# Introduction

This documentation describes the data model used in the project, including the schema structure, table relationships, and data dictionary. The data model has been staged and transformed through various stages, each of which is described in its own worksheet.

# Worksheets

The data transformation process is organized into different worksheets, each focusing on a specific task:

* STAGING Creation
* SENSORDATA CLEANUP
* CROPDATA CLEANUP
* WEATHERDATA CLEANUP
* LOCATIONDATA CLEANUP
* IRRIGATIONDATA CLEANUP
* PESTDATA CLEANUP
* SOILDATA CLEANUP
* DIMENSIONS
* FACTS

# Staging Tables

The data is initially staged in the DFA23RAWDATA.RAWDATA schema and later moved to the DFA23RAWDATA.DATAHACKERS schema. The tables at this stage are raw formats named as follows:

* STAGING\_WEATHERDATARAW
* STAGING\_CROPDATARAW
* STAGING\_IRRIGATIONDATARAW
* STAGING\_LOCATIONDATARAW
* STAGING\_PESTDATARAW
* STAGING\_SENSORDATARAW
* STAGING\_SOILDATARAW

# Cleaned Tables

The raw tables undergo cleaning and transformation to produce tables with a '\_CLEANED' suffix. These tables hold the transformed data and are derived from the '\_TYPED' tables, which are dropped after this stage.

* STAGING\_CROPDATARAW\_CLEANED
* STAGING\_IRRIGATIONDATA\_CLEANED
* STAGING\_LOCATIONDATA\_CLEANED
* STAGING\_PESTDATA\_CLEANED
* STAGING\_SENSORDATACLEANED
* STAGING\_SOILDATA\_CLEANED
* STAGING\_WEATHERDATARAW\_CLEANED

# Dimension Tables

Dimension tables hold unique, non-overlapping data that can be used to slice and dice the fact tables.

* LOCATION\_DIMENSION
* TIME\_DIMENSION
* CROP\_TYPE\_DIMENSION
* PEST\_TYPE\_DIMENSION
* IRRIGATION\_METHOD\_DIMENSION

# Fact Tables

Fact tables contain metrics that are subject to analysis and are linked to dimension tables through foreign keys.

* SENSOR\_DATA\_FACT
* CROP\_DATA\_FACT
* WEATHER\_DATA\_FACT
* SOIL\_CONDITION\_FACT

**DIMENSION TABLES DETAILS**

* LOCATION\_DIMENSION Details:
* Contains location-specific details. Fields include LOCATION\_ID, LOCATION\_NAME, LATITUDE, LONGITUDE, ELEVATION, and REGION.
* TIME\_DIMENSION Details:
* Captures time-related data. Fields include TIME\_ID, TIMESTAMP, YEAR, MONTH, DAY, and HOUR.
* CROP\_TYPE\_DIMENSION Details:
* Contains crop-related details. Fields include CROP\_TYPE\_ID, CROP\_TYPE, and GROWTH\_STAGE.
* PEST\_TYPE\_DIMENSION Details:
* Captures pest-related information. Fields include PEST\_TYPE\_ID, PEST\_TYPE, and PEST\_SEVERITY.
* IRRIGATION\_METHOD\_DIMENSION Details:
* Includes irrigation-related information. Fields include IRRIGATION\_METHOD\_ID, IRRIGATION\_METHOD, and WATER\_SOURCE.

**FACT TABLES RELATIONSHIPS**

* SENSOR\_DATA\_FACT Relationships:
* Linked to TIME\_DIMENSION through the TIME\_ID foreign key.
* CROP\_DATA\_FACT Relationships:
* Linked to TIME\_DIMENSION through the TIME\_ID foreign key and to CROP\_TYPE\_DIMENSION through the CROP\_TYPE\_ID foreign key.
* WEATHER\_DATA\_FACT Relationships:
* Linked to TIME\_DIMENSION through the TIME\_ID foreign key.
* SOIL\_CONDITION\_FACT Relationships:
* Linked to TIME\_DIMENSION through the TIME\_ID foreign key.

**URL TO THE WORKSHEETS**

* **STAGE DATA**  
  This links to the folder containing the sheet used to stage the data from DFA23RAWDATA.RAWDATA to DFA23RAWDATA.DATAHACKERS schema  
  **url** -> <https://app.snowflake.com/ugriiwa/or94187/#/stage-data-fWX4Dnj4o>
* **CLEAN AND TRANSFORM DATA**  
  This links to the folder containing the sheets for cleaning each of staged version for SensorData, WeatherData, SoilData, CropData, PestData, IrrigationData, LocationData in the DFA23RAWDATA.DATAHACKERS schema  
  **url ->** <https://app.snowflake.com/ugriiwa/or94187/#/staging-data-cleanup-fFO9pIkAL>
* **FACTS AND DIMENSIONS CREATION**  
  This links to the folder containing the creation of the facts and dimension tables  
  url -> https://app.snowflake.com/ugriiwa/or94187/#/facts-and-dimensions-fIvrY0OAO