Fact Constellation Schema

Dr. Shalini Gambhir

Introduction to Fact Constellation Schema

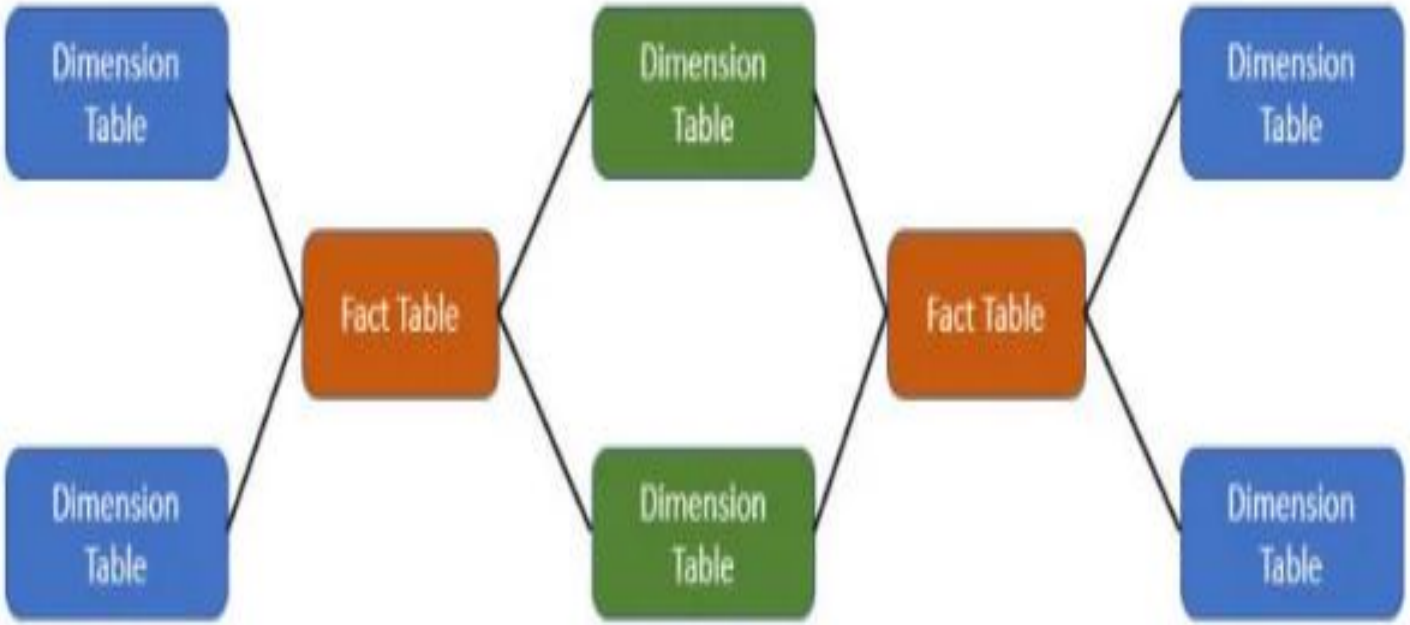
- Fact Constellation Schema is a schema that represents a multidimensional model of tables. This schema is a group of different fact tables that have few similar dimensional tables. It can be represented as a group of multiple star schemas and thus, it is also called a Galaxy schema. Fact schema is the most frequently used schema to design a Data warehouse and also, it is a little complicated than star and snowflake schema model.
- Fact constellation schema is a tool of **analytical processing** via online, which has a huge group of the number of **fact tables** that share **dimensional tables**, also aggregated as a group of stars. We can also call it as an extension of the **star constellation model**.

What is Fact Constellation Schema?

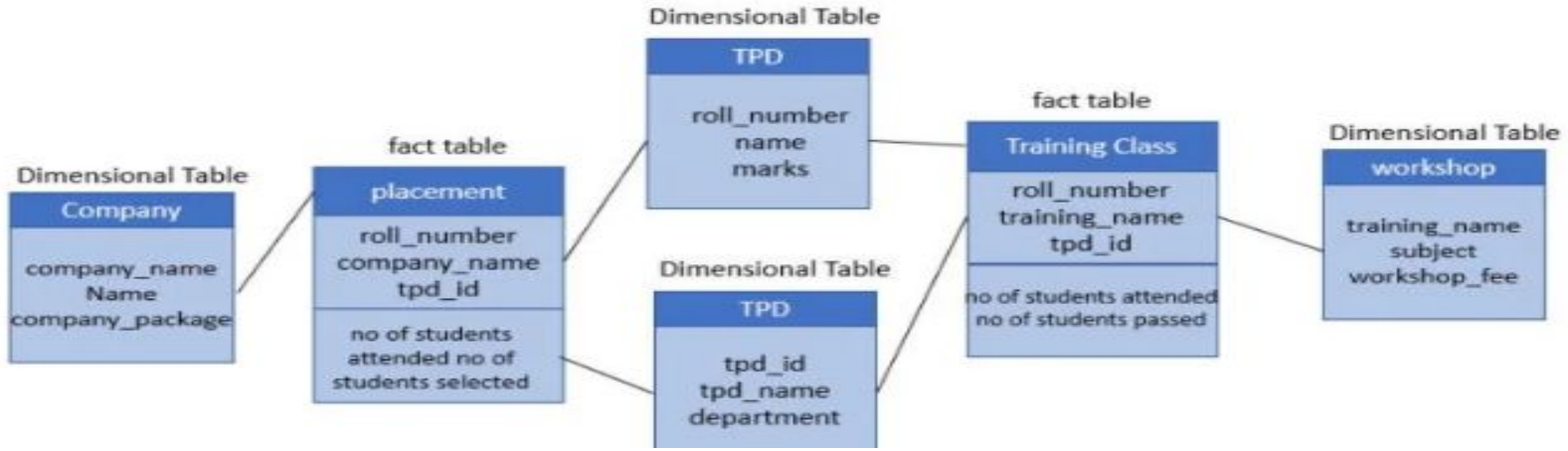
- A fact constellation schema can have multiple fact tables associated with it. It is commonly used a schema for designing data warehouses and it is much complicated than any other schema such as star and snowflake schema. We can create a fact constellation schema from a star schema by splitting them into one or more star schemas. It is also said that a fact constellation schema can have many fact tables and a shared dimensional table.

General Structure of Fact Constellation

- A combined group of star schema is called a fact constellation schema.



Fact Constellation Schema Architecture



***The student is a dimensional table where roll_number, name, marks are its attributes.**
***TPD is a dimensional table where tpd_id, tpd_name, department are its attributes.**
***Placement is a fact table where roll_number, company_name, tpd_id are its attributes with its facts number of students attended, number of students selected.**
***Training class is a fact table where roll_number, training_name, tpd_id are its attributes, and the number of students attended, the number of students selected is its facts.**
***The workshop is a dimensional table where training_name, subject, workshop_fee are its attributes.**
***Company is a dimensional table where company_name, name, company_package are its attributes.**
Here in the above schema, placement and training_class are two different star schemas having their dimensional table as student, TPD, and company in Star schema with placement as its fact table. The other star schema is student, TPD, and workshop as its dimensional table with training class as its fact table.

- Here in the previous example two-star schema has two-dimensional tables in common and thus they form a set of fact constellation schema.

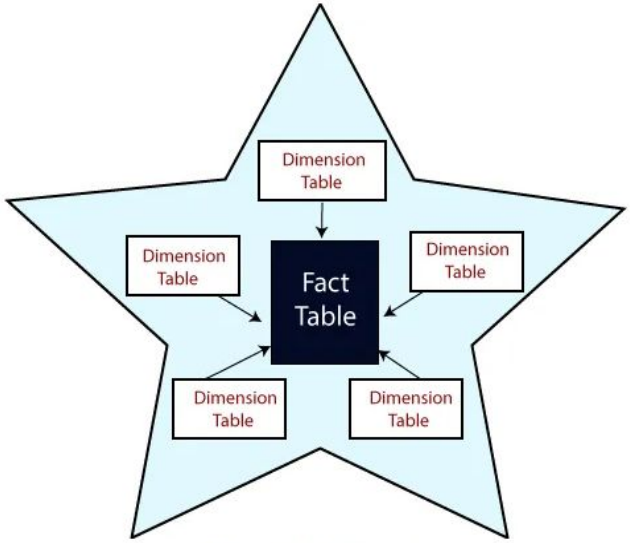
Advantages and Disadvantages

- Few advantages that make it more useful:
- Tables are subdivided into fact and dimensional to understand the relationship between them.
- It is a flexible schema that makes users use it.
- Here dimensional tables are shared by the number of fact tables.
- It is a normalized form of snowflake and star schema.
- We can access the data in the database using complex queries.

