

Data Warehousing - Schemas

- ❑ Schema is a logical description of the entire database.
- ❑ It includes the name and description of records of all record types including all associated data-items and aggregates.
- ❑ Like a database, a data warehouse also requires to maintain a schema.

Model

- ❑ A database uses relational model, while a data warehouse uses
 1. Star,
 2. Snowflake, and
 3. Fact Constellation schema.

Fact Table:

- ❑ A fact table is a table in a data warehouse that stores facts.
- ❑ Facts are numeric data that describe business events, such as sales transactions.
- ❑ A fact table is usually connected to dimension tables,
- ❑ Store the descriptive information about the events, such as the customer's name, the product's description, and the date of the sale.

Dimension Table:

- ❑ A dimension table is a database table referencing defining pieces of information or attributes for particular records in a primary database table.
- ❑ A dimension table is a database table that contains descriptive information.
- ❑ A dimension table has fields that are descriptive, such as text, date and time, or even a number.

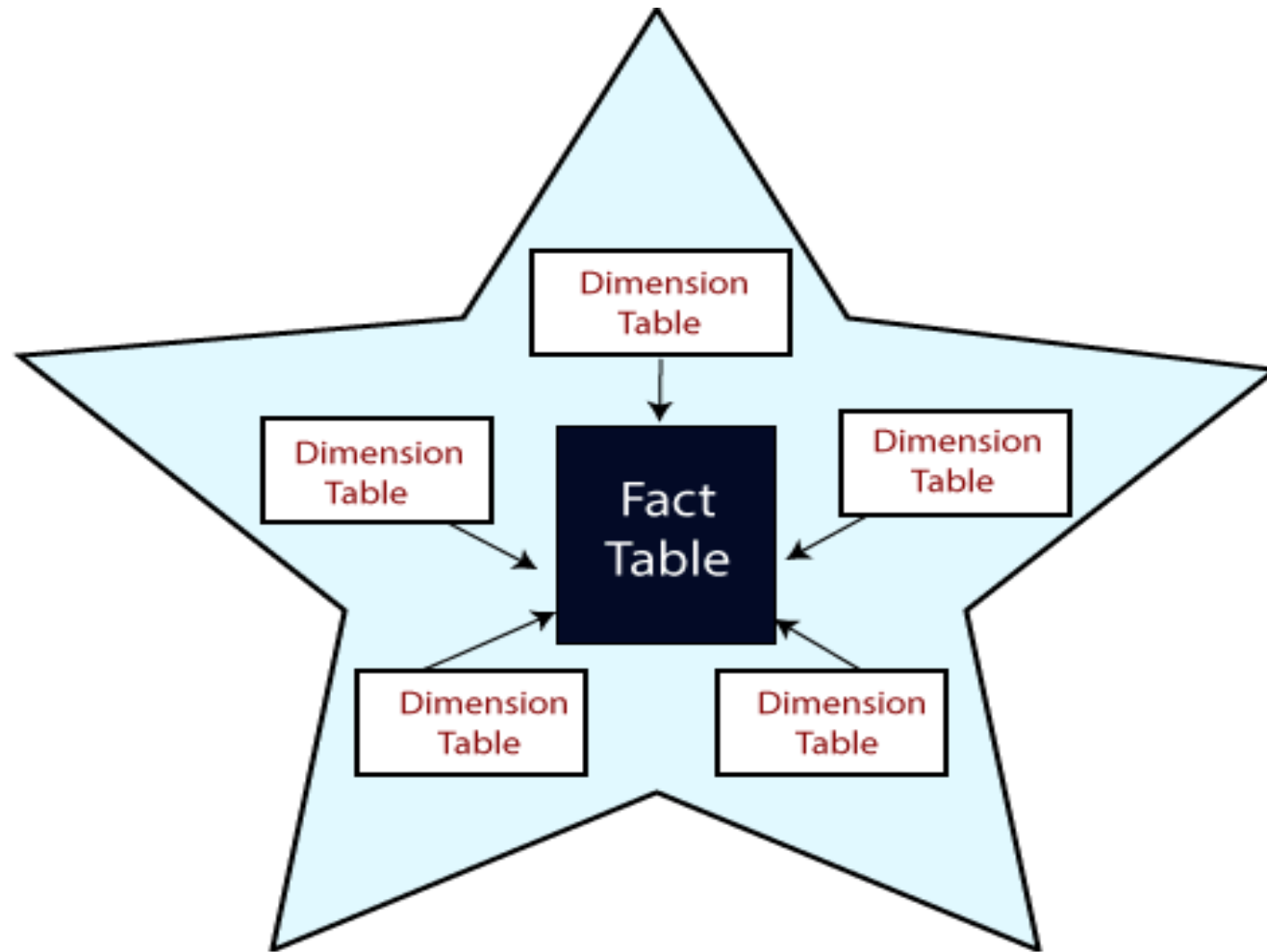
Difference between a Fact table and a Dimension table:

1. The fact table contains measurements, metrics, and facts about a business process, while the Dimension table is a companion to the fact table, which contains descriptive attributes to be used as query constraining.
2. The fact table is located at the center of a star or snowflake schema, whereas the Dimension table is located at the edges of the star or snowflake schema.
3. A fact table is defined by its grain or most atomic level, whereas a Dimension table should be wordy, descriptive, complete, and of assured quality.
4. The fact table helps to store report labels, whereas Dimension table contains detailed data.
5. The fact table does not contain a hierarchy, whereas the Dimension table contains hierarchies.

Star Schema:

- ❑ A star schema is the elementary form of a dimensional model, in which data are 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 about the 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, and the points of the star are the dimension tables.



Star Schema

- ❑ Example: Suppose a star schema is composed of a fact table, SALES, and several dimension tables connected to it for time, branch, item, and geographic locations.
- ❑ The TIME table has a column for each day, month, quarter, and year. The ITEM table has columns for each item_Key, item_name, brand, type, supplier_type. The BRANCH table has columns for each branch_key, branch_name, branch_type. The LOCATION table has columns of geographic data, including street, city, state, and country.

Dimension Table

time
time_key
day
day_of_the_week
month
Quarter
Year

Sales Fact Table

time_key
item_key
branch_key
location_key
unit_sold
dollars_sold

Dimension Table

item
item_key
item_name
brand
type
supplier_type

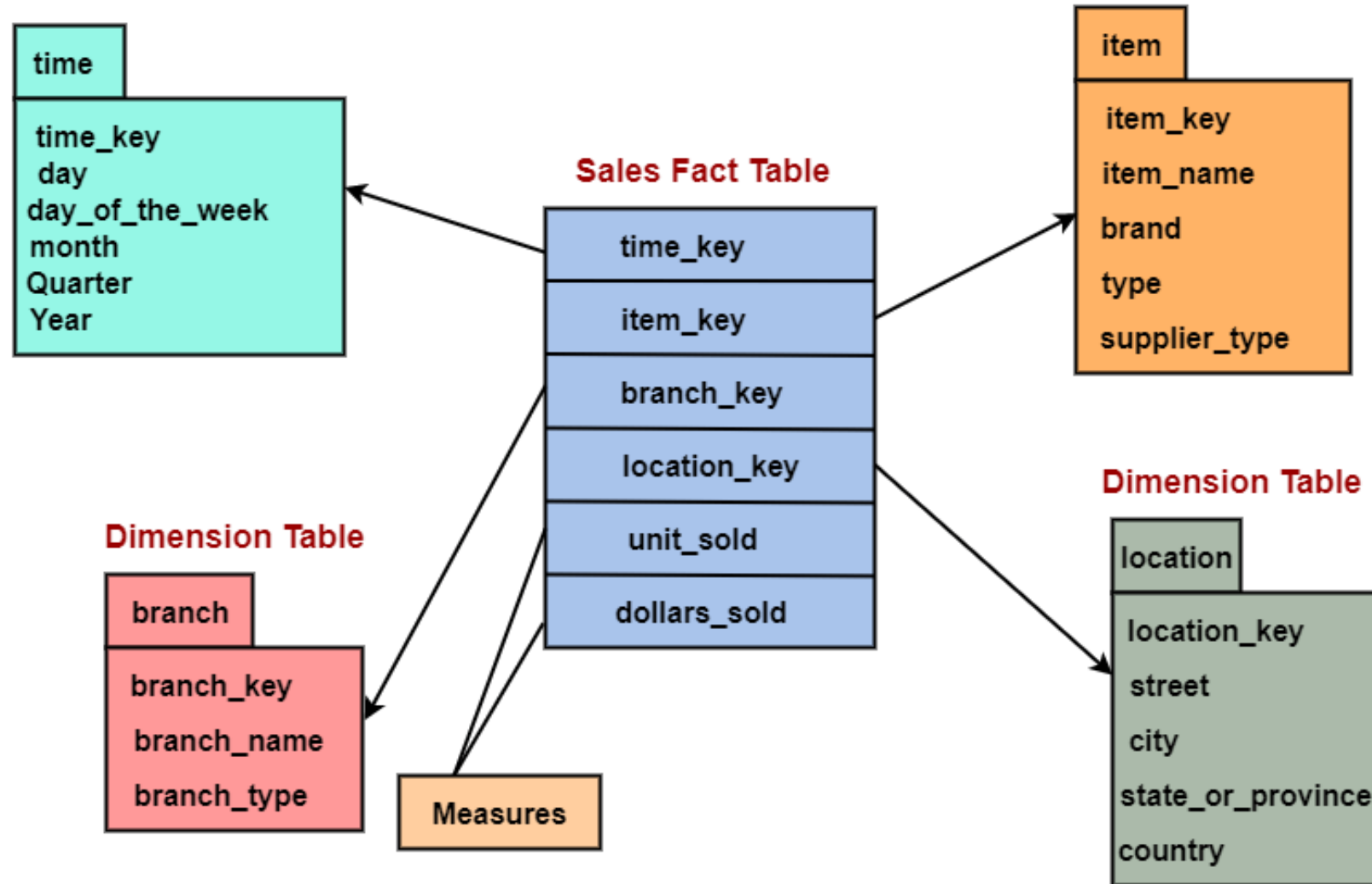
Dimension Table

branch
branch_key
branch_name
branch_type

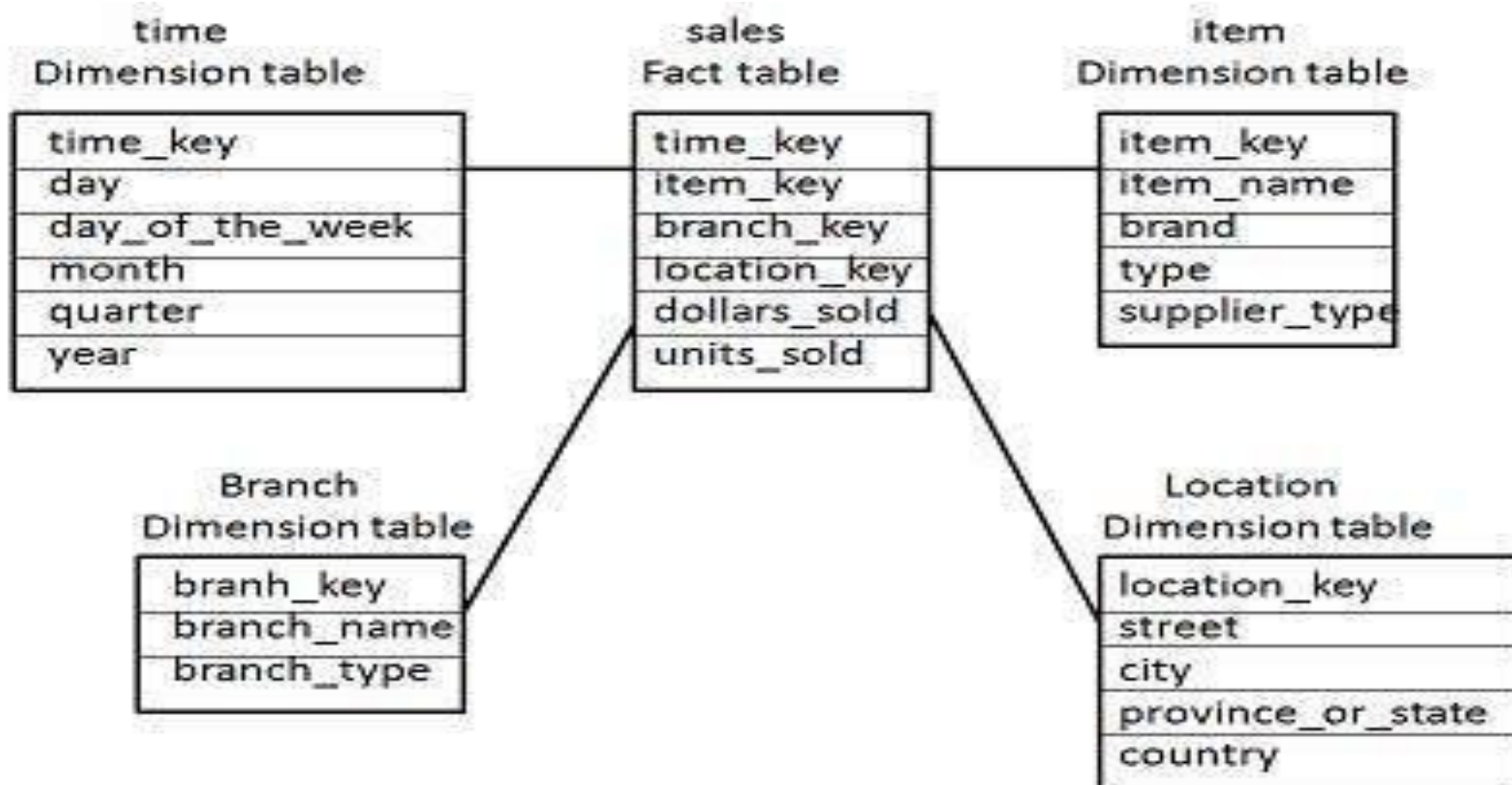
Dimension Table

location
location_key
street
city
state_or_province
country

Measures



- ❑ Each dimension in a star schema is represented with only one-dimension table.
- ❑ This dimension table contains the set of attributes.
- ❑ The following diagram shows the sales data of a company with respect to the four dimensions, namely time, item, branch, and location.
- ❑ There is a fact table at the center. It contains the keys to each of four dimensions.
- ❑ The fact table also contains the attributes, namely dollars sold and units sold.



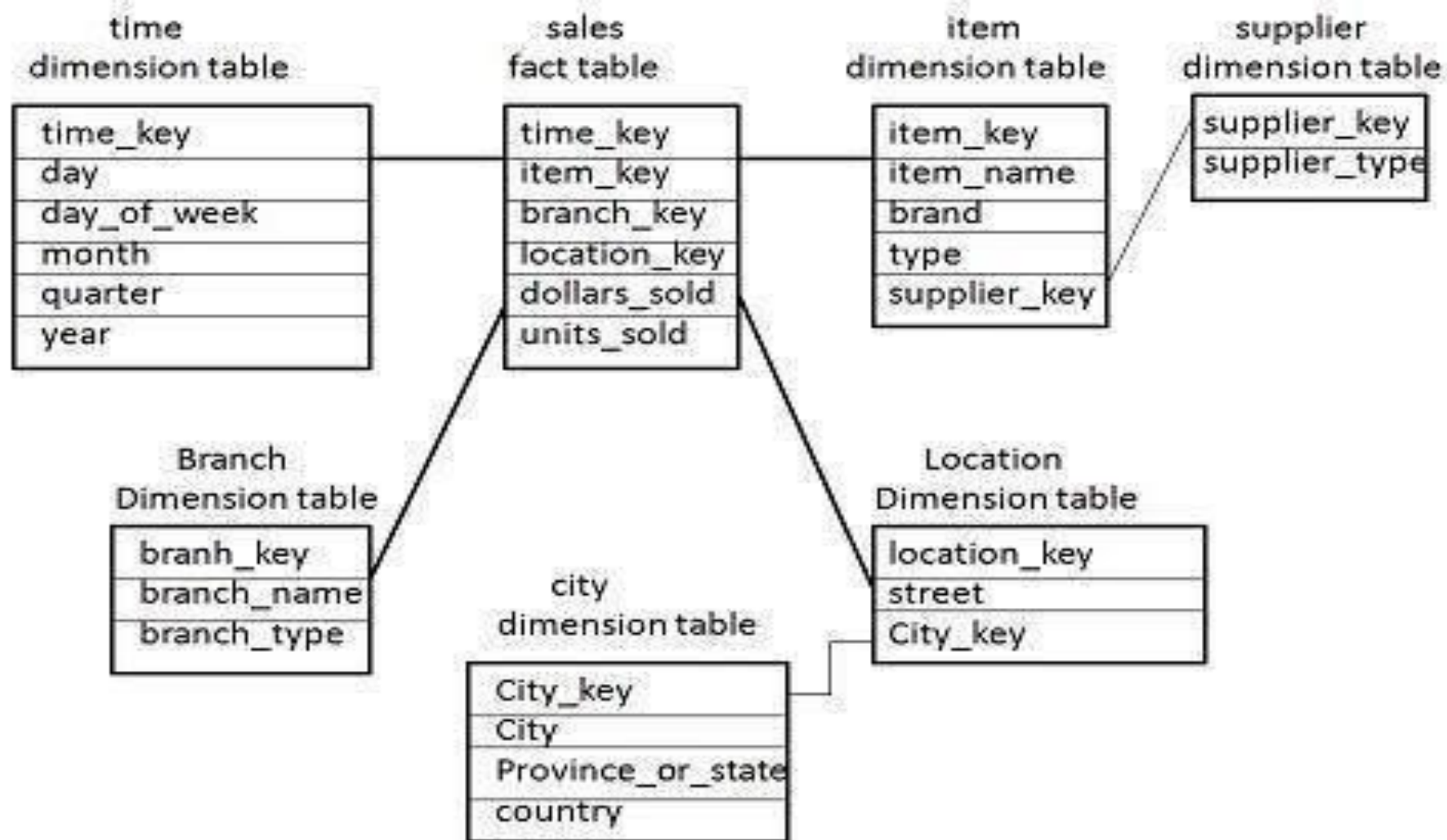
- ❑ In this scenario, the SALES table contains only four columns with IDs from the dimension tables, TIME, ITEM, BRANCH, and LOCATION, instead of four columns for time data, four columns for ITEM data, three columns for BRANCH data, and four columns for LOCATION data.
- ❑ Thus, the size of the fact table is significantly reduced.
- ❑ When we need to change an item, we need only make a single change in the dimension table, instead of making many changes in the fact table.

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 normalized dimension table is called a Snowflake.
- ❑ The snowflake schema is an expansion of the star schema where each point of the star explodes into more points.
- ❑ 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.

- ❑ Some dimension tables in the Snowflake schema are normalized.
- ❑ The normalization splits up the data into additional tables.
- ❑ Unlike Star schema, the dimensions table in a snowflake schema are normalized.
- ❑ For example, the item dimension table in star schema is normalized and split into two dimension tables, namely item and supplier table.



