

Schema Types In Data Warehouse Modeling

Data Warehouse Schema

In a data warehouse, a schema is used to define the way to organize the system with all the database entities (fact tables, dimension tables) and their logical association.

The different types of Schemas in DW:

1. Star Schema
2. Snow Flake Schema
3. Galaxy Schema
4. Star Cluster Schema

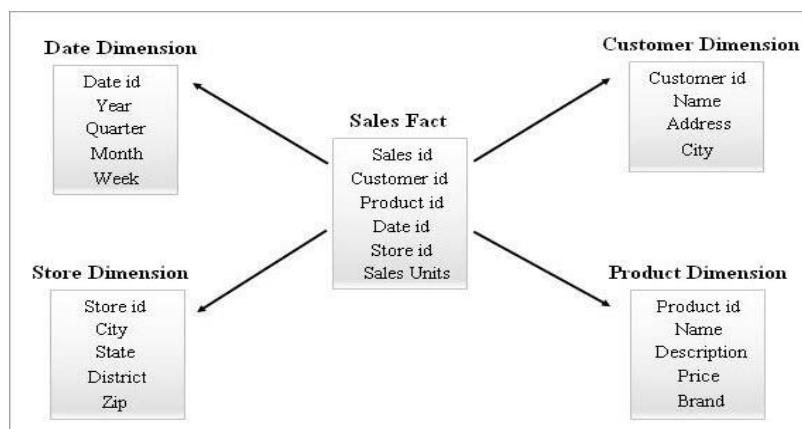
1) Star Schema

This is the simplest and most effective schema in a data warehouse. A fact table in the center surrounded by multiple dimension tables resembles a star in the Star Schema model. The fact table maintains one-to-many relations with all the dimension tables. Every row in a fact table is associated with its dimension table rows with a foreign key reference.

Due to the above reason, navigation among the tables in this model is easy for querying aggregated data. An end-user can easily understand this structure. Hence all the Business Intelligence (BI) tools greatly support the Star schema model.

While designing star schemas the dimension tables are purposefully de-normalized. They are wide with many attributes to store the contextual data for better analysis and reporting.

An example of a Star Schema:



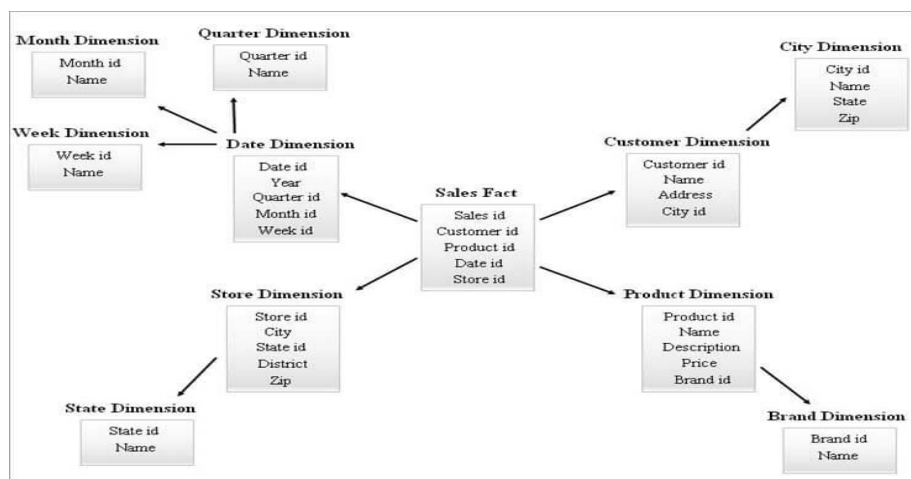
2) Snowflake Schema

Star schema acts as an input to design a Snowflake schema. Snow flaking is a process that completely normalizes all the dimension tables from a star schema.

The arrangement of a fact table in the center surrounded by multiple hierarchies of dimension tables looks like a Snowflake in the Snowflake schema model. Every fact table row is associated with its dimension table rows with a foreign key reference.

While designing Snowflake schemas the dimension tables are purposefully normalized. Foreign keys will be added to each level of the dimension tables to link to its parent attribute. The complexity of the Snowflake schema is directly proportional to the hierarchy levels of the dimension tables.

An example of a Snowflake Schema:



The Dimension Tables in the above Snowflake Diagram are normalized as:

- Date dimension is normalized into Quarterly, Monthly and Weekly tables by leaving foreign key ids in the Date table.
- The store dimension is normalized to comprise the table for State.
- The product dimension is normalized into Brand.
- In the Customer dimension, the attributes connected to the city are moved into the new City table by leaving a foreign key id in the Customer table.

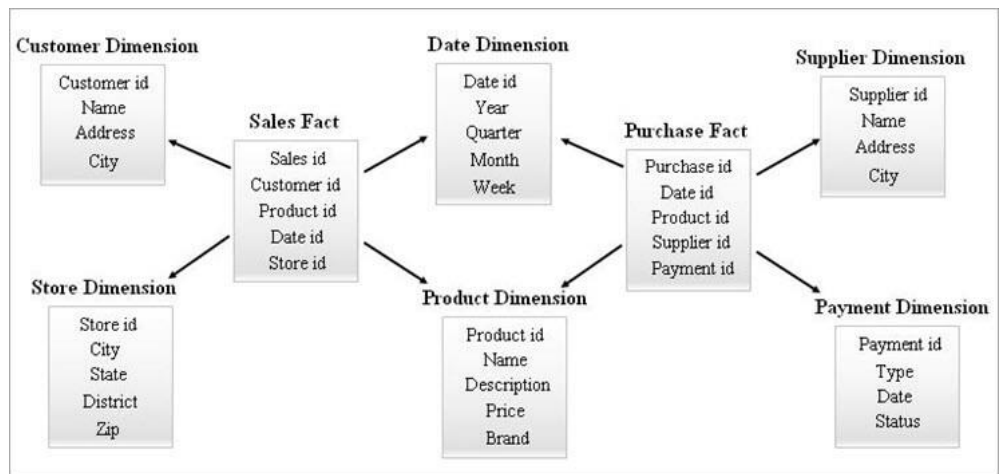
3) Galaxy Schema

A galaxy schema is also known as Fact Constellation Schema. In this schema, multiple fact tables share the same dimension tables. The arrangement of fact tables and dimension tables looks like a collection of stars in the Galaxy schema model.

The shared dimensions in this model are known as Conformed dimensions.

This type of schema is used for sophisticated requirements and for aggregated fact tables that are more complex to be supported by the Star schema (or) Snowflake schema. This schema is difficult to maintain due to its complexity.

An example of Galaxy Schema:



4) Star Cluster Schema

A Snowflake schema with many dimension tables may need more complex joins while querying. A star schema with fewer dimension tables may have more redundancy. Hence, a star cluster schema came into the picture by combining the features of the above two schemas.

Star schema is the base to design a star cluster schema and few essential dimension tables from the star schema are snowflaked and this, in turn, forms a more stable schema structure.

An example of a Star Cluster Schema:

