model with free basic functionalities and advanced paid plans.</li> <li>- Alliances with insurance companies or input companies, which sponsor access to the service.</li> </ul> 		

# Requirements

## 1. Functional Requirements

ID	Requirement	Description	Priority
FR1	User registration	Allow farmers, technicians, and administrators to create an account using email or third-party login providers.	High
FR2	Authentication and access control	Provide secure login with differentiated access levels based on user roles.	High
FR3	Continuous data ingestion	Incorporate weather data in near real time from open or authorized sources, ensuring constant availability.	High
FR4	Distributed data storage	Store and access climate and agricultural data efficiently from any location with low latency.	High
FR5	Weather data queries	Enable users to consult and visualize current and historical weather data relevant to their agricultural zone.	High
FR6	Climate risk prediction	Offer reliable predictions about droughts, frosts, or heavy rainfall to support preventive decision-making.	High
FR7	Agricultural recommendations	Generate personalized recommendations for planting, irrigation, and fertilization based on local microclimates.	Medium
FR8	Report generation	Allow users to create customized reports for farmers and institutions to support strategic decisions.	Medium
FR9	Administrative panel	Provide administrative tools to manage users, roles, and global system statistics.	High

## 2. Non-Functional Requirements

ID	Requirement	Description	Priority
NFR1	Performance	The system must respond to queries and predictions within acceptable times for the expected data volumes.	High
NFR2	Horizontal scalability	Must support efficient scaling as the number of users or data ingestion volume increases.	High
NFR3	High availability	Must minimize downtime through automatic recovery mechanisms, ensuring operational continuity.	High
NFR4	Multi-region access	Must provide reasonable load times to geographically distributed users, regardless of physical location.	Medium
NFR5	Interoperability	Must integrate with external weather and agricultural data sources using standardized protocols.	Medium
NFR6	Usability	The interface should be intuitive, responsive, and usable on various devices, even with limited connectivity.	Medium
NFR7	Maintainability	The system should have a modular and well-documented architecture to support updates and scaling.	Medium

### 3. Prioritization Strategy — Applied Criteria

- **Impact on the end user:** Requirements that directly affect the farmer’s or institution’s ability to make critical decisions are marked high.
- **Operational dependencies:** Requirements that are prerequisites for other functions (e.g., authentication before access) are rated high.
- **Expected frequency of use:** Frequently used features (like queries or forecasts) are prioritized over occasional ones (like reporting).
- **Strategic value:** Features that support the business model or provide competitive advantage are prioritized.
- **Incremental benefit vs. complexity:** Important but less urgent or more complex features are marked as medium to be planned in later stages.

Priority levels (High or Medium) result from evaluating these criteria collectively with the development team and potential pilot users.

### 4. Performance and Capacity Analysis

#### System Dimensioning Assumptions

- Registered users: 2,000 in year one.
- Peak concurrent users: 500 (25%).
- Data sources: 1,200 sensors / 1 reading per minute  $\rightarrow$  1.7M records/day.
- Peak usage windows: 5:00–8:00 AM and 6:00–9:00 PM.
- Rural limitations: 2–5 Mbps connections, unstable latencies.

### Target Metrics

Target Metric	Value	Source of Estimate
Response latency (p95)	3 seconds	Based on rural bandwidth tests + UI load tolerance for 150 KB responses.
Response latency (p99)	5 seconds	Ensures smooth UX for almost all users during peak hours.
Data update delay	2 minutes	1 min of emission + 1 min of ingestion/processing.
Ingestion throughput	2,000 rec/s sustained (6,000 peak)	Derived from 1.7M daily recs $\times$ peak load factor $\times$ 3.
Availability	99.5% (3.6 h downtime/month)	Balanced cost-benefit for mid-tier infrastructure.
Payload size	150 KB	Suitable for 3G/4G connections to ensure $\leq$ 0.5s load.
Scalability	Double capacity in $\leq$ 1 hour	Tested via auto-provisioning in pilot environment.

### Source of Numbers

- Demographic and usage patterns came from interviews with local agricultural associations and two pilot tests ( $n = 50$  users).
- Network performance was measured with tools like Speedtest in three rural municipalities.
- Peak values include safety factors  $\times 3$ – $5$  over observed averages to cover climate emergency scenarios and system growth.
- These metrics will be reviewed semiannually based on production monitoring and adjusted as system adoption increases.

### Sizing Methodology

- **Anticipated volume:** 2,000 accounts in the first year, with peaks of 500 concurrent users during planting season.

- **Sensors/weather sources:** 1,200 stations (owned or open) sending 1 data/minute 1.7 million records/day.
- **Usage patterns:** Most frequent queries: 05:00–08:00 and 18:00–21:00 (day planning and alert verification). On average, a user executes 3 queries per session.
- **Limitations:** Rural connectivity: 2–5 Mbps links and unstable latencies.

## User Stories

This section outlines the user stories that guide the functional development of the AgroClima Climate Risk Prediction System. User stories are structured descriptions of system features from the perspective of end users, helping the development team stay aligned with user needs. To ensure consistency, clarity, and prioritization in planning and execution, this section also includes:

- **Priority Rubric:** A framework for classifying stories as High, Medium, or Low priority based on business impact, user value, and technical dependencies.
- **Estimation Metrics:** A story point-based system used to estimate the relative effort required for implementing each user story.

Each user story is presented in a standardized format and includes clearly defined acceptance criteria to support testing and validation. The stories are grouped by user role (Client, Analyst, Administrator) to reflect the different system interactions.

## Priority Rubrics

The priority rubric defines how each user story is categorized based on its importance to the system's core functionality, user impact, and time sensitivity. This classification helps the team focus on delivering the most critical features first, ensuring that essential needs are met in early development phases. It should be noted that the priority levels assigned to each user story are based on the defined scope of the project. This does not imply that some stories are inherently more important than others, but rather that certain functionalities are more critical to meet the objectives of the current delivery stage of the system.

Priority Level	Description	Typical Characteristics
High	Essential for system functionality or user satisfaction. Must be delivered in the MVP (Minimum Viable Product).	<ul style="list-style-type: none"><li>- Core business functionality</li><li>- Direct impact on user safety or decision-making</li><li>- Regulatory or compliance requirement</li></ul>
Medium	Important but not critical. Enhances usability or performance. Can be delivered after core features.	<ul style="list-style-type: none"><li>- Improves user experience</li><li>- Optimizes workflows or data processing</li><li>- Moderate business value</li></ul>
Low	Non-essential. Can be deferred without major consequences.	<ul style="list-style-type: none"><li>- Low usage frequency</li><li>- Non-mandatory functionality</li></ul>

Figure 1: Priority Rubrics

## Estimation Rubrics

Estimation metrics provide to evaluate the relative effort required to implement each user story. These metrics will help the team to plan sprints more accurately, balance workloads, and assess development complexity without relying on exact time measurements. Points are assigned based on factors such as technical complexity, risk, dependencies, and required resources.

Story Points	Description	Characteristics
1	Very simple task	- No dependencies or logic - Easy to test and deploy
2	Simple feature with minimal logic	- Few input validations - Slight data manipulation - Low technical risk
3	Moderate complexity	- Requires backend/frontend interaction - Conditional logic - Complex read/write transactions with database
4	High complexity	- Multiple components involved - Integration with APIs - High performance in the database

Figure 2: Estimation metrics

## User Stories

Based on the priority rubrics and estimation metrics defined earlier, we developed a set of user stories for the AgroClima system to identify and address the primary needs of its users. The user stories are organized by user roles within the system: the Client, typically a farmer or crop owner; the Analyst, who processes, analyzes, and visualizes environmental data; and the Administrator, who manages user access and ensures proper and secure system operation.

### User Stories (Client)

<b>Title:</b> Register as a Client	<b>Priority:</b> High	<b>Estimate:</b> 3
<b>User Story:</b> As a client, I want to register using my email account so that I can access climate data and receive tailored recommendations.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"> <li>The system allows registration via email, Google, or Microsoft.</li> <li>Client roles are assigned automatically upon successful registration.</li> <li>A confirmation email is sent upon account creation.</li> </ul>		

Figure 3: User Story 1



<b>Title:</b> View Local Weather Conditions	<b>Priority:</b> Medium	<b>Estimate:</b> 4
<b>User Story:</b> As a client, I want to view real-time and historical weather data for my region so that I can plan activities effectively.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"> <li>• Weather data is displayed for the client's location.</li> <li>• Data includes temperature, humidity, precipitation, and forecast.</li> <li>• A time filter (e.g., last 7 days, 1 month) is available.</li> </ul>		

