# MySQL Database Table Documentation

## farmer

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| start\_date | DATE NOT NULL | Date at which data was collected |
| farmer\_id | INT(AUTO\_INCREMENT) | Primary key, unique identifier. |
| first\_name | VARCHAR(100) | Farmer's first name. |
| last\_name | VARCHAR(100) | Farmer's last name. |
| nationality | VARCHAR(50) | Nationality of the farmer. |
| sex | CHAR(1) | Gender of the farmer. |
| dob | DATE | Date of birth. |
| id\_type | VARCHAR(50) | Type of identification document. |
| id\_number | VARCHAR(100) | Identification number. |
| id\_doc\_number | VARCHAR(50) | Additional document identifier. |
| id\_image\_url | VARCHAR(255) | Link to the ID image. |
| farmer\_image\_url | VARCHAR(255) | Link to the farmer image |
| marital\_status | ENUM | Marital status (e.g., Single, Married). |
| contact\_number | VARCHAR(20) | Primary contact number. |
| mobile\_money\_number | VARCHAR(20) | Mobile money account number. |
| is\_mobile\_money\_same | BOOLEAN | Indicates if mobile money number is same as contact number (default TRUE). |
| association | VARCHAR(100) | Affiliation or cooperative association. |
| location | VARCHAR(100) | Location or address of the farmer. |
| global\_gap\_certified | ENUM ('Yes','No but interested','No and not interested') | Certification status for Global GAP. |
| usda\_organic\_certified | ENUM ('Yes','No but interested','No and not interested') | Certification status for USDA Organic. |

## farm

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| farm\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farmer\_id`. |
| farm\_number | INT | Unique farm number. |
| acreage | DECIMAL(10,2) | Acreage of farm |
| farm\_location | VARCHAR(100) | Description of the farm's location. |
| geo\_coordinates | POINT | Central geographic coordinates of the farm. |
| geo\_boundaries | POLYGON | Polygon describing the farm boundaries. |
| longitude | DECIMAL(10,7) | Longitude coordinates of farm |
| latitude | DECIMAL(10,7) | Latitude coordinates of farm |
| road\_accessibility | ENUM ('accessible', 'difficult', 'not accessible') | Accessibility level (e.g., accessible, difficult). |

## intercropping\_data

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| crop\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farmer\_id`. |
| intercropped\_crops | TEXT | List of intercrops (e.g., Okra, Cocoa). |

## dependents

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| dependent\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farmer\_id`. |
| age\_range | VARCHAR(20) | Age range of dependents (e.g., 0-5). |
| num\_dependents | INT | Number of dependents. |

## farm\_inputs

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| input\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farmer\_id`. |
| input\_type | VARCHAR(50) | Type of input (e.g., Fertilizer). |
| details | VARCHAR(255) | Specific details about the input. |

## impact\_assessment

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| assessment\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farm\_id`. |
| disability | BOOLEAN | Indicates if the farmer has a disability. |
| farming\_income\_type | ENUM ('Primary', 'Secondary') | Farming's role in income (e.g., Primary). |
| primary\_income\_source | VARCHAR(255) | Description of income source if farming is secondary. |
| post\_harvest\_loss\_percent | INT | Post-harvest losses due to spoilage, crops being uneaten, etc. |
| relies\_on\_rainfall | ENUM(Yes, No) | Indicates reliance on rainfall. |
| other\_irrigation\_methods | TEXT | Description of other irrigation methods. |

## main\_crop\_data

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| crop\_data\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farmer\_id` |
| farm\_id | INT | Foreign key referencing `farm.farm\_id`. |
| crop\_name | VARCHAR(50) | Name of the main crop. |
| variety | VARCHAR(100) | Variety of crop |
| flowering\_image\_url | VARCHAR(255) | Link to flowering stage image. |
| avg\_fruit\_set\_percentage | DECIMAL(5,2) | Average fruit set percentage of sampled trees |
| avg\_fruit\_count | INT | Average fruit count of sampled trees |
| initial\_planting\_year | YEAR | Year trees were planted on each farm |
| tree\_spacing | VARCHAR(50) | Tree spacing in feet or meters (e.g. 10x10 m, 6x6 ft) |
| tree\_flower\_percentage | INT | Aggregated percentage of total flowers on each tree |
| predicted\_harvest\_time | VARCHAR(255) | Predicted harvest timing |

