# **Data Model Documentation**

*Data in the database is broken into 3 parts: Raw data, Cleaned data, Dimension and fact tables.*

*Below are the schemas of all the tables.*

## **RAW DATA SCHEMA:**

This schema contains data directly sourced from the sensors repository. It contains raw unprocessed data.

### Tables in this schema:

1. CROPDATA Table:
   * Stores data related to crop information and yields (unfiltered)
2. LOCATIONDATA Table:
   * Contains information about sensor locations (unfiltered).
3. PESTDATA Table:
   * Stores data related to pests affecting crops (unfiltered).
4. SENSORDATA Table:
   * Contains sensor data, including temperature, humidity, soil moisture, and more (unfiltered).
5. SOILDATA Table:
   * Stores soil-related data such as composition, pH, and nutrient levels (unfiltered).
6. WEATHERDATA Table:
   * Contains weather data, including conditions, wind speed, and precipitation (unfiltered).

### RAW DATA DICTIONARY:

* CROPDATA Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | VARCHAR(16777216) | Timestamp of crop data |
| CROP\_TYPE | VARCHAR(16777216) | Type of crop |
| CROP\_YIELD | VARCHAR(16777216) | Crop yield in some unit |
| GROWTH\_STAGE | VARCHAR(16777216) | Growth stage of the crop |
| PEST\_ISSUE | VARCHAR(16777216) | Pest issue affecting crop |

* LOCATIONDATA Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| SENSOR\_ID | VARCHAR(16777216) | Sensor identifier |
| LOCATION\_NAME | VARCHAR(16777216) | Name of the location |
| LATITUDE | VARCHAR(16777216) | Latitude coordinates |
| LONGITUDE | VARCHAR(16777216) | Longitude coordinates |
| ELEVATION | VARCHAR(16777216) | Elevation above sea level |
| REGION | VARCHAR(16777216) | Geographic region of location |

* PESTDATA Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | VARCHAR(16777216) | Timestamp of pest data |
| PEST\_TYPE | VARCHAR(16777216) | Type of pest |
| PEST\_DESCRIPTION | VARCHAR(16777216) | Description of the pest |
| PEST\_SEVERITY | VARCHAR(16777216) | Severity of the pest issue |

* SENSORDATA Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| SENSOR\_ID | VARCHAR(16777216) | Sensor identifier |
| TIMESTAMP | VARCHAR(16777216) | Timestamp of sensor data |
| TEMPERATURE | VARCHAR(16777216) | Temperature reading |
| HUMIDITY | VARCHAR(16777216) | Humidity reading |
| SOIL\_MOISTURE | VARCHAR(16777216) | Soil moisture reading |
| LIGHT\_INTENSITY | VARCHAR(16777216) | Light intensity reading |
| BATTERY\_LEVEL | VARCHAR(16777216) | Battery level of the sensor |

* SOILDATA Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | VARCHAR(16777216) | Timestamp of soil data |
| SOIL\_COMP | VARCHAR(16777216) | Soil composition reading |
| SOIL\_MOISTURE | VARCHAR(16777216) | Soil moisture reading |
| SOIL\_PH | VARCHAR(16777216) | Soil pH reading |
| NITROGEN\_LEVEL | VARCHAR(16777216) | Nitrogen level reading |
| PHOSPHORUS\_LEVEL | VARCHAR(16777216) | Phosphorus level reading |
| ORGANIC\_MATTER | VARCHAR(16777216) | Organic matter content |

* WEATHERDATA Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | VARCHAR(16777216) | Timestamp of weather data |
| WEATHER\_CONDITION | VARCHAR(16777216) | Weather condition |
| WIND\_SPEED | VARCHAR(16777216) | Wind speed |
| PRECIPITATION | VARCHAR(16777216) | Precipitation amount |

## **CLEANED DATA SCHEMA:**

This schema contains data processed in our python pipeline and can be safely used.

### Tables in this schema:

1. CROPCLEAN Table:
   * Stores data related to crop information and yields.
2. LOCATIONCLEAN Table:
   * Contains information about sensor locations.
3. PESTCLEAN Table:
   * Stores data related to pests affecting crops.
4. SENSORCLEAN Table:
   * Contains sensor data, including temperature, humidity, soil moisture, and more.
5. SOILCLEAN Table:
   * Stores soil-related data such as composition, pH, and nutrient levels.
6. WEATHERCLEAN Table:
   * Contains weather data, including conditions, wind speed, and precipitation.

### CLEANED DATA DICTIONARY:

* CROPCLEAN Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of crop data |
| CROP\_TYPE | VARCHAR(20) | Type of crop |
| CROP\_YIELD | FLOAT | Crop yield in some unit |
| GROWTH\_STAGE | VARCHAR(20) | Growth stage of the crop |
| PEST\_ISSUE | VARCHAR(20) | Pest issue affecting crop |

* LOCATIONCLEAN Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| SENSOR\_ID | VARCHAR(30) | Sensor identifier |
| LOCATION\_NAME | VARCHAR(30) | Name of the location |
| LATITUDE | NUMBER(8,6) | Latitude coordinates |
| LONGITUDE | NUMBER(9,6) | Longitude coordinates |
| ELEVATION | NUMBER(8,2) | Elevation above sea level |
| REGION | VARCHAR(20) | Geographic region of location |

* PESTCLEAN Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of pest data |
| PEST\_TYPE | VARCHAR(20) | Type of pest |
| PEST\_DESCRIPTION | VARCHAR(16777216) | Description of the pest |
| PEST\_SEVERITY | VARCHAR(20) | Severity of the pest issue |

* SENSORCLEAN Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| SENSOR\_ID | VARCHAR(10) | Sensor identifier |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of sensor data |
| TEMPERATURE | NUMBER(5,2) | Temperature reading |
| HUMIDITY | NUMBER(5,2) | Humidity reading |
| SOIL\_MOISTURE | NUMBER(5,2) | Soil moisture reading |
| LIGHT\_INTENSITY | NUMBER(8,2) | Light intensity reading |
| BATTERY\_LEVEL | NUMBER(5,2) | Battery level of the sensor |

* SOILCLEAN Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of soil data |
| SOIL\_COMP | NUMBER(5,2) | Soil composition reading |
| SOIL\_MOISTURE | NUMBER(5,2) | Soil moisture reading |
| SOIL\_PH | NUMBER(5,2) | Soil pH reading |
| NITROGEN\_LEVEL | NUMBER(5,2) | Nitrogen level reading |
| PHOSPHORUS\_LEVEL | NUMBER(5,2) | Phosphorus level reading |
| ORGANIC\_MATTER | NUMBER(5,2) | Organic matter content |

* WEATHERCLEAN Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of weather data |
| WEATHER\_CONDITION | VARCHAR(30) | Weather condition |
| WIND\_SPEED | NUMBER(5,2) | Wind speed |
| PRECIPITATION | NUMBER(5,2) | Precipitation amount |

## **Dimension and Facts SCHEMA:**

This schema contains dimensions and facts tables extracted from the clean data.

### Tables in this schema:

1. CROP\_DIMENSION Table:
   * Stores data related to crop dimension.
2. LOCATION\_DIMENSION Table:
   * Contains information about location dimension.
3. PEST\_DIMENSION Table:
   * Stores data related to pests dimension.
4. SENSOR\_DIMENSION Table:
   * Contains information about sensor dimension.
5. SOIL\_DIMENSION Table:
   * Contains information about soil dimension.
6. WEATHER\_DIMENSION Table:
   * Contains information about weather dimension.

### DIMENSION AND FACT, DATA DICTIONARY:

* CROP\_DIMENSION Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| CROP\_TYPE | VARCHAR(20) | Type of crop |
| GROWTH\_STAGE | VARCHAR(20) | Growth stage of the crop |
| PEST\_ISSUE | VARCHAR(20) | Pest issue affecting crop |
| primary key (CROP\_TYPE) | | |

* LOCATION\_DIMENSION Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| SENSOR\_ID | VARCHAR(30) | Sensor identifier |
| LOCATION\_NAME | VARCHAR(30) | Name of the location |
| LATITUDE | NUMBER(8,6) | Latitude coordinates |
| LONGITUDE | NUMBER(9,6) | Longitude coordinates |
| ELEVATION | NUMBER(8,2) | Elevation above sea level |
| REGION | VARCHAR(20) | Geographic region of location |
| primary key (SENSOR\_ID) | | |

* PEST\_DIMENSION Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| PEST\_TYPE | VARCHAR(20) | Type of pest |
| PEST\_DESCRIPTION | VARCHAR(16777216) | Description of the pest |
| PEST\_SEVERITY | VARCHAR(20) | Severity of the pest issue |
| primary key (PEST\_TYPE) | | |

* SENSOR\_DIMENSION Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| SENSOR\_ID | VARCHAR(10) | Sensor identifier |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of sensor data |
| TEMPERATURE | NUMBER(5,2) | Temperature reading |
| HUMIDITY | NUMBER(5,2) | Humidity reading |
| SOIL\_MOISTURE | NUMBER(5,2) | Soil moisture reading |
| LIGHT\_INTENSITY | NUMBER(8,2) | Light intensity reading |
| BATTERY\_LEVEL | NUMBER(5,2) | Battery level of the sensor |
| primary key (SENSOR\_ID) | | |

* SOIL\_DIMENSION Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of soil data |
| SOIL\_COMP | NUMBER(5,2) | Soil composition reading |
| SOIL\_MOISTURE | NUMBER(5,2) | Soil moisture reading |
| SOIL\_PH | NUMBER(5,2) | Soil pH reading |
| NITROGEN\_LEVEL | NUMBER(5,2) | Nitrogen level reading |
| PHOSPHORUS\_LEVEL | NUMBER(5,2) | Phosphorus level reading |
| ORGANIC\_MATTER | NUMBER(5,2) | Organic matter content |
| primary key (TIMESTAMP) | | |

* WEATHER\_DIMENSION Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of weather data |
| WEATHER\_CONDITION | VARCHAR(30) | Weather condition |
| WIND\_SPEED | NUMBER(5,2) | Wind speed |
| PRECIPITATION | NUMBER(5,2) | Precipitation amount |
| primary key (TIMESTAMP) | | |

* CROP\_YIELD Table:

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| TIMESTAMP | TIMESTAMP\_NTZ(9) | Timestamp of crop data |
| CROP\_TYPE | VARCHAR(20) | Type of crop |
| SENSOR\_ID | VARCHAR(10) | Sensor identifier |
| CROP\_YIELD | FLOAT | Crop yield in some unit |
| primary key (TIMESTAMP) | | |
| foreign key (CROP\_TYPE) references DFA23RAWDATA.VANGUARD.CROP\_DIMENSION(CROP\_TYPE) | | |
| foreign key (SENSOR\_ID) references DFA23RAWDATA.VANGUARD.SENSOR\_DIMENSION(SENSOR\_ID) | | |