

# Enhancing Crop Yield Prediction with Non-Relational Database Solution.

## Introduction:

In this scenario, we will explore the challenges faced by a hypothetical agricultural research organization, Agri-Tech Research Institute, in predicting crop yields. The current system relies on traditional relational databases, but as the complexity and volume of agricultural data increase, there is a need for a more flexible and scalable solution. We will discuss why deploying a non-relational database solution would be the best option for addressing the specific difficulties encountered by Agri-Tech.

## Organization Overview:

The Agri-Tech Research Institute is committed to developing agricultural practices via research and development. To provide vital insights to farmers and policymakers, the organization collects huge volumes of data pertaining to soil conditions, weather patterns, crop types, and cultivation techniques. The current relational database struggles to handle the diverse and dynamic nature of agricultural data, hindering the accuracy and efficiency of crop yield predictions.

## Challenges:

### 1. Diverse Data Sources:

- **Issue:** Agri-Tech collects data from a variety of sources, including IoT sensors, satellite imaging, weather stations, and soil sensors. Each source generates data in a variety of forms and structures, making integration within a relational database more difficult.
- **Impact:** The relational model's fixed schema isn't easily adaptable to accommodate varying data types and structures. It requires significant data preprocessing, transformation, and normalization, leading to inefficiencies in data management and analysis.

### 2. Real-Time Data Processing:

- **Issue:** Access to real-time data in a timely manner is crucial for accurate crop yield predictions and prompt decision-making by farmers. The existing system, on the other hand, struggles to handle and evaluate data quickly.
- **Impact:** Delays in data processing hinder the system's ability to provide farmers with up-to-date, actionable insights. This lag affects the agility required for timely interventions in crop management.

### 3. Scalability:

- **Issue:** As precision farming and research projects expand, the volume of collected data grows significantly. The relational database encounters scalability issues, struggling to handle this increasing influx of data without sacrificing performance.

- **Impact:** Performance bottlenecks emerge as data volumes rise, impacting the system's responsiveness and query execution times. This restricts the system's ability to efficiently handle the growing data load.

#### **4. Flexible Data Model:**

- **Issue:** Agricultural data is very dynamic, with changing data requirements, fluctuating structures, and new forms of data being presented on a regular basis. However, the relational database's inflexible schema makes accommodating these modifications difficult.
- **Impact:** Any alterations or additions to the schema demand significant modifications to the database structure, causing disruptions in ongoing operations. This inflexibility impedes the swift adaptation required to incorporate new data variables or sources.

Addressing these challenges often requires a more adaptable and scalable database solution like MongoDB, which excels in handling diverse data formats, supporting real-time data processing, scaling seamlessly with growing data volumes, and offering a flexible schema that accommodates evolving data needs without disrupting ongoing operations.

## **Data Requirements:**

Data collection is instrumental in the scenario of enhancing crop yield prediction in agriculture, especially considering the challenges faced by the Agri-Tech Research Institute. Here's how data collection can be beneficial:

### **1. CROP DATA COLLECTION:**

**Data-Driven Insights:** Gathering data on crop types, planting dates, and harvest dates provides significant insights into crop growth patterns and life cycles. This information assists in making informed crop management decisions.

Farmers and Researchers can use precision farming techniques by gathering data on certain crops. This involves accurate irrigation, fertilization, and pesticide administration customized to each crop type's demands, resulting in resource efficiency.

### **2. IoT SENSOR DATA:**

**Real-Time Monitoring:**

**Continuous Data Stream:**

IoT sensors give real-time data on a variety of characteristics such as soil moisture, temperature, and humidity. This allows for continual monitoring of critical environmental parameters for crop growth.

**Irrigation with Accuracy:**

**Improved Water Management:** Soil moisture sensors aid in precisely determining soil moisture levels. This data is used to optimize irrigation operations, ensuring crops receive the appropriate amount of water and reducing water waste.

### 3. WEATHER DATA:

**Understanding Seasonal Variations:** Analyzing historical crop data allows for a better understanding of how different crops perform under varying weather conditions. This information can guide farmers in selecting the most suitable crops for specific seasons.