Figure 4: User Story 2

<b>Title:</b> Receive Weather Alerts	<b>Priority:</b> Medium	<b>Estimate:</b> 4
<b>User Story:</b> As a client, I want to receive automatic alerts for upcoming extreme weather events in my area so that I can take preventive actions.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"> <li>• Alerts are triggered based on predictive models.</li> <li>• Alerts are displayed in the user dashboard.</li> <li>• Each alert includes event type, severity and estimated time.</li> </ul>		

Figure 5: User Story 3

<b>Title:</b> Access Historical Recommendations	<b>Priority:</b> High	<b>Estimate:</b> 3
<b>User Story:</b>  As a client, I want to view past recommendations and actions taken so I can evaluate their effectiveness over time.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"> <li>• A history tab shows all past recommendations received.</li> <li>• Each entry includes date, context, and whether the advice was followed.</li> <li>• The user can filter history by date range or type.</li> </ul>		

Figure 6: User Story 4

<b>Title:</b> Receive Farming Recommendations	<b>Priority:</b> Low	<b>Estimate:</b> 4
<b>User Story:</b>  As a client, I want to receive personalized recommendations based on weather predictions and soil conditions so I can optimize my crop yield.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"> <li>• Recommendations are shown on the dashboard.</li> <li>• Data includes suggestions on irrigation, fertilization, and planting.</li> <li>• Alerts are sent for critical weather risks (e.g., frost, drought).</li> </ul>		

Figure 7: User Story 5

## User Stories (Analyst)

Title:	Priority:	Estimate:
Access and Analyze Climate Data	Medium	4
<b>User Story:</b>  As an analyst, I want to access real-time and historical climate data from all monitored stations so I can analyze environmental trends.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"><li>• A dashboard shows station data in real time.</li><li>• Filters allow data segmentation by location and date.</li></ul>		

Figure 8: User Story 6

Title:	Priority:	Estimate:
Generate Analytical Reports	Medium	3
<b>User Story:</b>  As an analyst, I want to generate reports about climate conditions and risk trends to support decision-making for stakeholders.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"><li>• Reports can be generated monthly or weekly.</li><li>• Data visualizations (charts, graphs) are included.</li></ul>		

Figure 9: User Story 7

## User Stories (Administrator)

Title:	Priority:	Estimate:
Manage Users and Roles	High	2
<b>User Story:</b>  As an administrator, I want to manage user accounts and assign roles so that access control is enforced throughout the platform.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"><li>• Admin dashboard lists all users with filters by role.</li><li>• Roles can be assigned or changed.</li><li>• User accounts can be deactivated or deleted.</li></ul>		

Figure 10: User Story 8

<b>Title:</b> Audit User Activity Logs	<b>Priority:</b> Low	<b>Estimate:</b> 4
<b>User Story:</b>  As an administrator, I want to access logs of user activities so I can ensure compliance and investigate anomalies.		
<b>Acceptance Criteria:</b> <ul style="list-style-type: none"> <li>• Logs include login times, data access, configuration changes.</li> <li>• Filters by user, date, and activity type are available.</li> <li>• Logs can be exported in CSV format.</li> </ul>		

Figure 11: User Story 9

# Initial Database Architecture

This section presents a conceptual design of the initial database architecture proposed for the AgroClima system. The system is intended to collect, process, store, and provide access to large volumes of data related to agricultural environments, including weather conditions and user activity. The proposed architecture is designed to be scalable, modular, and capable of supporting predictive analytics and real-time decision-making, while maintaining a clear separation of concerns across its components.

## High-Level Architecture Proposal

The architecture of the AgroClima platform is designed to meet the specific needs of a system focused on predicting climate-related agricultural risks. This requires the integration of diverse environmental data sources, the continuous processing of high-volume geospatial and time-series data, and the delivery of accurate and timely recommendations to a heterogeneous group of users, including farmers, analysts, and administrators.

This High-level architectural proposal is structured not around specific technologies, but around solving the core functional and non-functional requirements of the system. These include scalability, data integrity, low-latency access to forecasts, support for historical data analysis, and secure multi-user access.

### 1. Data Sources and Acquisition

The system must gather data from multiple heterogeneous sources, which vary in structure, frequency, and origin. These sources include:

- Real-time environmental measurements, such as temperature, humidity, and soil moisture from in-field sensors or public meteorological networks.
- Historical climate datasets, used for training predictive models and evaluating long-term patterns.
- External services, such as weather forecast APIs, satellite data repositories, or governmental platforms.

The data acquisition layer must support both continuous ingestion of real-time data and batch ingestion of historical or bulk data, ensuring flexibility in handling structured, semi-structured, and unstructured formats.

### 2. Data Ingestion and Preprocessing

To ensure consistency and reliability in downstream processes, all incoming data must be validated, cleaned, and transformed. The ingestion layer should support:

- The decoupling of data producers from processors, allowing asynchronous and scalable data flows.
- The detection and correction of anomalies or missing values before data enters the core system.

This layer serves as a critical buffer between raw data collection and structured data storage, facilitating quality control and enabling fault tolerance.

### 3. Data Storage Architecture

Given the volume, velocity, and variety of the data, the storage architecture must accommodate multiple data types and access patterns. The storage design is composed of two main components:

- A raw data repository that temporarily stores ingested data in its original format for traceability, reprocessing, or auditing purposes.
- A processed data store that holds structured information, supporting efficient querying.

The processed store should enable time-efficient access to both historical records and the latest readings.

#### 4. Data Processing and Analytics Layer

Once the data is curated and structured, it must be processed to extract actionable insights. This includes:

- Feature extraction and data enrichment, such as calculating vegetation indices, aggregating precipitation levels, or linking sensor readings to crop stages.
- Execution of predictive models, which estimate the likelihood and severity of climate events based on historical and current inputs.
- Triggering of alerts and recommendations, based on the outputs of the predictive algorithms.

This layer must support both real-time processing (for alerts and urgent decisions) and batch processing (for retrospective analysis and reporting).

#### 5. Application and Serving Layer

This layer is responsible for providing personalized, timely, and reliable access to the system's outputs. It serves:

- Visual dashboards for analysts, with access to aggregated data, prediction outputs, and historical comparisons.
- Advisory interfaces for farmers, focused on immediate recommendations, alert notifications, and field-specific information.
- Administrative tools for managing user roles, permissions, and system configurations.

The application layer must enforce role-based access control, ensure data security, and offer modular interfaces that adapt to different user profiles and usage scenarios.

#### 6. Access and Integration Interfaces

To allow external systems, applications, and users to interact with AgroClima, a set of structured interfaces is exposed through secure APIs. These interfaces must:

- Support modular access to data and services, depending on the user's role and needs.
- Enable integration with third-party systems, such as government platforms, mobile apps, or agricultural decision-support tools.

#### 7. Data Flow Overview

The overall flow of data through the system proceeds as follows:

- Acquisition: Data is collected from external APIs, sensors, and user uploads.
- Landing and Validation: Incoming data is stored temporarily and validated for structure and quality.
- Processing: Data is transformed, enriched, and passed through prediction and analysis pipelines.
- Storage: Structured outputs are stored in databases optimized for querying and access.
- Interaction: End users access insights via dashboards, alerts, or API endpoints, depending on their profile.

This flow ensures traceability, adaptability, and responsiveness throughout the system.

## ER Diagram (Initial Version)

The following section presents an initial version of the Entity-Relationship (ER) diagram for the AgroClima system. This diagram has been developed by applying a structured 10-step design methodology, which facilitates a systematic understanding of the domain and its key components. The ER model serves as a conceptual representation of the system's data structure, capturing the main entities, their attributes, and the relationships between them. This model provides a foundation for the logical and physical design of the database and ensures alignment between system requirements and data organization.

### 1. Components Definition

The first step consists of identifying the core components involved in the domain of the system. AgroClima focuses on collecting environmental data, processing it through prediction models, and generating actionable recommendations for agricultural users. Thus, the essential components are:

- System users
- Agricultural fields and cultivated crops
- Meteorological data
- Predictive models and generated risk events
- Alerts and recommendations for decision-making
- Geographical locations and external data sources

These components provide the foundation for defining the data entities required to represent and support the system's operations.

