Schemas for Multidimensional Databases

What is Schema?

- Database uses relational model while data warehouse requires Schema.
- Schema is a logical description of the entire database.
- It includes the name and description of records.
- Much like a database, a data warehouse also requires to maintain a schema.

Types of Schemas

- Star Schema
- Snowflake Schema
- Galaxy Schema

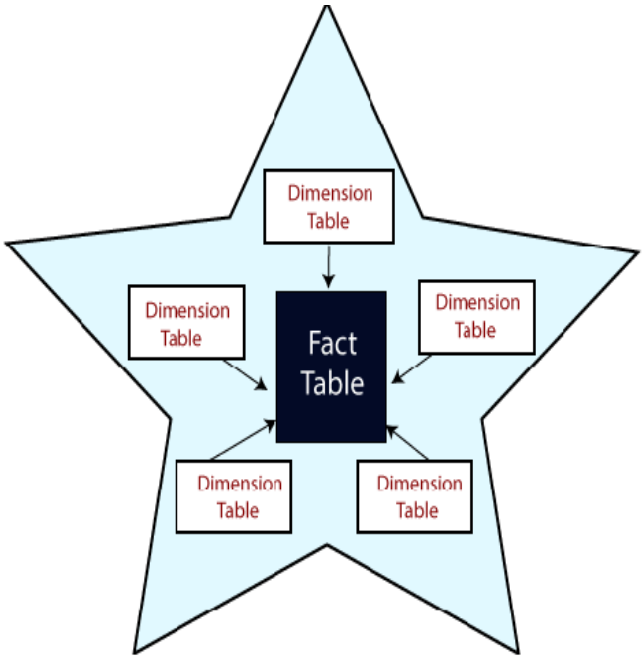
Dimension Table

- Contain information about a particular dimension.
- Textual information of the business
- Stores attributes or dimensions that describes the objects in a fact table.
- Dimension table has a surrogate key column that uniquely identifies each dimension record.
- It is de-normalized because built to analyze data as easily as possible.

What is Star Schema?

A star schema is the elementary form of a dimensional model, in which dataare organized into facts and dimensions. A fact is an event that is counted or measured, such as a sale or log in. A dimension includes reference data aboutthe fact, such as date, item, or customer.

A star schema is a relational schema where a relational schema whose design represents a multidimensional data model. The star schema is the explicit data warehouse schema. It is known as star schema because the entity- relationship diagram of this schemas simulates a star, with points, diverge from a central table. The center of the schema consists of a large fact table, andthe points of the star are the dimension tables.



Star Schema

Characteristics of Star Schema

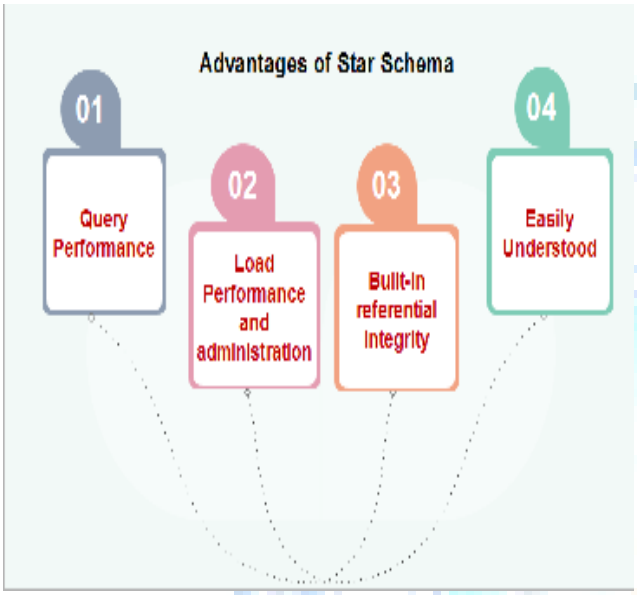
The star schema is intensely suitable for data warehouse database design because ofthe following features:

- It creates a DE-normalized database that can quickly provide query responses.
- It provides a flexible design that can be changed easily or added to throughoutthe development cycle, and as the database grows.
- It provides a parallel in design to how end-users typically think of and use thedata.
- It reduces the complexity of metadata for both developers and end-users.

Advantages of Star Schema

Star Schemas are easy for end-users and application to understand and navigate. With a well-designed schema, the customer can instantly analyze large, multidimensional data sets.

The main advantage of star schemas in a decision-support environment are:



What is Snowflake Schema?