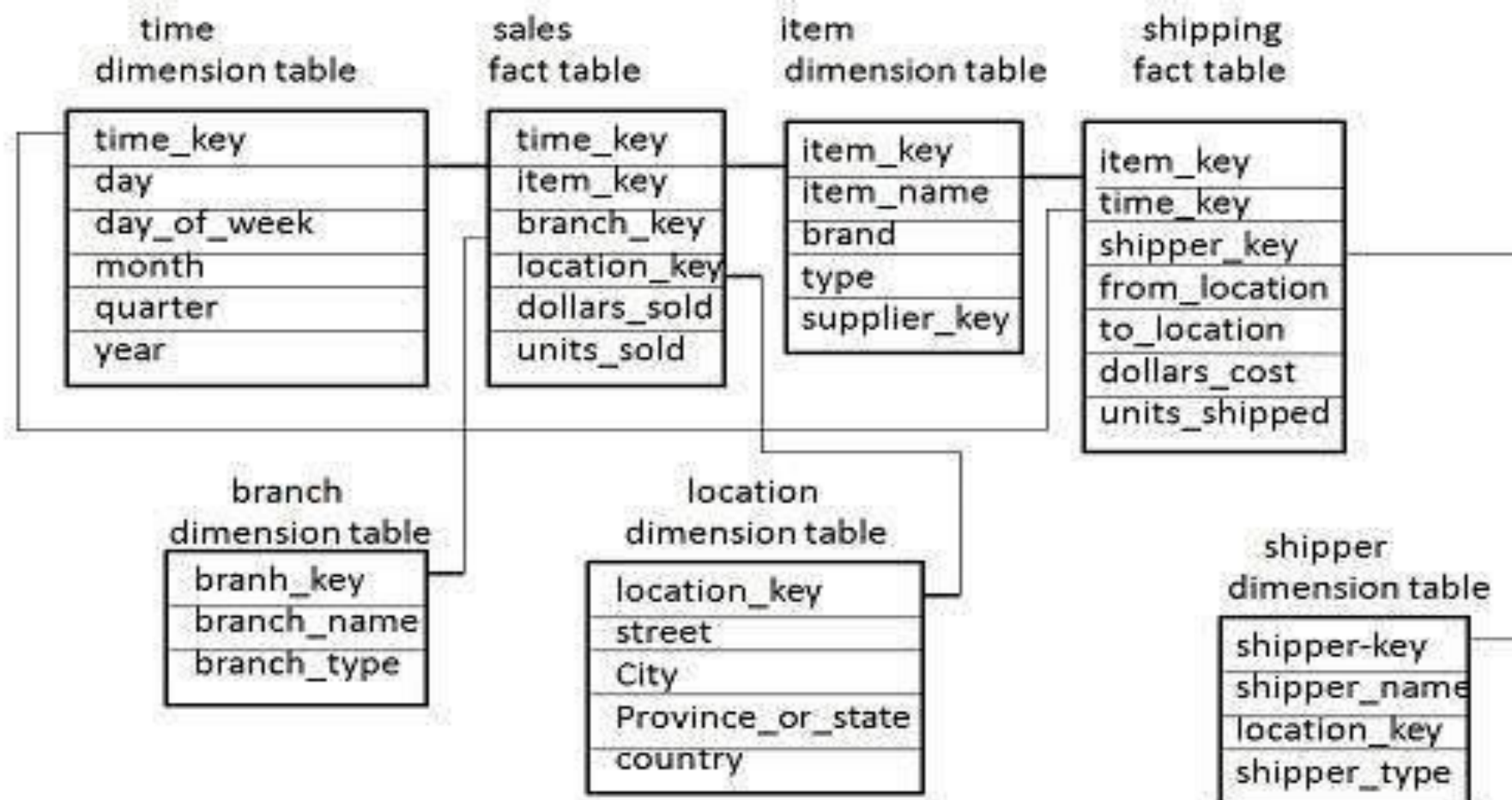
- ❑ Now the item dimension table contains the attributes item_key, item_name, type, brand, and supplier-key.
- ❑ The supplier key is linked to the supplier dimension table.
- ❑ The supplier dimension table contains the attributes supplier_key and supplier_type.
- ❑ Due to normalization in the Snowflake schema, the redundancy is reduced and therefore, it becomes easy to maintain and the save storage space.

Fact constellation schema

- ❑ In a data warehouse, a fact table is a table that stores the measurements, metrics, or facts related to a business operation.
- ❑ It is located at the center of a star or snowflake schema and is surrounded by dimension tables.
- ❑ When multiple fact tables are used, they can be organized using a "fact constellation schema."

Fact Constellation Schema

- ❑ A fact constellation has multiple fact tables.
- ❑ It is also known as galaxy schema.
- ❑ The following diagram shows two fact tables, namely sales and shipping.



- ❑ The sales fact table is same as that in the star schema.
- ❑ The shipping fact table has the five dimensions, namely item_key, time_key, shipper_key, from_location, to_location.
- ❑ The shipping fact table also contains two measures, namely dollars sold and units sold.
- ❑ It is also possible to share dimension tables between fact tables.
- ❑ For example, time, item, and location dimension tables are shared between the sales and shipping fact table.