### 2. Entities Definition

Entities represent real-world objects or concepts that need to be stored and managed in the database. From the components above, we identify the following entities:

- **User**: Represents the individuals registered in the AgroClima platform. Users may have different roles such as farmers, analysts, or administrators and are the recipients of climate alerts and predictions.
- **Agriculture\_field**: Represents a physical agricultural field owned or managed by a user. Each field is geolocated and can be associated with one or more crops. This entity allows AgroClima to tailor predictions and recommendations to specific land plots.
- **Crop**: Stores the crops cultivated in each agricultural field. This allows the platform to provide crop-specific recommendations and monitor weather impacts on individual crop types.
- **Prediction**: Stores the results generated by machine learning models that predict extreme weather events that may affect specific locations.
- **Alert**: Stores records of alerts sent to users when a relevant climate event is predicted. Alerts are based on predictions generated by the system.
- **Recommendation**: Stores system-generated recommendations based on predictions and field conditions. These are delivered to farmers to support agricultural decisions such as irrigation, fertilization, or harvesting.
- **Weather\_Station**: Represents the weather stations (physical or virtual) that collect meteorological data used by the system to feed predictions.
- **Weather\_Data**: Stores individual weather readings collected by a weather station at a specific timestamp. These readings are key inputs for generating predictions.
- **Location**: Represents geographic locations associated with weather stations or prediction areas. Enables spatial organization of environmental data.

- ***Event\_type***: Defines the types of weather events that the system can predict.
- ***Severity\_level***: Categorizes the severity level of a predicted event, helping to prioritize alerts and responses.
- ***Model***: Contains metadata about the predictive models used by AgroClima, including versioning and training information.
- ***Data\_source***: Defines the external or internal sources of meteorological data used in the system. This allows tracking of data provenance and managing integration with third-party APIs.

Each entity will later be expanded with attributes and relationships.

### 3. Define Attributes per Entity

In this step, each entity is described by a set of attributes that define its properties:

- ***User***: user\_id, user\_name, email, role, password\_hash
- ***Agriculture\_field***: field\_id, field\_type
- ***Crop***: crop\_id, crop\_name
- ***Prediction***: prediction\_id, timestamp, probability
- ***Alert***: alert\_id, timestamp\_sent, status
- ***Recommendation***: recommend\_id, recommend\_description
- ***Weather\_Station***: station\_id, name
- ***Weather\_Data***: data\_id timestamp, temperature, humidity, wind\_speed
- ***Location***: location\_id, name, latitude, longitude, type
- ***Event\_type***: event\_type\_id, event\_name, event\_description
- ***Severity\_level***: severity\_level\_id, severity\_type, level\_description
- ***Model***: model\_id, model\_name, version, training\_timestamp
- ***Data\_source***: source\_id, name, type, url

Each attribute is assigned a data type and constraints in later stages.

### 4. Relationships Definition

The following table presents the relationship matrix among the identified entities in the AgroClima system. This matrix facilitates the systematic identification of associations between pairs of entities, helping to visualize how data elements interact within the system. A check mark indicates that there is a direct relationship between the corresponding entities, typically supported by a foreign key in the database schema.



	User	Field	Crop	Location	Station	Weather Data	Data Source	W-Data Source	Model	Prediction	Event Type	Severity Level	Alert	Recommendation
User	—	✓								✓			✓	
Agriculture Field	✓	—	✓	✓						✓				✓
Crop		✓	—											
Location		✓		—	✓					✓				
Weather Station				✓	—	✓	✓	✓						
Weather Data					✓	—								
Data Source					✓		—	✓						
Model									—	✓				
Prediction	✓	✓		✓					✓	—	✓	✓	✓	✓
Event Type										✓	—			
Severity Level										✓		—		
Alert	✓									✓			—	
Recommendation		✓								✓				—

Table 1: Relationship Matrix

## 5. Relationships Types Definition

Each identified relationship is classified by cardinality:

- **One-to-Many:**

- A user → many fields
- A user → many alerts
- A field → one crop
- A prediction → many alerts
- A prediction → many recommendations
- A station → many weather data entries

- **Many-to-One:**

- A weather station → one location
- A prediction → one location, one event type, one severity level, one model

- **Many-to-Many:**

- A weather station ↔ multiple data sources

This classification supports correct implementation of referential integrity and constraint rules.

## 6. First Entity-Relationship Model Draw

At this stage, an initial version of the Entity-Relationship (ER) model is constructed based on the previously defined entities, attributes, and relationships. This first diagram provides a conceptual visualization of the system's data structure, capturing the logical connections between entities. Although preliminary, it serves as a foundation for further refinement and normalization in subsequent steps.

