Data Engineer

Galaxy – Schema:

WH Star und Snowflakeschema

Conceptual Modeling of Data Warehouses

- Modeling data warehouses: multidimensional model
  - Star schema: A fact table in the middle connected to a set of dimension tables
  - Snowflake schema: A refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake
  - Fact constellations: Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called galaxy schema or fact constellation

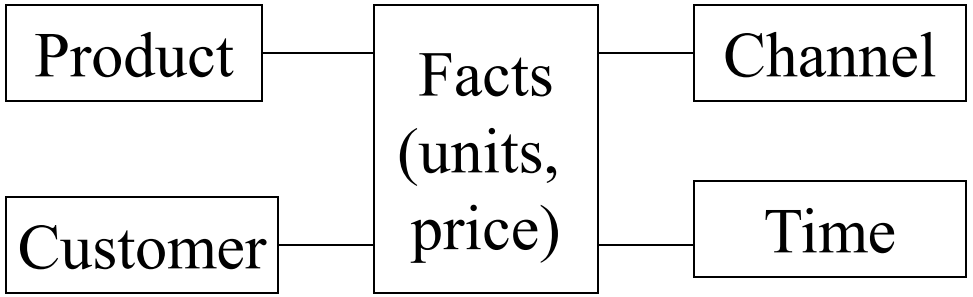
Creating the Dimensional Model

- Identify fact tables
  - Translate business measures into fact tables
  - Analyze source system information for additional measures
  - Identify base and derived measures
  - Document additivity of measures
- Identify dimension tables
- Link fact tables to the dimension tables
- Create views for users

Dimension Tables

Dimension tables have the following characteristics:

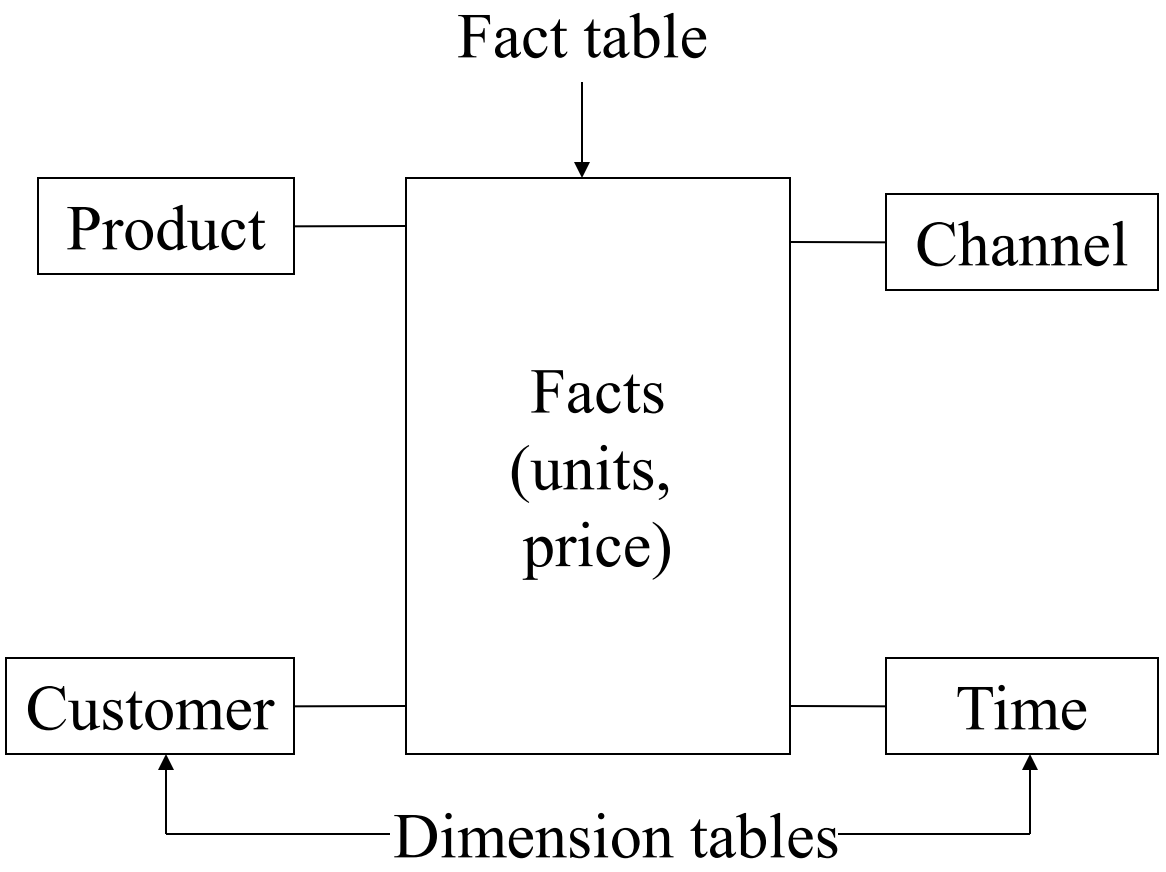
- Contain textual information that represents the attributes of the business
- Contain relatively static data
- Are joined to a fact table through a foreign key reference



Fact Tables

Fact tables have the following characteristics:

- Contain numeric measures (metric) of the business
- May contain summarized (aggregated) data
- May contain date-stamped data
- Are typically additive
- Have key value that is typically a concatenated key composed of the primary keys of the dimensions
- Joined to dimension tables through foreign keys that reference primary keys in the dimension tables