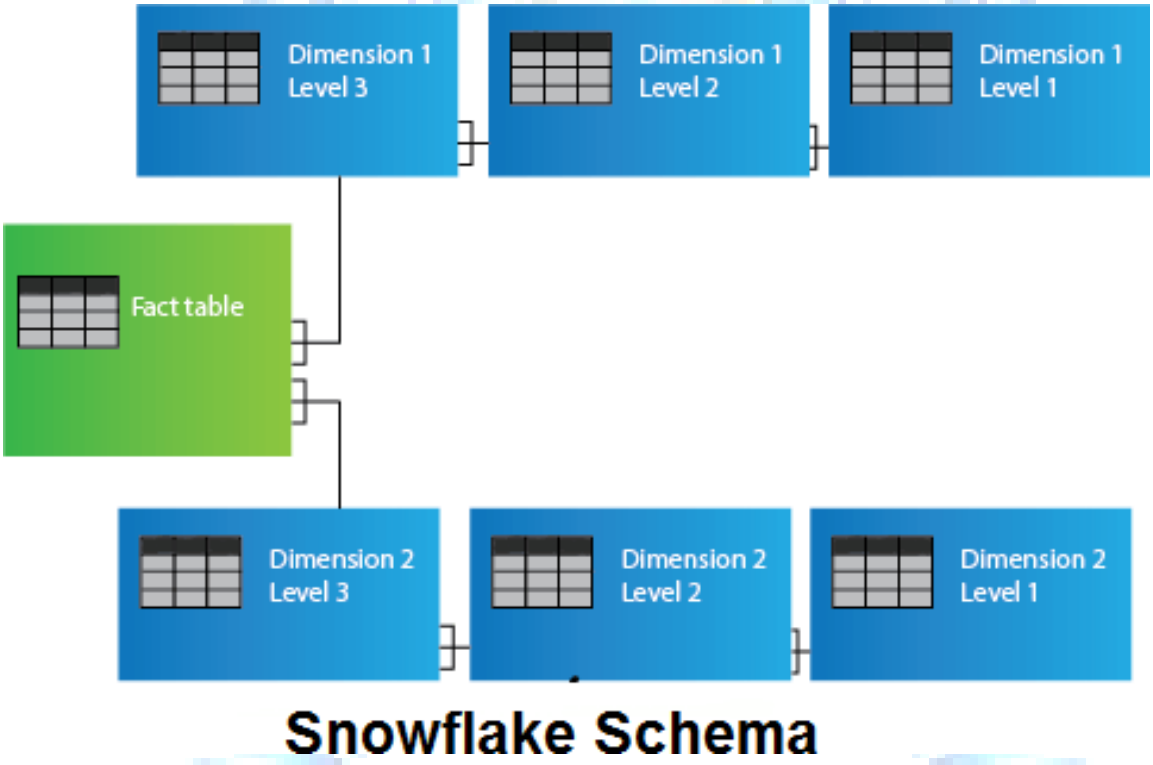
- A snowflake schema is equivalent to the star schema. "A schema is known as a snowflake if one or more dimension tables do not connect directly to the fact table but must join through other dimension tables."
- The snowflake schema is an expansion of the star schema where each point of the star explodes into more points. It is called snowflake schema because the diagram of snowflake schema resembles a snowflake.
- Snowflaking is a method of normalizing the dimension tables in a STAR schemas. When we normalize all the dimension tables entirely, the resultant structure resembles a snowflake with the fact table in the middle.

Snowflaking is used to develop the performance of specific queries. The schema is diagramed with each fact surrounded by its associated dimensions, and those dimensions are related to other

dimensions, branching out into a snowflake pattern.

The snowflake schema consists of one fact table which is linked to many dimension tables, which can be linked to other dimension tables through a many-to-one relationship. Tables in a snowflake schema are generally normalized to the third normal form. Each dimension table performs exactly one level in a hierarchy.

The following diagram shows a snowflake schema with two dimensions, each having three levels. A snowflake schemas can have any number of dimension, and each dimension can have any number of levels.

