



Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence & Data Science

| | |
|-----------------------------|--|
| Name: | |
| Roll No: | |
| Class/Sem: | TE/V |
| Experiment No.: | 1 |
| Title: | Data Warehouse Construction – Star schema and Snowflake schema |
| Date of Performance: | |
| Date of Submission: | |
| Marks: | |
| Sign of Faculty: | |



Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence & Data Science

Aim: To Build a Data Warehouse – Star Schema, Snowflake Schema and Fact Constellation Schema

Objective: A data warehouse is a large store of data collected from multiple sources within a business. The objective of a data warehouse system is to provide consolidated, flexible, meaningful data storage to the end user for reporting and analysis.

Theory:

In general, the warehouse design process consists of the following steps:

1. Choose a business process to model (e.g., orders, invoices, shipments, inventory, account administration, sales, or the general ledger). If the business process is organizational and involves multiple complex object collections, a data warehouse model should be followed. However, if the process is departmental and focuses on the analysis of one kind of business process, a data mart model should be chosen.
2. Choose the business process grain, which is the fundamental, atomic level of data to be represented in the fact table for this process (e.g., individual transactions, individual daily snapshots, and so on).
3. Choose the dimensions that will apply to each fact table record. Typical dimensions are time, item, customer, supplier, warehouse, transaction type, and status.
4. Choose the measures that will populate each fact table record. Typical measures are numeric additive quantities like dollars sold and units sold.

Steps to Draw Star, Snowflake, and Fact Constellation Schemas

1. Star Schema:

- Step 1: Identify the central fact table, which contains quantitative data (e.g., sales, revenue).
- Step 2: Determine the dimension tables related to the fact table, such as time, product, customer, etc.
- Step 3: Define the relationships between the fact table and each dimension table, usually a one-to-many relationship.
- Step 4: Draw the fact table at the center and connect it to each dimension table using lines, creating a star-like structure.



Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence & Data Science

2. Snowflake Schema:

- Step 1: Start with the fact table as in the Star Schema.
- Step 2: Identify the dimension tables and further normalize them by breaking them into multiple related tables (e.g., split "Location" into "Country" and "City").
- Step 3: Establish relationships between these normalized dimension tables and the fact table.
- Step 4: Draw the fact table at the center, then connect it to the dimension tables, which in turn connect to their sub-tables, forming a snowflake-like structure.

3. Fact Constellation Schema:

- Step 1: Identify multiple fact tables representing different processes or subjects (e.g., sales and inventory).
- Step 2: Identify the shared dimension tables that will connect to these fact tables.
- Step 3: Define relationships between each fact table and the shared dimension tables.
- Step 4: Draw all fact tables and connect them to the shared dimension tables, creating a constellation of facts and dimensions.

Problem Statement:

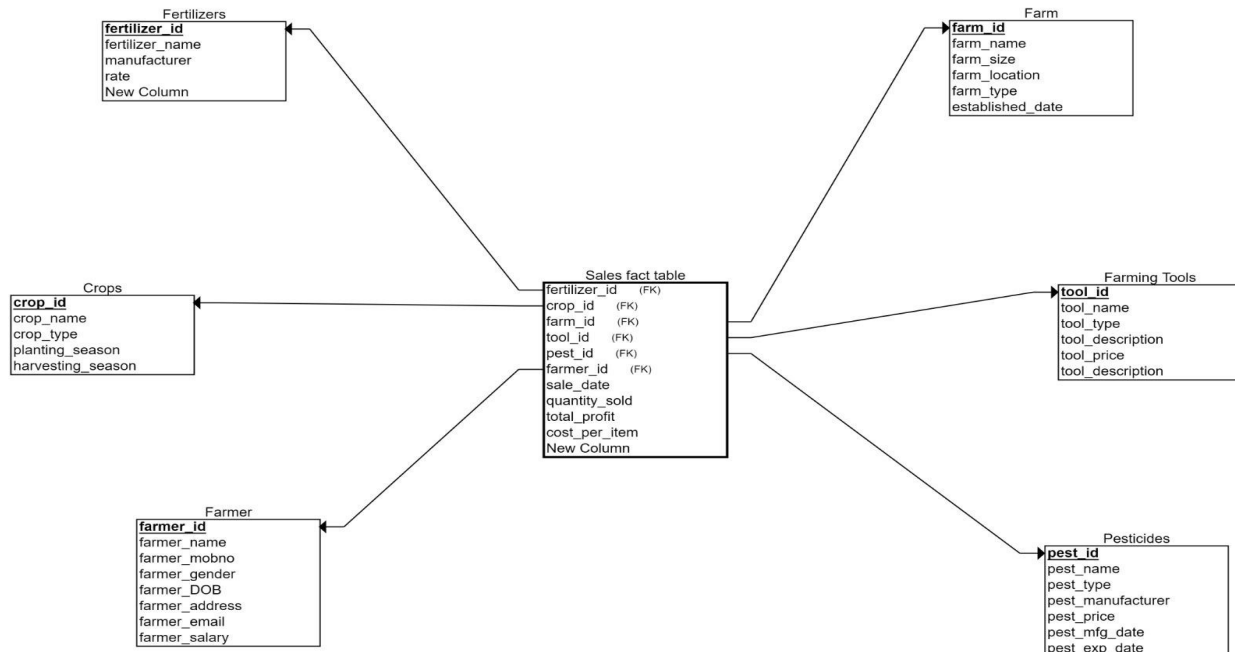
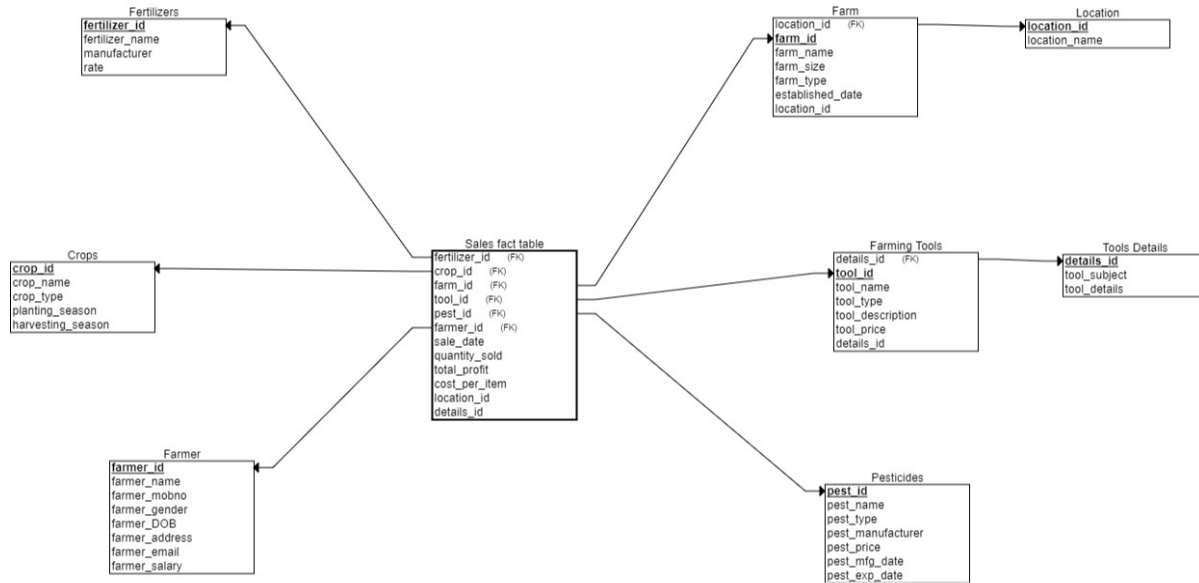
The existing data architecture makes it difficult for business analysts to quickly access and analyze sales data across different dimensions, such as time, location, and product. The lack of an integrated view of sales metrics and associated attributes results in slow report generation and poor decision-making capabilities.



Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence & Data Science

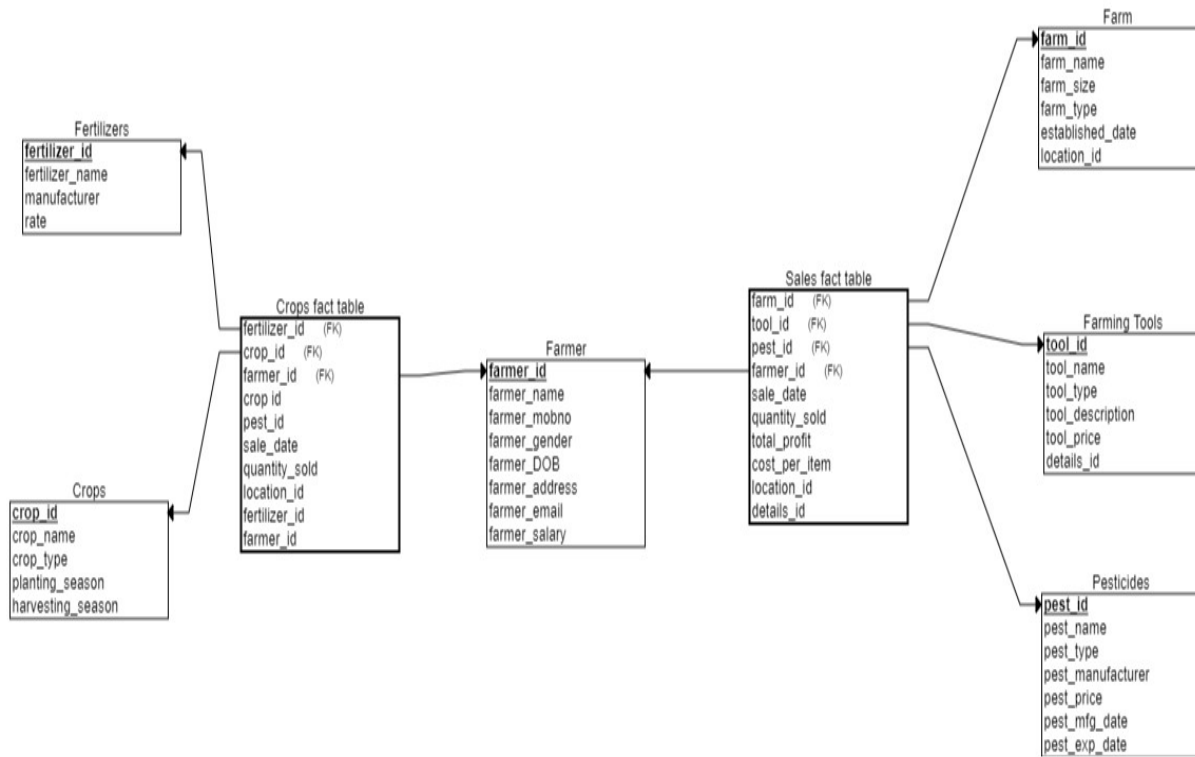
Construction of Star schema, Snowflake schema and Fact Constellation Schema:





Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence & Data Science



Conclusion:

1. How does the Snowflake Schema compare to the Star Schema in terms of ease of maintenance and scalability?
2. How do you manage the complexity of ETL (Extract, Transform, Load) processes in a Fact Constellation Schema?