Fact table
Product
Channel
Facts
(units,
price)
Customer
Time
Dimension tables

Star Schema Model

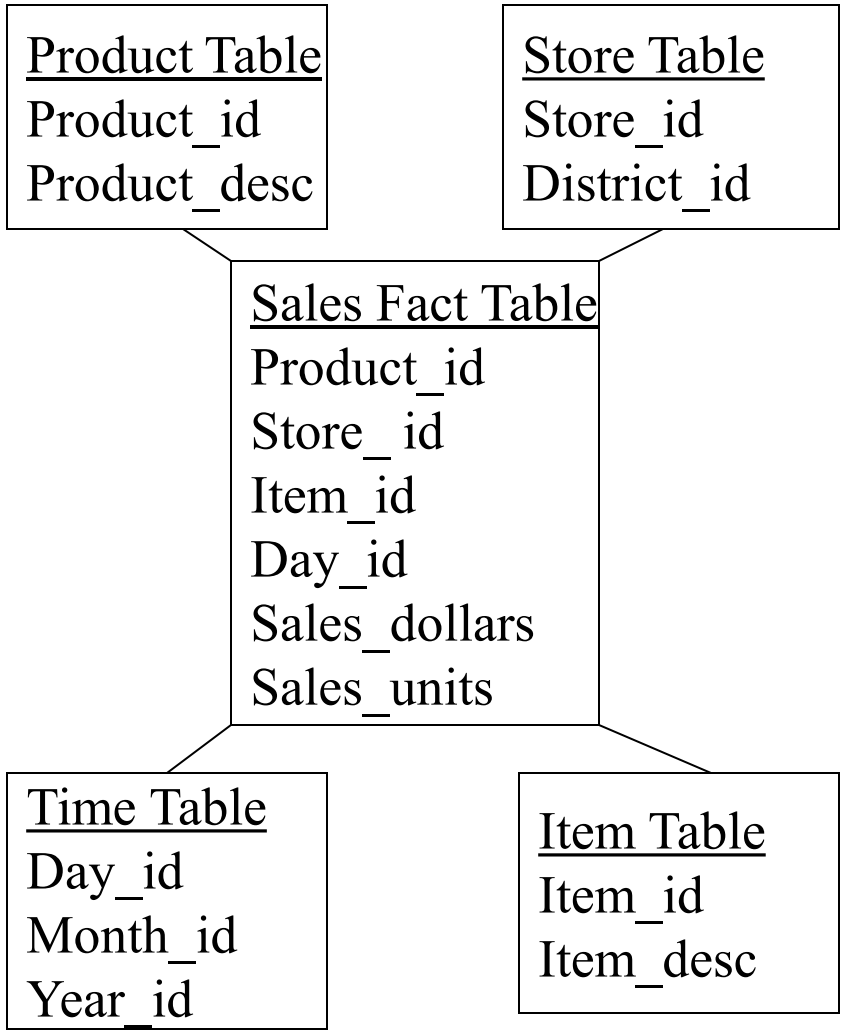
- The star schema separates business process data into ***facts,*** which hold the measurable, quantitative data about a business, and ***dimensions***, which are descriptive attributes related to fact data.

Facts and Dimension attribute

- Examples of fact data include
  - sales price, sale quantity, and time, distance, speed, and weight measurements.
- Related dimension attribute examples include
  - product models, product colors, product sizes, geographic locations, and salesperson name

Star Schema Model

- Central fact table
- Radiating dimensions
- Denormalized model



Features of Star Schema Model

- Easy for users to understand
- Fast response to queries
- Simple metadata
- Supported by many front end tools
- Less robust to change
- Slower to build
- Does not support history

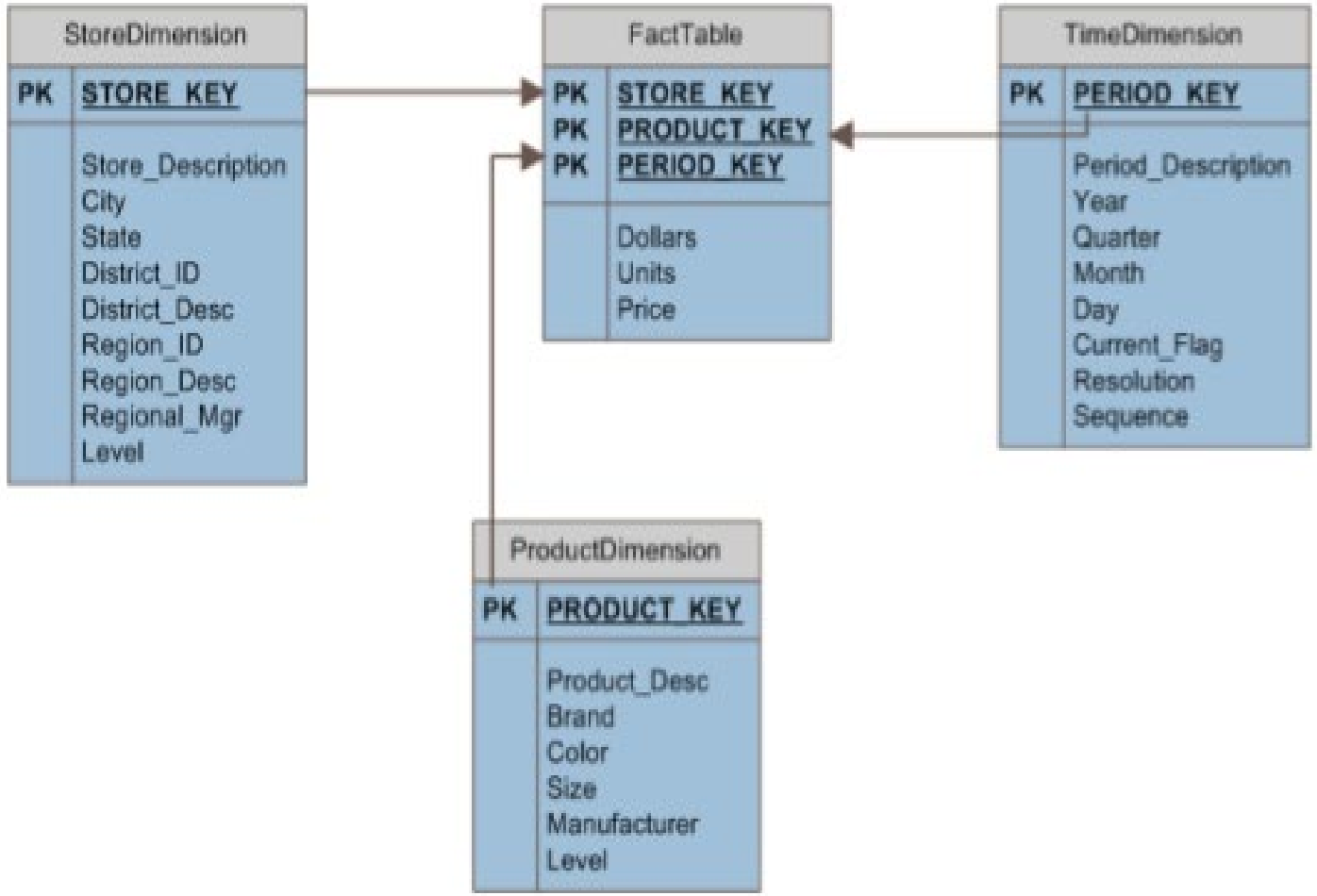
Classic star schema

- A single fact table containing a compound primary key, with one segment for each dimension and additional columns of additive numeric facts
- A single dimensional table (for each dimension) with a generated key, and a level indicator that describe the attribute level of each record

- The single fact will contain the *detail data*, such as sales dollars ***for a given store***, for a given product, in a given time period
- The fact table may also contain *partially consolidated data*,

  such as sales dollars ***for a region***, for a given product, in a given time period