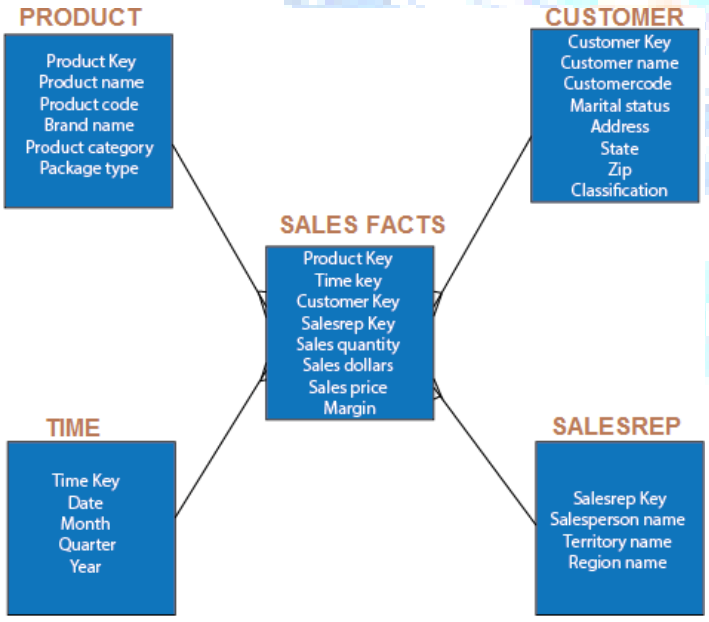


Snowflake Schema

Example: Figure shows a snowflake schema with a Sales fact table, with Store, Location, Time, Product, Line, and Family dimension tables. The Market dimension has two dimension tables with

Store as the primary dimension table, and Location asthe outrigger dimension table. The product dimension has three dimension tables with Product as the primary dimension table, and the Line and Family table are the outrigger dimension tables.

- A star schema store all attributes for a dimension into one denormalized table. This needed more disk space than a more normalized snowflake schema. Snowflaking normalizes the dimension by moving attributes with low cardinality into separate dimension tables that relate to the core dimension table by using foreign keys. Snowflaking for the sole purpose of minimizing disk space is not recommended, because it can adversely impact query performance.
- In snowflake, schema tables are normalized to delete redundancy. In snowflake dimension tables are damaged into multiple dimension tables.
- Figure shows a simple STAR schema for sales in a manufacturing company.The sales fact table include quantity, price, and other relevant metrics. SALESREP, CUSTOMER, PRODUCT, and TIME are the dimension tables.



STAR Schema

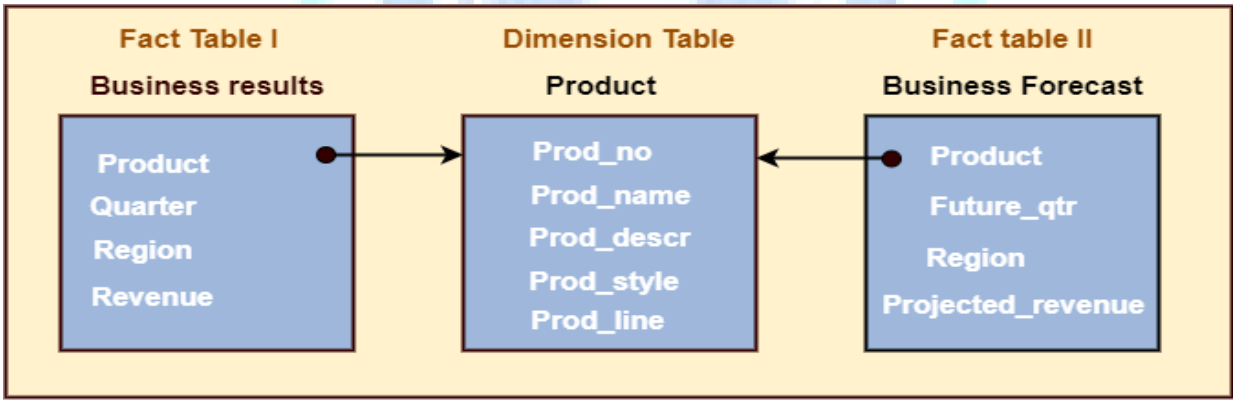
The STAR schema for sales, as shown above, contains only five tables, whereas the normalized

version now extends to eleven tables. We will notice that in the snowflake schema, the attributes with low cardinality in each original dimension tables are removed to form separate tables.

What is Fact Constellation Schema?

- A Fact constellation means two or more fact tables sharing one or moredimensions.
- It is also called Galaxy schema.

- Fact Constellation Schema describes a logical structure of data warehouse ordata mart.
- Fact Constellation Schema can design with a collection of de-normalized FACT,Shared, and Conformed Dimension tables.
- Fact Constellation Schema is a sophisticated database design that is difficult tosummarize information.
- Fact Constellation Schema can implement between aggregate Fact tables ordecompose a complex Fact table into independent simplex Fact tables.
-



FACT Constellation Schema