- Few disadvantages are listed below:
- It is difficult to understand as it is a very complex schema to implement.
- It uses more space in the database comparative to the star schema.
- It has many joins between dimensional and fact tables and thus it is difficult to understand.
- This is difficult to maintain and operate.

- Fact constellation schema is also known as galaxy schema. It is possible to share two or more dimensional table with one fact table in a group of the star schema. It is a normalized form of snowflake schema and star schema. The main disadvantage of it is that it is a very challenging design architecture as it might have joined, aggregation relationship between dimensional and fact tables.

Star, Snowflake, and Fact Constellation(Galaxy Schema) Schemas



Star Schema

Star Schema:The star schema is a widely used schema design in data warehousing. It features a central fact table that holds the primary data or measures, such as sales, revenue, or quantities. The fact table is connected to multiple dimension tables, each representing different attributes or characteristics related to the data in the fact table. The dimension tables are not directly connected to each other, creating a simple and easy-to-understand structure.

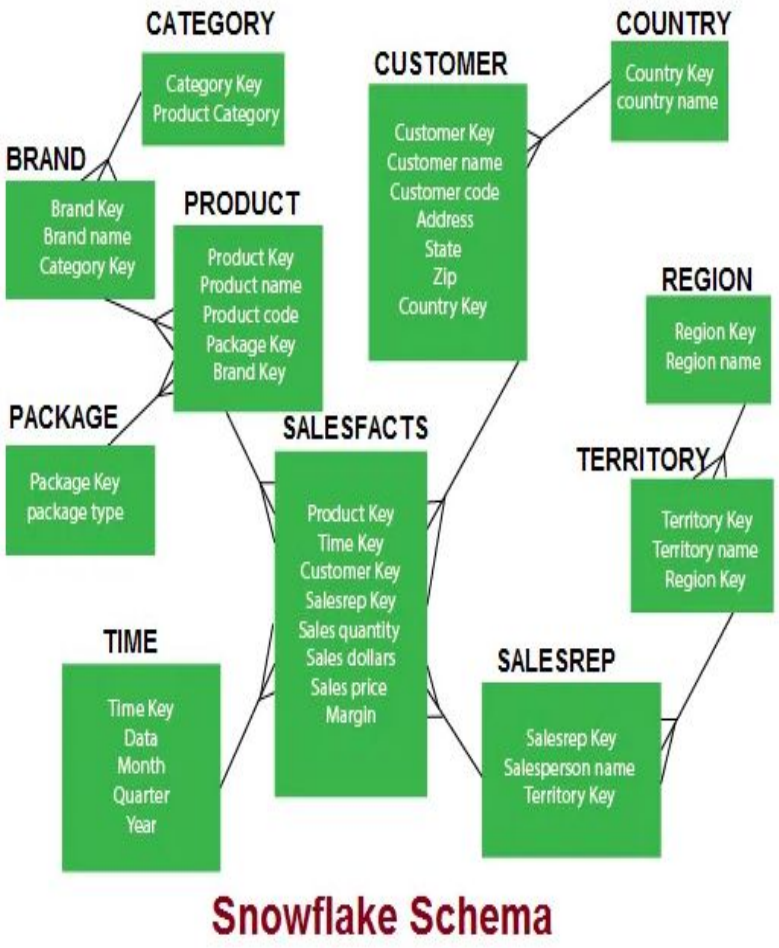
Simplicity*: Star schema is the simplest and most straightforward schema design, with fewer tables and relationships. It provides ease of understanding, querying, and report generation.*

Denormalization*: Dimension tables in star schema are often denormalized, meaning they may contain redundant data to optimize query performance.*

Star-Schema Example

- Example: Consider a retail data warehouse. The fact table might contain sales data with measures like “Total Sales” and “Quantity Sold.” The dimension tables could include “Product” with attributes like “Product ID,” “Product Name,” and “Category,” and “Time” with attributes like “Date,” “Month,” and “Year.”
- The fact table connects to these dimension tables through foreign keys, allowing analysts to perform queries like “Total Sales by Product Category” or “Quantity Sold by Date.”