Classic Star Schema Model



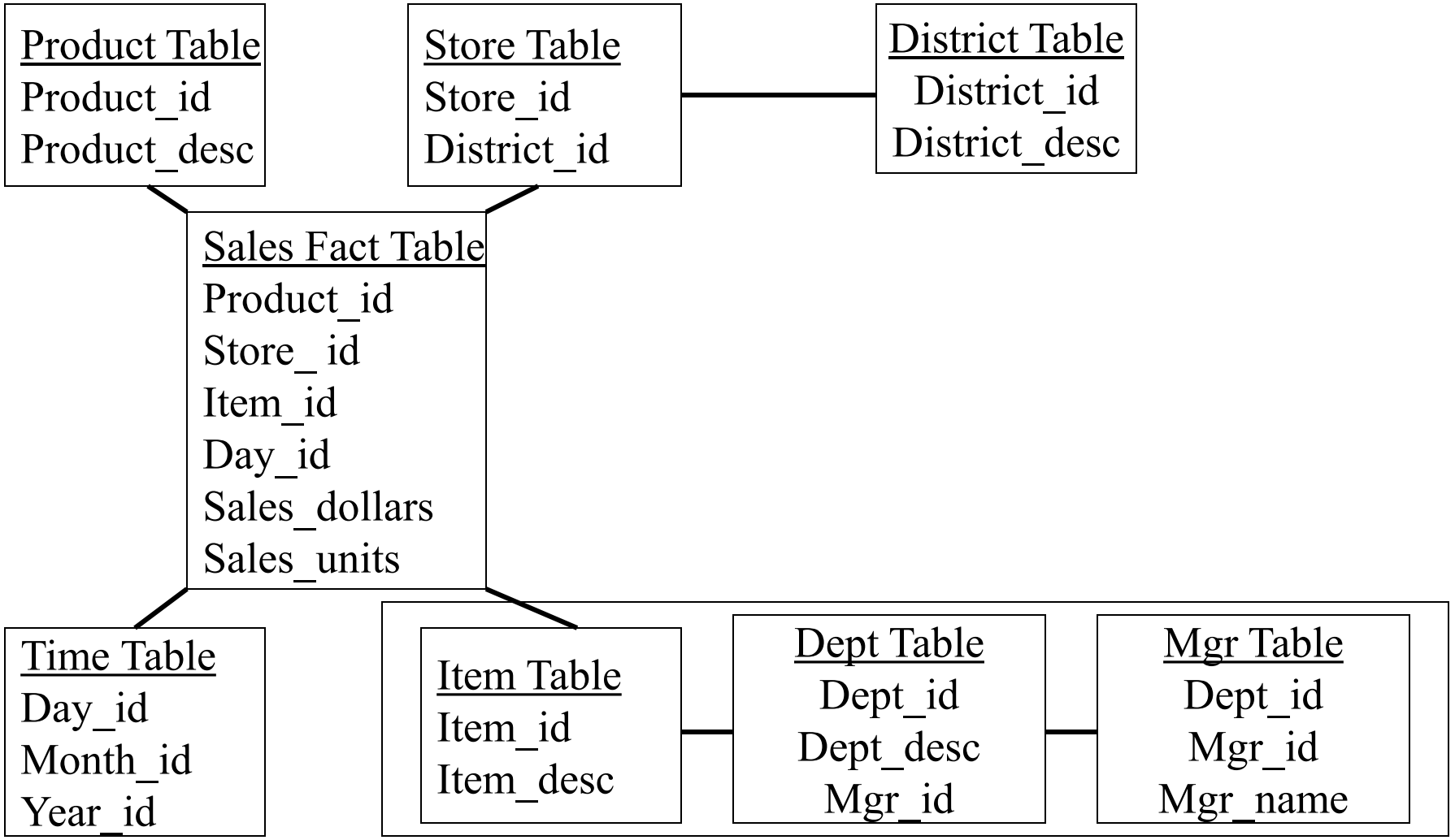
Limitation

- Star schema must contain
  - either All of the combination of aggregated data or
  - At least views of every combination
- Dimension table must carry a level indicator for every record and every query must use it
  - Select a.store_key, a.period_key, a.dollars From Fact table a Where a.store_key in (select store_key from store_dimension Where region = “Nrth’ and level=3);

Snowflake Schema Model

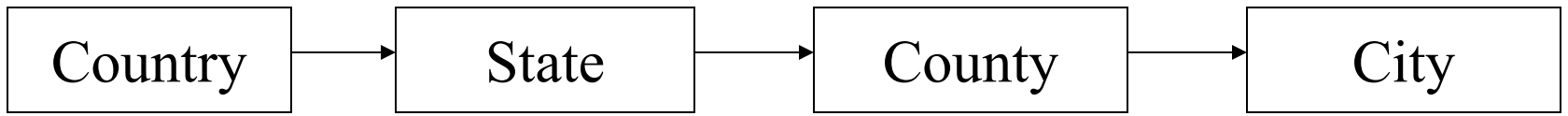
- There is no level in the dimension table
- Dimension tables are normalized by decomposing at the attribute level
- Each dimension table has only one key for each level of dimension’s hierarchy
- The lowest level key joins the dimension table to both the fact table and the lower level attribute table.