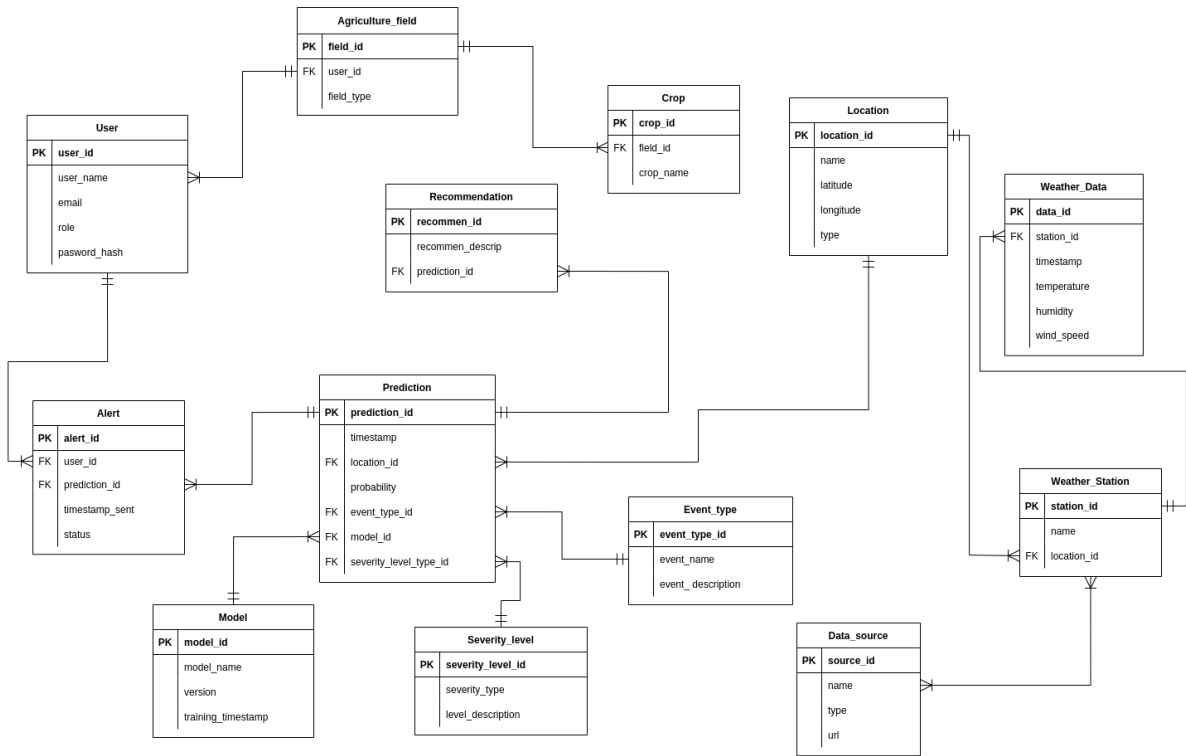


Figure 1: First ER Model Draw

## 7. Split Many-toMany Relationships

To comply with best practices in relational modeling, any many-to-many relationships are broken down. Specifically:

The relationship between weather stations and data sources is transformed into an associative entity, *Weather\_data\_source*, which includes:

- *station\_id* (FK)
- *source\_id* (FK)
- *start\_date* (attribute)