**Environmental Impact on Crops:**

**Temperature Variation:** Monitoring temperature helps in understanding how different crops respond to variations in temperature. This data is crucial for predicting growth patterns and identifying temperature-sensitive crops.

**Humidity Levels:** Humidity affects plant transpiration and disease susceptibility. By tracking humidity levels, farmers can implement measures to prevent excessive moisture-related issues.

**Crop Growth Modeling:**

**Machine Learning Models:** Historical weather data, combined with crop-specific information, can be used to train machine learning models. These models can predict crop growth patterns based on environmental conditions, enhancing the accuracy of crop yield predictions.

### 4. RESEARCH AND DEVELOPMENT:

**Data for Research Projects:** Researchers at the Agri-Tech Research Institute can use crop data to conduct studies on the performance of different crops. This aids in developing new cultivation techniques, identifying resilient crop varieties, and improving overall agricultural practices.

## Proposed Non-Relational Database Solution:

Document Database (MongoDB):

### Justification for MongoDB:

MongoDB is recommended for the following reasons:

**1. Versatile Data Model:** MongoDB's document-oriented structure fits the complex and ever-changing nature of agricultural data. It enables Agri-Tech to save data in a format that reflects the complexities of real-world agricultural processes.

**2. Real-Time Data Processing:** The in-memory storage engine in MongoDB improves real-time data processing capabilities. This functionality ensures that Agri-Tech can quickly evaluate data and provide farmers with current insights for informed decision-making.

**3. Scalability:** Because of MongoDB's horizontal scaling characteristics, it is well suited to handling the expanding volume of agricultural data. Agri-Tech can extend their research programs without concerns about database performance issues.

- 4. Flexibility for Research Projects:** Document databases, particularly MongoDB, allow Agri-Tech to adapt to changes in data requirements without disrupting ongoing research projects. This flexibility is crucial for accommodating new variables, sensors, and research methodologies.
- 5. Community Support and Ecosystem:** MongoDB has a vibrant community and a rich ecosystem of tools and connectors. This ensures that Agri-Tech can leverage a wide range of resources for the development, maintenance, and optimization of its crop yield prediction system.

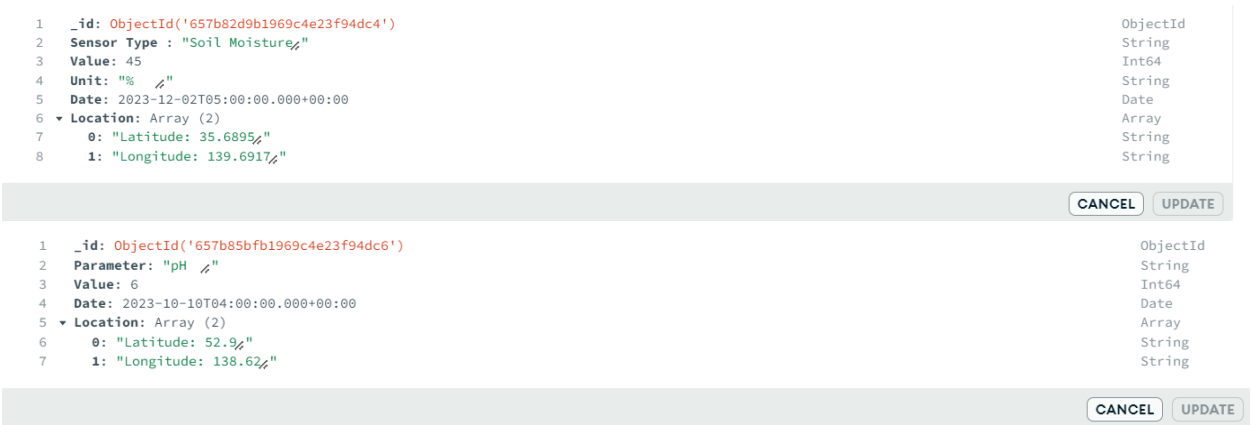
## Prototype Implementation:

I have developed a prototype to showcase MongoDB's capabilities in managing diverse agricultural data for crop yield predictions. The prototype will include data on soil conditions, weather patterns, crop types, and cultivation techniques. Screenshots of the prototype will be provided to demonstrate the functionality of the proposed solution.