## Harvesting

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| harvesting\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farmer\_id | INT | Foreign key referencing `farmer.farmer\_id`. |
| collected\_or\_harvested | ENUM(collected, harvested) | Indicates if the crop was collected/harvested. |
| harvesting\_process | VARCHAR(100) | Description of the process (e.g., Hand-picking). |
| harvest\_responsibility | VARCHAR(100) | Party responsible for harvesting. |
| laborers\_count | INT | Number of laborers involved. |

## tree\_data

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| tree\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farm\_id | INT | Foreign key referencing `farm.farm\_id`. |
| picture\_date | DATE | Date at which tree photo was taken. |
| tree\_number | INT | Unique identifier for the tree. |
| tree\_geopoint | POINT | Geographical point of the tree. |
| picture\_1\_url | VARCHAR(255) | URL for the first picture of the tree. |
| picture\_2\_url | VARCHAR(255) | URL for the second picture of the tree. |
| picture\_3\_url | VARCHAR(255) | URL for the third picture of the tree. |
| picture\_4\_url | VARCHAR(255) | URL for the fourth picture of the tree. |
| picture\_5\_url | VARCHAR(255) | URL for the fifth picture of the tree. |
| picture\_6\_url | VARCHAR(255) | URL for the sixth picture of the tree. |
| picture\_7\_url | VARCHAR(255) | URL for the seventh picture of the tree. |
| picture\_8\_url | VARCHAR(255) | URL for the eighth picture of the tree. |
| fruit\_count | INT | Number of fruits counted by object detection model |

## Yield and Revenue

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| yield\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| farm\_id | INT | Foreign key referencing `farm.farm\_id`. |
| yield\_date | DATE | Date yield was recorded |
| season | ENUM(Major, Minor) | Season for which yield was recorded |
| initial\_yield\_mt | DECIMAL(10,2) | Initial yield before losses were incurred (MT) |
| total\_yield\_mt | DECIMAL(10,2) | Total yield of farm after losses (MT) |
| loss\_percentage | DECIMAL(10,2) | Percentage loss recorded before total yield |
| tonnage\_sold | DECIMAL(10,2) | Tonnage of crop sold out of total yield (MT) |
| percentage\_sold | DECIMAL(10,2) | Percentage of tonnage sold out of total yield |
| total\_production\_cost\_ghs | DECIMAL(10,2) | Total production cost in GHS for total acreage |
| price\_per\_kilo\_ghs | INT | Price per kilogram of crop |
| revenue\_from\_sales\_ghs | DECIMAL(10,2) | Revenue from tonnage sold (GHS) |
| income\_lost\_ghs | DECIMAL(10,2) | Income lost from losses |

## Users

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| project\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| project\_name | VARCHAR(255) | Name of project |
| crop\_name | VARCHAR(100) | Name of crop |
| product\_type | VARCHAR(50) | Type of product (Fresh, Flour, Dried, Chips) |
| quantity\_required | DECIMAL(10,2) | Quantity of crops ordered |
| quantity\_unit | VARCHAR(20) | Unit in which crop quantity was ordered (Metric Tons, Kilograms, Pounds) |
| harvest\_time\_start | DATE | Date the user is expecting harvest to start |
| harvest\_time\_end | DATE | Date the user is expecting the harvest to end |
| created\_by | INT | User ID of user who created the project |
| created\_at | TIMESTAMP | Current timestamp of project creation |
| status | VARCHAR(20) | Status of project (active, completed, cancelled) |
| quality\_grade | VARCHAR(50) | Quality grade of crop sourced |
| notes | TEXT | Notes taken about the project |

## Project

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| project\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| project\_name | VARCHAR(255) | Name the user assigns to the created project |
| crop\_name | VARCHAR(100) | Name of crop involved in the project |
| product\_type | VARCHAR(50) | Type of product (Fresh, Flour, Dried, Chips) |
| quantity\_required | DECIMAL(10,2) | Quantity of product/crop required |
| quantity\_unit | VARCHAR(20) | Unit of quantity required (Metric tons, kilograms, pounds) |
| harvest\_time\_start | DATE | Date of harvest start selected by the user |
| harvest\_time\_end | DATE | Date of harvest end selected by the user |
| created\_by | INT | User who created the project |
| created\_at | TIMESTAMP DEFAULT CURRENT\_TIMESTAMP | Current timestamp of project creation |
| status | VARCHAR(20) DEFAULT “active” | Status of project (active, completed, cancelled) |
| quality\_grade | VARCHAR(50) | Quality grade of product |
| notes | TEXT | Notes taken about project |