Snowflake Schema Model

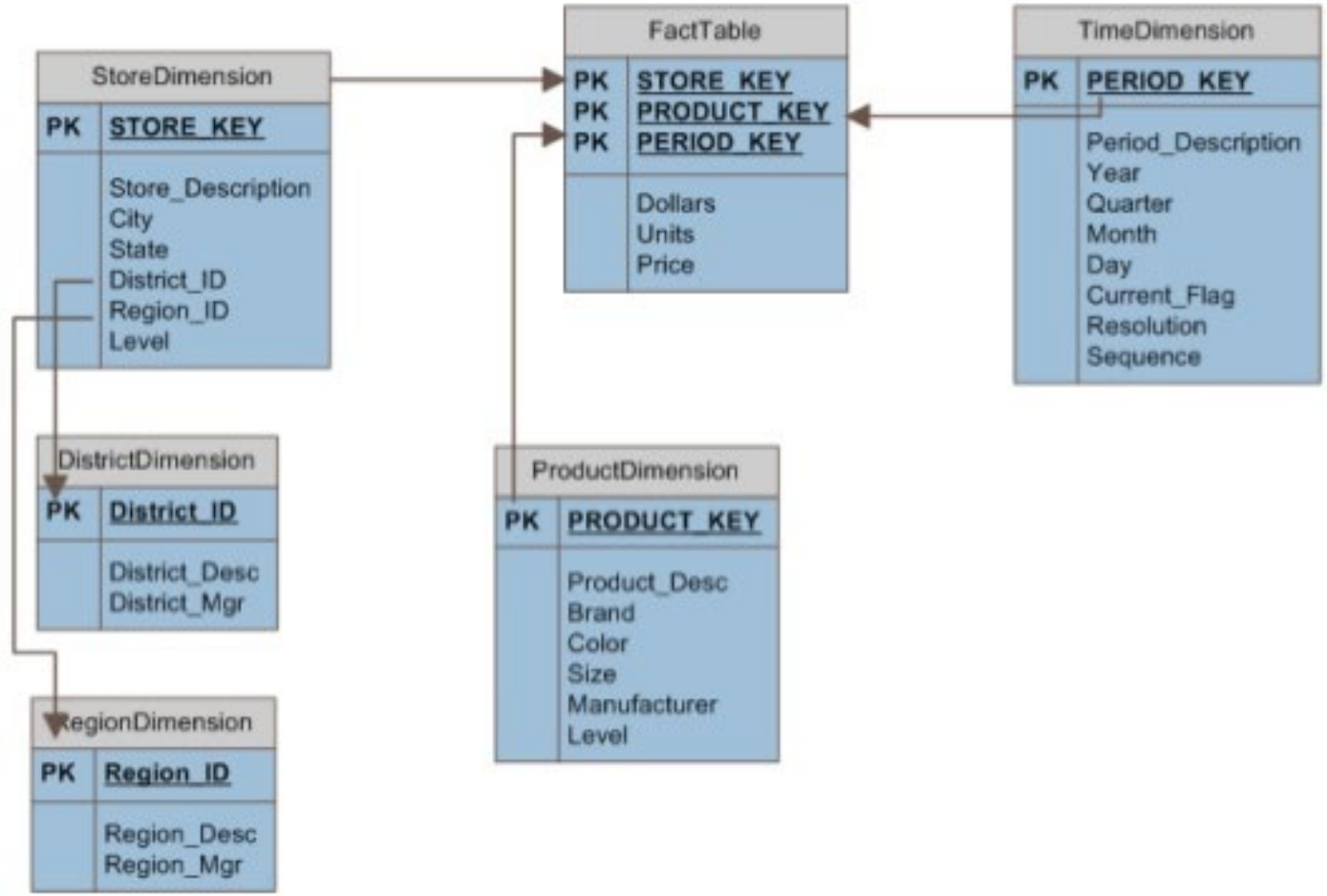


Snowflake Schema Model

- Direct use by some tools
- More flexible to change
- Provides for speedier data loading
- May become large and unmanageable
- Degrades query performance
- More complex metadata



Snowflake Schema

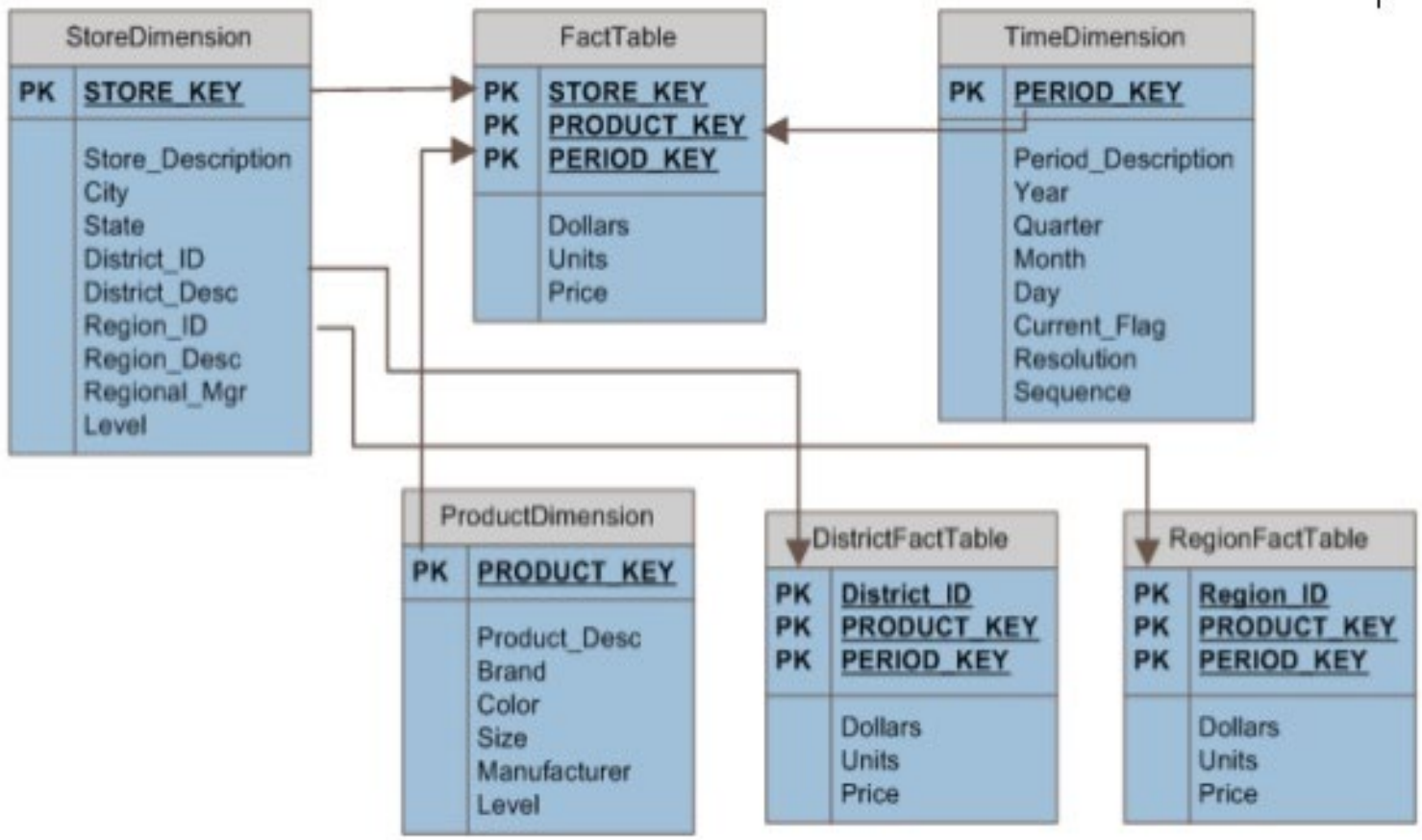


- The snowflake structure can reduce the effectiveness of browsing, since more joins will be needed to execute a query.

Fact constelation Schema Galaxy Schema

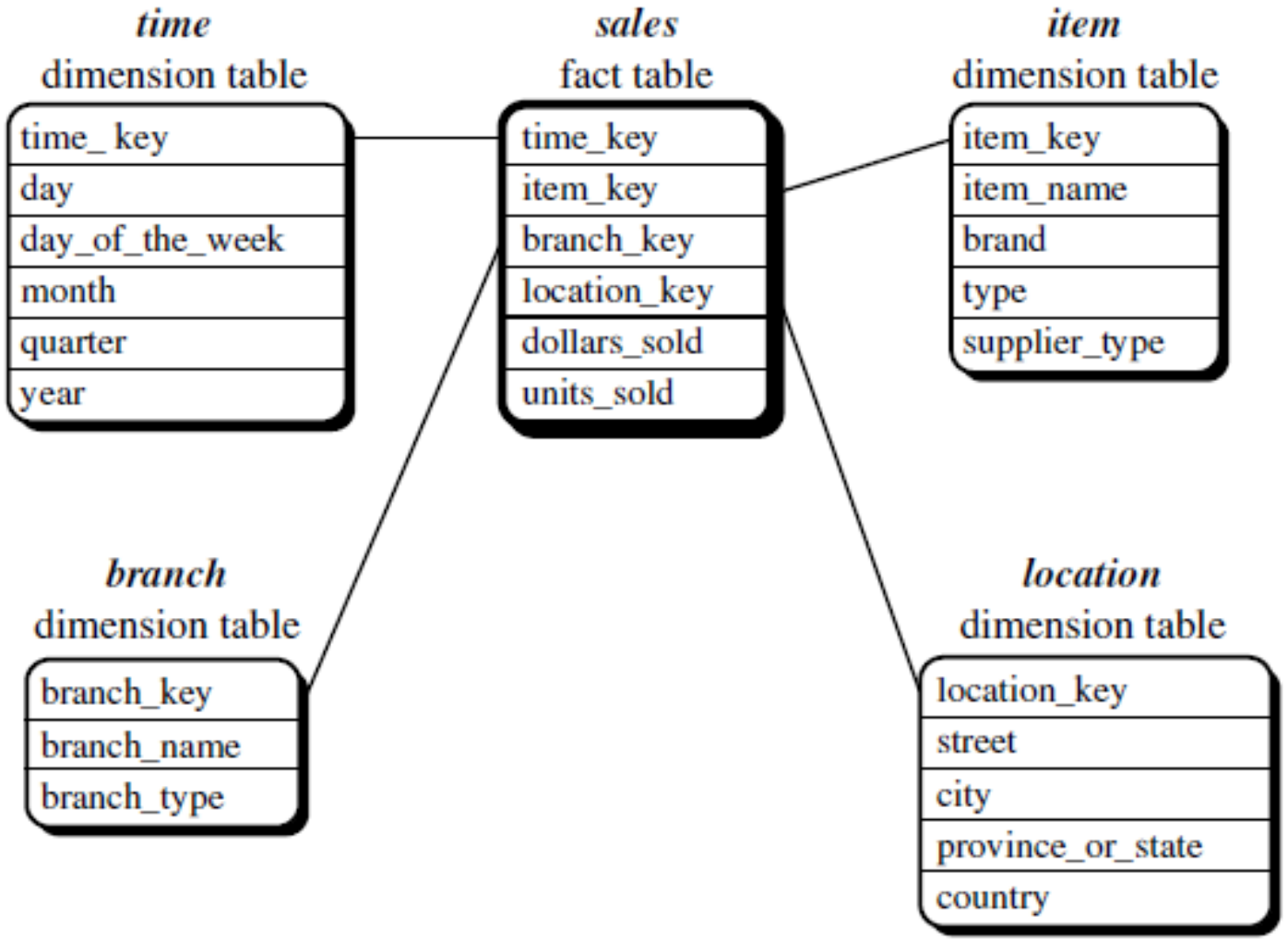
- Sophisticated applications may require multiple fact tables to share dimension tables. This kind of schema can be viewed as a collection of stars, and hence is call a galaxy schema or fact constellation

Fact Constellation Schema



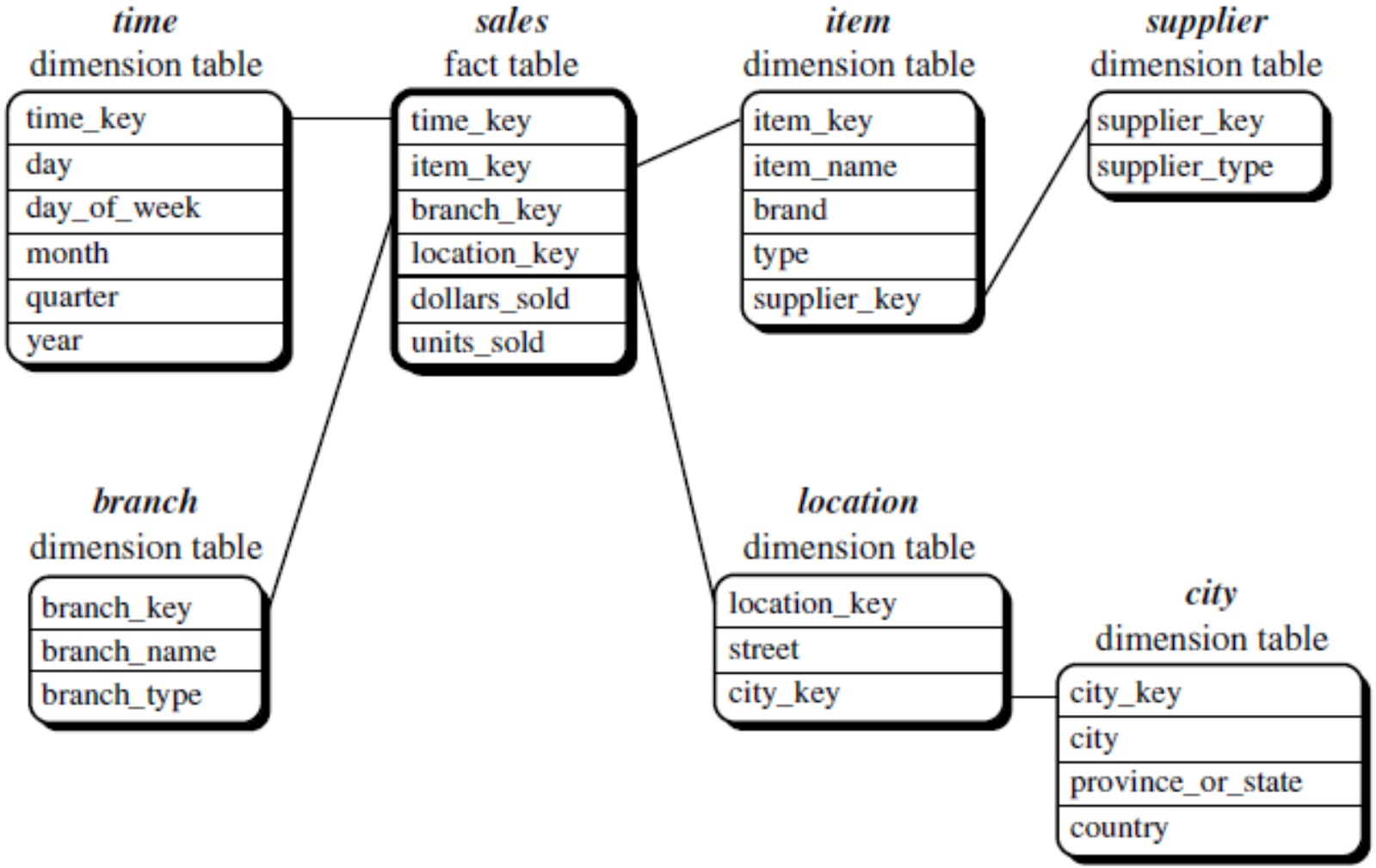
- For Data Marts, star or snowflake schema are commonly used, since both are generated toward modeling single subject
- Star schema is more popular and efficient

Example of Star Schema



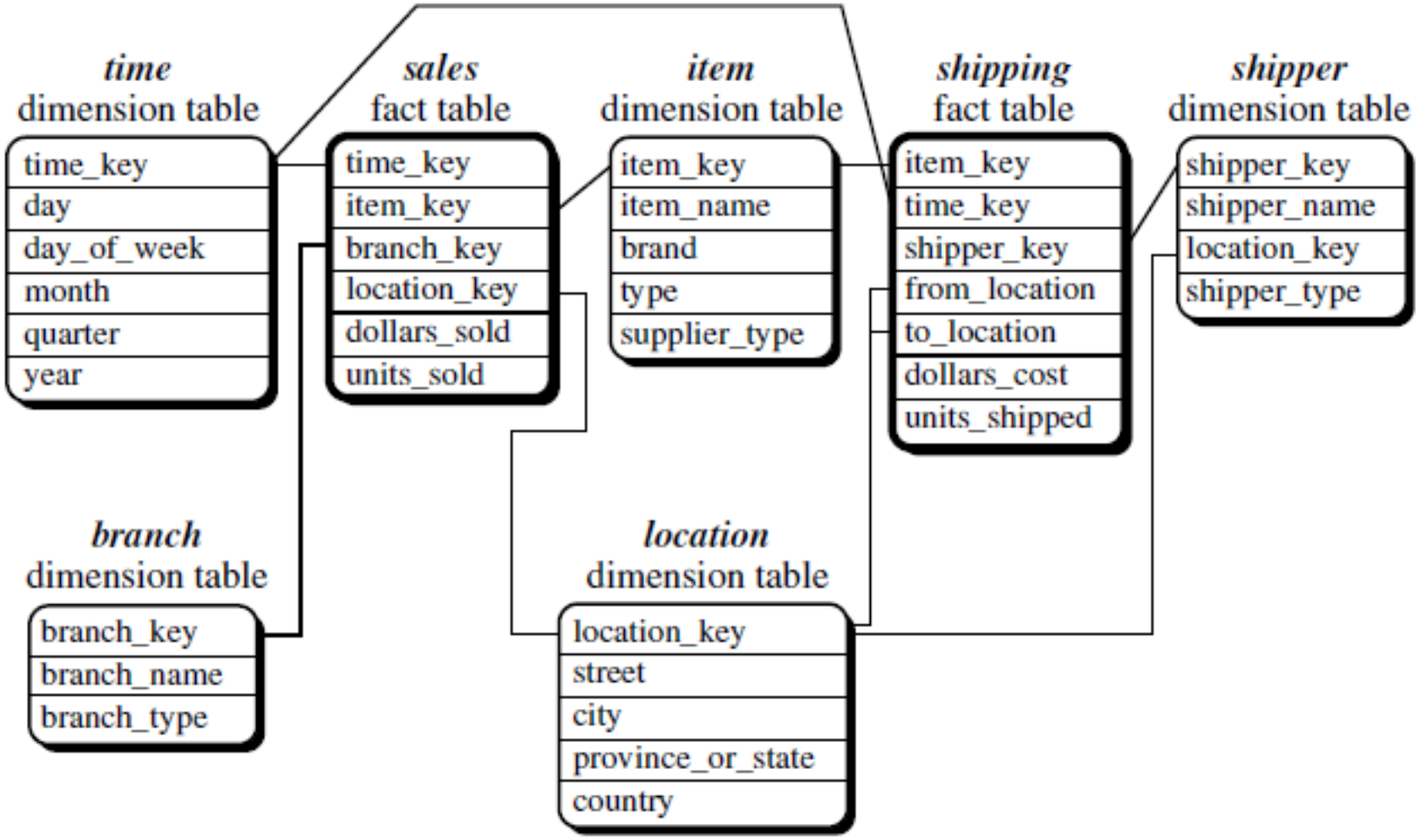
Star schema of a data warehouse for sales