### 1. IoT Sensor Data:

**Attributes:**

- `_id`: Unique Identifier for the IoT Sensor.
- `Sensor Type`: Type of sensor (Example: Soil Moisture, Temperature)
- `Value`: Sensor reading Value.
- `Unit`: Measurement unit (In this case Celsius)
- `Date`: Date of the sensor reading.
- `Location`: Geospatial Data (Latitude, Longitude)



1	_id: ObjectId('657b8655b1969c4e23f94dc7')	ObjectId
2	Parameter: "pH"	String
3	Value: 7	Int64
4	Date: 2023-10-13T04:00:00.000+00:00	Date
5	Location: Array (2)	Array
6	0: "Latitude: 57.65"	String
7	1: "Longitude: 198.7"	String

CANCEL UPDATE

1	_id: ObjectId('657b8521b1969c4e23f94dc5')	ObjectId
2	Sensor Type: "Soil Moisture"	String
3	Value: 35	Int64
4	Unit: "%"	String
5	Date: 2023-11-11T05:00:00.000+00:00	Date
6	Location: Array (2)	Array
7	0: "Latitude: 67.89"	String
8	1: "Longitude: 176.54"	String

CANCEL UPDATE

## 2. Weather Data Collection:

### Attributes:

- Sensor type: Type of weather sensor (e.g., temperature, humidity).
- value: Sensor reading value.
- units: Measurement unit (e.g., °C, %).
- Date: Date of the weather reading.
- location: Geospatial data (latitude and longitude).

1	_id: ObjectId('657b8730b1969c4e23f94dc9')	ObjectId
2	Sensor Type: "Temperature"	String
3	Value: 28	Int64
4	Unit: "celsius"	String
5	Location: Array (2)	Array
6	0: "Latitude: 34.9"	String
7	1: "Longitude: 37.9"	String

CANCEL UPDATE

1	_id: ObjectId('657b87d7b1969c4e23f94dca')	ObjectId
2	Sensor Type: "Temperature"	String
3	Value: 42	Int64
4	Unit: "Celsius"	String
5	Location: Array (2)	Array
6	0: "Latitude: 43.86"	String
7	1: "Longitude: 176.98"	String

CANCEL UPDATE

1	_id: ObjectId('657b8851b1969c4e23f94dcb')	ObjectId
2	Sensor Type: "Humidity"	String
3	Value: 35	Int64
4	Location: Array (2)	Array
5	0: "Latitude: 47.98"	String
6	1: "Longitude: 156.6"	String

CANCEL UPDATE

1	_id: ObjectId('657b88d4b1969c4e23f94dcc')	ObjectId
2	Sensor Type : "Rainfall"	String
3	Value: 5	Int64
4	Location: Array (2)	Array
5	0: "Latitude: 76.39"	String
6	1: "Longitude: 198.34"	String
7	Units: "mm"	String

CANCEL UPDATE

1	_id: ObjectId('657b9464b1969c4e23f94e0a')	ObjectId
2	Temperature: 28	Int64
3	Humidity: 45	Int64
4	Precipitation: 5	Int64

CANCEL UPDATE

### 3. Crop Data Collection:

#### Attributes:

- crop type: Type of crop.
- planting date: Date when the crop was planted.
- harvest date: Date when the crop was harvested.
- sensor data: References to relevant IoT sensor and weather data.

Here, I have given Sensor data Id and Weather data Id as it gives the details about the Soil moisture, pH and temperature, Humidity of a particular crop.