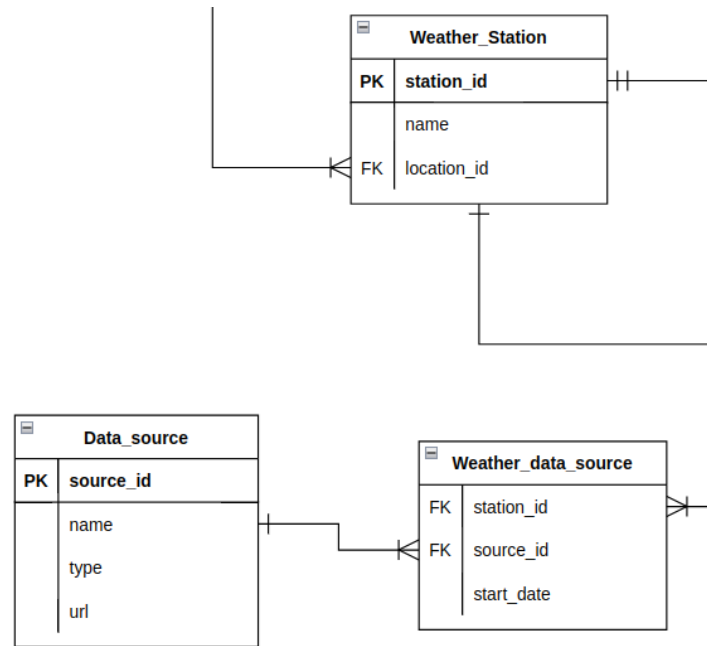


Figure 2: Many-to-Many relationship fix

This allows tracking the history of data provider associations for each station.

## 8. Second Entity-Relationship Model Draw

This step presents the refined version of the Entity-Relationship (ER) model, incorporating adjustments derived from the analysis of relationship types and attribute distribution. The updated diagram resolves any previously identified many-to-many relationships through associative entities and improves overall clarity and consistency.

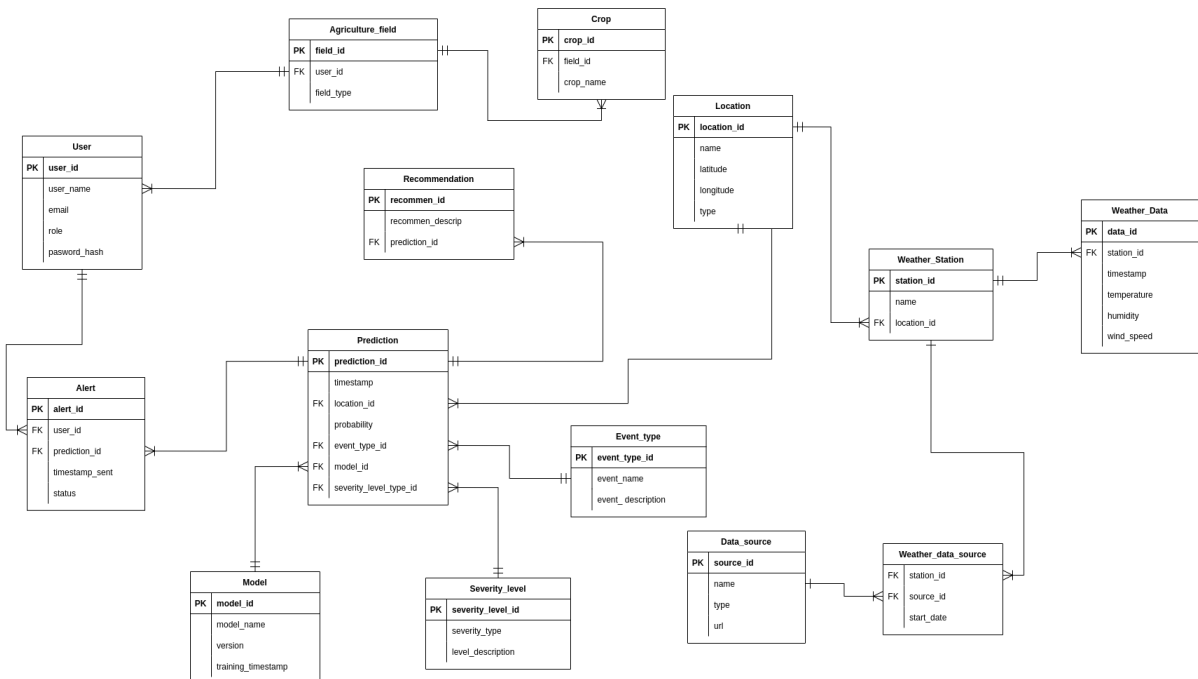


Figure 3: Second ER Model Draw

## 9. Get Data-Structure Entity-Relationship Model

In this step, the conceptual design of the system is translated into a structured data model suitable for implementation in a relational database management system. This data-structure ER model includes detailed definitions of data types and normalization refinements. It serves as a bridge between the abstract representation of the system and its physical realization, ensuring data consistency, integrity, and scalability.



Figure 4: Entities Data-structure

## 10. Final Initial ER Model Draw

The final step in the entity-relationship modeling process involves defining the constraints and structural properties that govern how data is stored, validated, and maintained within the database system.