Snowflake schema



Snowflake Schema

Snowflake Schema: The snowflake schema is an extension of the star schema, designed to further reduce data redundancy by normalizing the dimension tables. In a snowflake schema, dimension tables are broken down into multiple related sub-tables. This normalization creates a more complex structure with additional levels of relationships, reducing storage requirements but potentially increasing query complexity due to the need for additional joins.

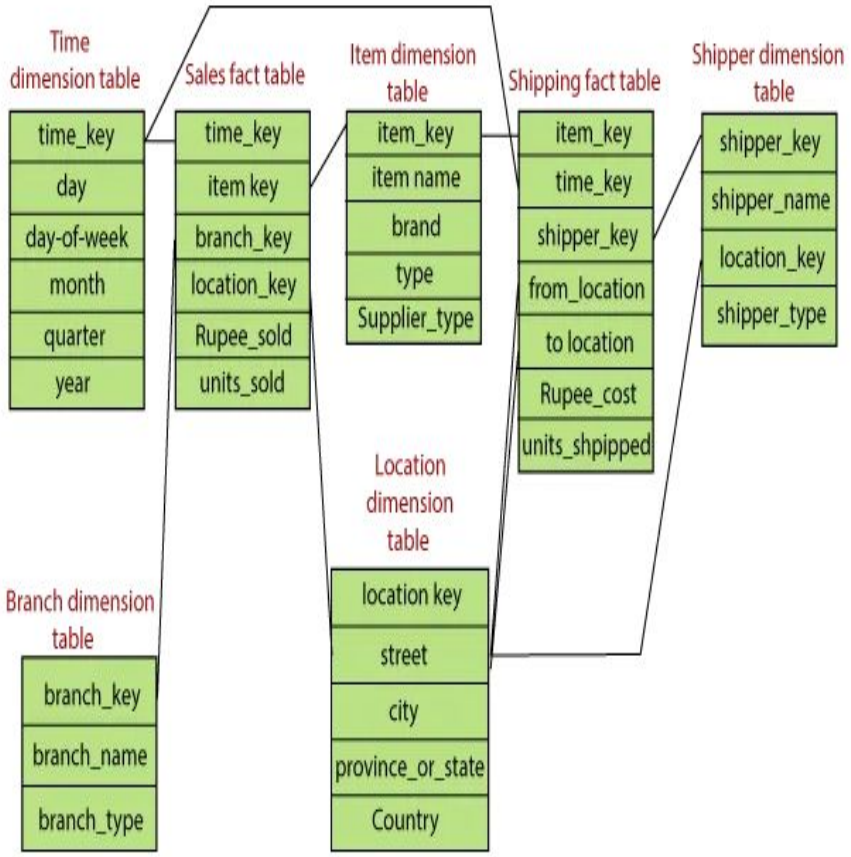
Normalization*: Snowflake schema normalizes dimension tables, resulting in more tables and more complex relationships compared to the star schema.*

Space Efficiency*: Due to normalization, the snowflake schema may require less storage space for dimension data but may lead to more complex queries due to additional joins*

Example-Snowflake schema

- Example: Continuing with the retail data warehouse example, in a snowflake schema, the “Product” dimension may be normalized into sub-tables like “Product_Category,” “Product_Subcategory,” and “Product_Details,” each holding specific attributes related to the product. This normalization allows for efficient storage of data, but it may require more complex queries to navigate through the snowflake structure.

Fact Constellation Schema



Fact Constellation (Galaxy Schema):The fact constellation schema, also known as a galaxy schema, is a more complex design that involves multiple fact tables sharing dimension tables. It is used when there are multiple fact tables with different measures and each fact table is related to several common dimension tables.

Complexity: *Fact constellation schema is the most complex among the three designs, as it involves multiple interconnected star schemas.*

Flexibility: *This schema design offers more flexibility in modeling complex and diverse business scenarios, allowing multiple fact tables to coexist and share dimensions.*

Example-Fact Constellation schema

- Example: In a data warehouse for a healthcare organization, there could be multiple fact tables representing different metrics like patient admissions, medical procedures, and medication dispensing. These fact tables would share common dimension tables like “Patient,” “Doctor,” and “Date.” The fact constellation schema allows analysts to analyze different aspects of healthcare operations while efficiently reusing shared dimension tables.