1	_id: ObjectId('657b8a6ab1969c4e23f94dd0')	ObjectId
2	Planting Date: 2023-10-10T04:00:00.000+00:00	Date
3	Harvest Date: 2023-12-10T05:00:00.000+00:00	Date
4	Sensor Data ID: 657b82d9b1969c4e23f94dc4	ObjectId
5	Weather data ID: 657b8730b1969c4e23f94dc9	ObjectId
6	▼ Cultivation Techniques: Array (3)	Array
7	0: "Irrigation: Sprinkler Irrigation"	String
8	1: "Fertilization: Synthetic Fertilizer"	String
9	2: "Pest control: Chemical pest control"	String
10	Crop: "Wheat"	String
<div>CANCEL UPDATE</div>		
1	_id: ObjectId('657b8c39b1969c4e23f94de3')	ObjectId
2	Planting Date: 2023-07-08T04:00:00.000+00:00	Date
3	Harvest Date: 2023-12-08T05:00:00.000+00:00	Date
4	Sensor Data ID: 657b8521b1969c4e23f94dc5	ObjectId
5	Weather Data ID: 657b87d7b1969c4e23f94dca	ObjectId
6	▼ Cultivation Techniques: Array (3)	Array
7	0: "Irrigation: Drip Irrigation"	String
8	1: "Fertilization: Chemical Fertilizer"	String
9	2: "Pest control: Perimethrin"	String
10	Crop Type: "Corn"	String
<div>CANCEL UPDATE</div>		
1	_id: ObjectId('657b8cf8b1969c4e23f94de8')	ObjectId
2	Planting Date: 2023-06-12T04:00:00.000+00:00	Date
3	Harvest Date: 2023-12-09T05:00:00.000+00:00	Date
4	Sensor Data ID: 657b85bfb1969c4e23f94dc6	ObjectId
5	Weather Data ID: 657b8851b1969c4e23f94dcb	ObjectId
6	▼ Cultivation Techniques: Array (3)	Array
7	0: "Irrigation: Flooded Irrigation"	String
8	1: "Fertilizer: Chemical Fertilizer"	String
9	2: "Pest Control: Integrated Pest Management"	String
10	Crop Type: "Maize"	String
<div>CANCEL UPDATE</div>		
1	_id: ObjectId('657b8e37b1969c4e23f94ded')	ObjectId
2	Crop Type : "Sugar Cane"	String
3	Planting Date: 2023-07-16T04:00:00.000+00:00	Date
4	Harvest Date: 2023-12-05T05:00:00.000+00:00	Date
5	Sensor Data ID: 657b865b1969c4e23f94dc7	ObjectId
6	Weather Data ID: 657b88d4b1969c4e23f94dcc	ObjectId
7	▼ "cultivation_techniques": : Array (3)	Array
8	0: "Irrigation : Drip Irrigation"	String
9	1: "Fertilization : Organic Fertilizer"	String
10	2: "Pest Control: Integrated Pest Management"	String
<div>CANCEL UPDATE</div>		

In the below provided Screenshots,

The Weather document (ObjectId("657b9464b1969c4e23f94e0a")) is referenced by both Field documents representing different agricultural fields and crops.

This referencing mechanism allows us to maintain a single source of weather data while associating it with multiple fields or crops.

If there are updates to weather information, they can be made in one place (the Weather document) and automatically reflect in all documents referencing it.

This approach supports data normalization, reduces redundancy, and ensures consistency across documents referencing shared information

1    _id: ObjectId('657b94d6b1969c4e23f94e0c')	ObjectId
2    Planting Date: 2023-06-11T04:00:00.000+00:00	Date
3    Harvest Date: 2023-10-10T04:00:00.000+00:00	Date
4    Weather Data ID : 657b9464b1969c4e23f94e0a	ObjectId
5    ▼ Cultivation Techniques: Array (3)	Array
6       0: "Irrigation: Sprinkler Irrigation"	String
7       1: "Fertilizer: Organic Fertilizer"	String
8       2: "Pest Control: Chemical Pest Control"	String
9    Crop Type: "Rice"	String

1    _id: ObjectId('657b9616b1969c4e23f94e0f')	ObjectId
2    Planting Date: 2023-08-08T04:00:00.000+00:00	Date
3    Harvest Date: 2023-12-10T05:00:00.000+00:00	Date
4    ▼ Cultivation Techniques: Array (3)	Array
5       0: "Irrigation : Drip Irrigation"	String
6       1: "Fertilizer: Chemical Fertilizer"	String
7       2: "Pest Control: Integrated Pest Management"	String
8    Weather Data ID: 657b9464b1969c4e23f94e0a	ObjectId
9    Crop Type: "Millets"	String

## 4. Research Project Data Collection:

### Attributes:

researcher: Name of the researcher leading the project.

data\_types: Types of data used in the project (e.g., soil moisture, temperature)

Here, I've used Crop data Id since it provides detailed information on the crop, such as the crop's name, planting date, harvest date, and cultivation practices. If someone wants to know the data about a specific crop, they should seek up the crop data Id so that they can learn about all the qualities of a crop and delve deep into the soil moisture and temperature of the crop by using the weather data id and sensor data id they are looking for.

1	<code>_id: ObjectId('657b8fb4b1969c4e23f94dfb')</code>	ObjectId
2	<code>Researcher: "Niharika"</code>	String
3	<code>Data Types: "Soil Moisture, Temperature"</code>	String
4	<code>Crop Data ID: 657b8a6ab1969c4e23f94dd0</code>	ObjectId
		<input type="button" value="CANCEL"/> <input type="button" value="UPDATE"/>
1	<code>_id: ObjectId('657b905eb1969c4e23f94dfc')</code>	ObjectId
2	<code>Researcher: "Kavya"</code>	String
3	<code>Data Type: "Soil Moisture, Temperature"</code>	String
4	<code>Crop Data ID: 657b8c39b1969c4e23f94de3</code>	ObjectId
		<input type="button" value="CANCEL"/> <input type="button" value="UPDATE"/>
1	<code>_id: ObjectId('657b9097b1969c4e23f94dfd')</code>	ObjectId
2	<code>Researcher: "M.S. Swaminathan"</code>	String
3	<code>Data Type: "pH, Humidity"</code>	String
4	<code>Crop Data ID: 657b8cf8b1969c4e23f94de8</code>	ObjectId
		<input type="button" value="CANCEL"/> <input type="button" value="UPDATE"/>
1	<code>_id: ObjectId('657b9126b1969c4e23f94dfe')</code>	ObjectId
2	<code>Researcher: "Dr, Anuj"</code>	String
3	<code>Data Type: "pH, Humidity"</code>	String
4	<code>Crop Data ID: 657b8e37b1969c4e23f94ded</code>	ObjectId
		<input type="button" value="CANCEL"/> <input type="button" value="UPDATE"/>

## Conclusion:

In conclusion, the hypothetical scenario of Agri-Tech Research Institute highlights the challenges faced by organizations in the agriculture industry in managing and analyzing diverse and dynamic data for accurate crop yield predictions. The adoption of a non-relational database solution, specifically MongoDB, is justified based on its flexible data model, real-time data processing capabilities, scalability, and flexibility for research projects. The prototype implementation will provide a visual representation of how MongoDB can address the specific information needs of Agri-Tech and enhance the accuracy and efficiency of its crop yield prediction system.

## Citations:

- <https://www.mongodb.com/products/system-integrators/infosys>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7070948/>
- <https://www.mdpi.com/2071-1050/14/23/16131>
- [https://www.researchgate.net/publication/333477491\\_Designing\\_and\\_Implementing\\_Data\\_Warehouse\\_for\\_Agricultural\\_Big\\_Data](https://www.researchgate.net/publication/333477491_Designing_and_Implementing_Data_Warehouse_for_Agricultural_Big_Data)



- [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&q=mongodb+in+agriculture&oq=#d=gs\\_qabs&t=1702537533349&u=%23p%3DJ5In-oPx98QJ](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=mongodb+in+agriculture&oq=#d=gs_qabs&t=1702537533349&u=%23p%3DJ5In-oPx98QJ)
- [https://scholar.google.com/scholar?hl=en&as\\_sdt=0%2C5&q=mongodb+uses+in+agriculture&btnG=#d=gs\\_qabs&t=1702538028914&u=%23p%3DN8xveH65AagJ](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=mongodb+uses+in+agriculture&btnG=#d=gs_qabs&t=1702538028914&u=%23p%3DN8xveH65AagJ)
- <https://www.semanticscholar.org/paper/An-Efficient-Data-Warehouse-for-Crop-Yield-Ngo-Le-Khac/d2aa0548d1709dd524bdbb66ec3a0b99e2f88f3e>
- [https://www.researchgate.net/publication/326145036\\_An\\_Efficient\\_Data\\_Warehouse\\_for\\_Crop\\_Yield\\_Prediction](https://www.researchgate.net/publication/326145036_An_Efficient_Data_Warehouse_for_Crop_Yield_Prediction)