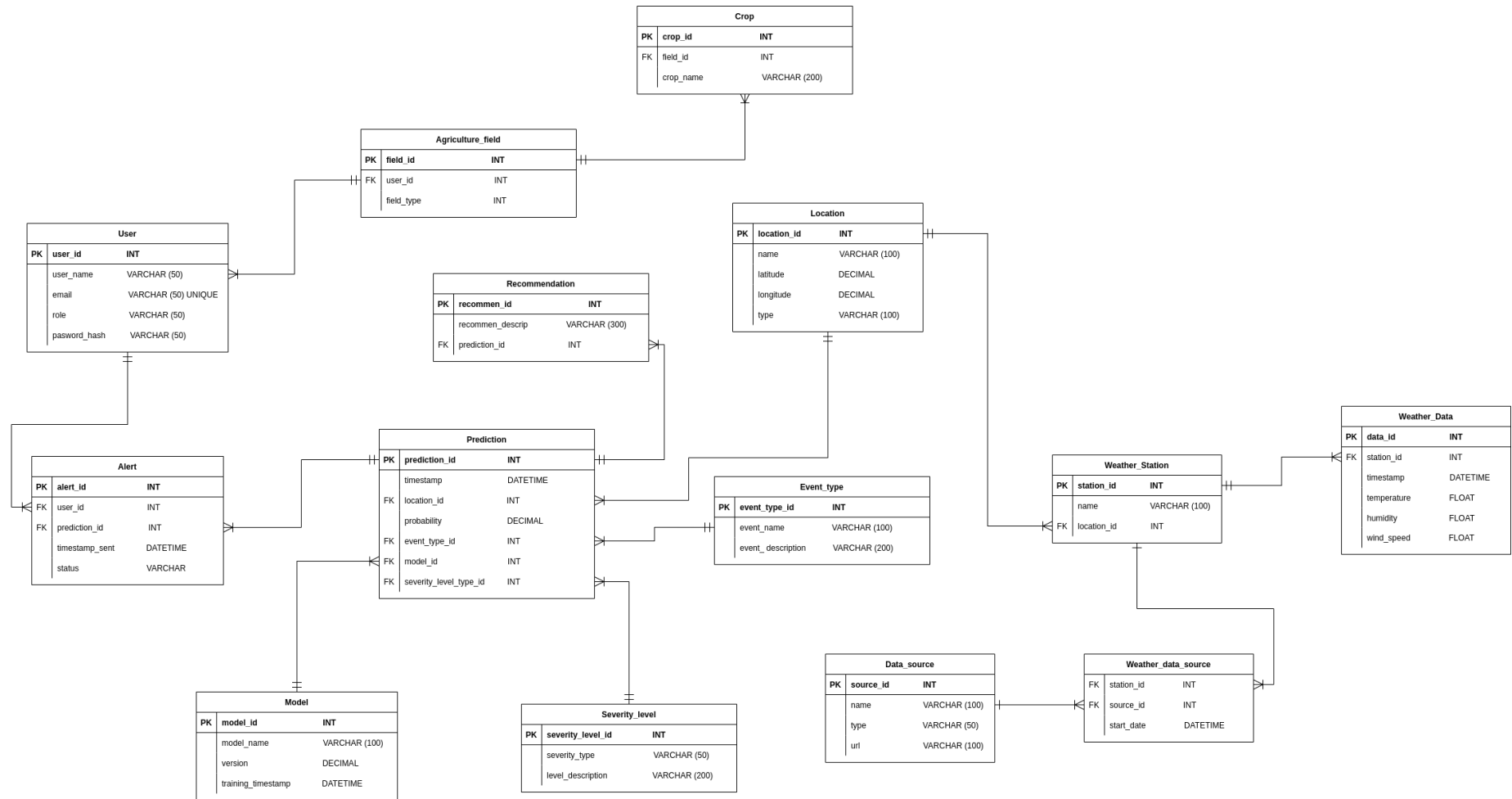


Figure 5: Initial ER Model Draw

## References

- 1 R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, 7th ed. Boston, MA, USA: Pearson, 2016.
- 2 C. A. Sierra, "INTRODUCTION TO DATABASES / Database Foundations," Lecture, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, 2024.