## Project Farm Assignment

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| assignment\_id | INT (AUTO\_INCREMENT) | Primary key, unique identifier. |
| project\_id | INT | Foreign key referencing project table |
| farm\_id | INT | Foreign key referencing farm table |
| farmer\_id | INT | Foreign key referencing farmer table |
| contribution\_amount | DECIMAL(10,2) | Amount each farmer contributes to a project |
| contribution percentage | DECIMAL(5,2) | Percentage of contribution to the project |
| estimated\_revenue | DECIMAL(12,2) | Estimated revenue the farmer would get from the order/project |
| assigned\_at | TIMESTAMP | Timestamp of farmer assignment |
| status | VARCHAR(20) | Status of assignment (assigned, fulfilled, cancelled) |

**DATABASE OPTIMISATION AND TRIGGERS**

**Database Optimizations**

**1. Indexing**

The database implements several strategic indexes to optimize query performance:

**Users Table Indexes:**

* idx\_last\_login: Indexes the last login timestamp for quick user activity queries
* idx\_username\_active: Composite index on username and active status
* idx\_email\_active: Composite index on email and active status

**Project Table Indexes:**

* idx\_project\_name: Indexes project names for quick lookups
* crop\_product: Composite index on crop name and product type
* harvest\_start: Indexes harvest start dates
* harvest\_end: Indexes harvest end dates

**Project-Farm Assignment Indexes:**

* farmer\_farm\_contribution: Composite index on farmer\_id, farm\_id, and contribution amount

**2. Spatial Data Optimization**

* Implements proper SRID (Spatial Reference System Identifier) using WGS84 (SRID 4326)
* Uses appropriate spatial data types (POINT, POLYGON) for geographic data
* Includes spatial indexing for geo\_boundaries and geo\_coordinates

**Database Triggers**

**1. Production Cost Calculation Triggers**

* Automatically calculates total production cost based on farm acreage
* Updates costs when acreage changes
* Uses a fixed rate of 16,000 GHS per acre

**2. Revenue Calculation Triggers**

* Calculates revenue from sales based on tonnage and price per kilo
* Converts tonnage to kilograms (× 1000) before calculation
* Updates automatically on insert and update operations

**3. Income Loss Calculation Triggers**

* Calculates lost income based on:
* Loss percentage
* Total yield
* Tonnage sold
* Price per kilo
* Updates automatically on insert and update operations

**4. Yield Calculation Triggers**

* Calculates total yield in metric tons
* Applies loss percentage to initial yield
* Rounds results to 2 decimal places
* Updates automatically on insert and update operations

**Python Script for Automatic Acreage Calculation**

A Python script was written to ensure automatic calculation of acreages when geoboundary data is collected. This eliminates the need for manual entries. Formally, one would have to login to Kobo and get acreage data from submissions and manually input those values in Excel. The units of measurement Kobo provides are either in hectares or square meters, so one would have to manually convert to acres.

**Key Functions**

1. **calculate\_geodesic\_area(coordinates)**

* Calculates accurate area measurements on Earth's surface
* Uses WGS84 ellipsoid for precise calculations
* Converts coordinates to a closed polygon if necessary
* Returns area in square meters

2. **validate\_polygon(coordinates)**

* Validates polygon data structure
* Checks for minimum number of points (3)
* Verifies polygon closure
* Returns validation status and error message if applicable

**3. update\_acreage\_values()**

Main function that:

* Connects to the MySQL database
* Retrieves farm data with geographic boundaries
* For each farm:
* Converts WKB data to shapely polygon
* Validates polygon structure
* Calculates area using geodesic calculations
* Converts square meters to acres
* Updates database with calculated acreage
* Includes error handling and logging
* Manages database connections properly

**Technical Details**

* Uses mysql.connector for database operations
* Implements shapely for geometric calculations
* Uses pyproj for geodesic calculations
* Includes proper error handling and connection management
* Converts between different units (square meters to acres)

**Security Features**

* Database credentials are properly managed
* Connection is properly closed after operations
* Includes error handling for invalid data
* Validates geometric data before calculations

This script is crucial for maintaining accurate farm size data in the database, which is used by various triggers for cost and yield calculations.