Example of Snowflake Schema



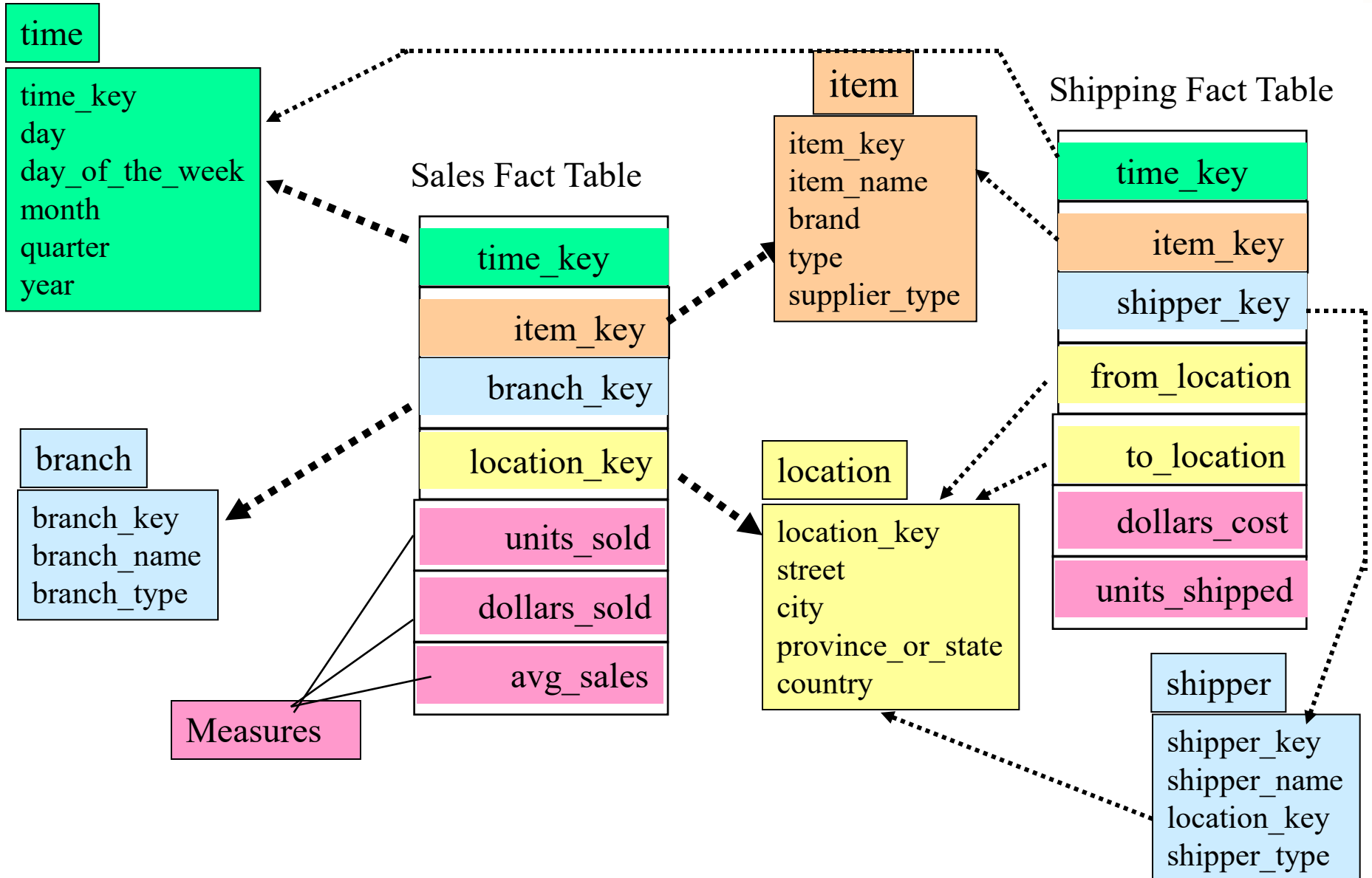
Snowflake schema of a data warehouse for sales

Example of Fact Constellation

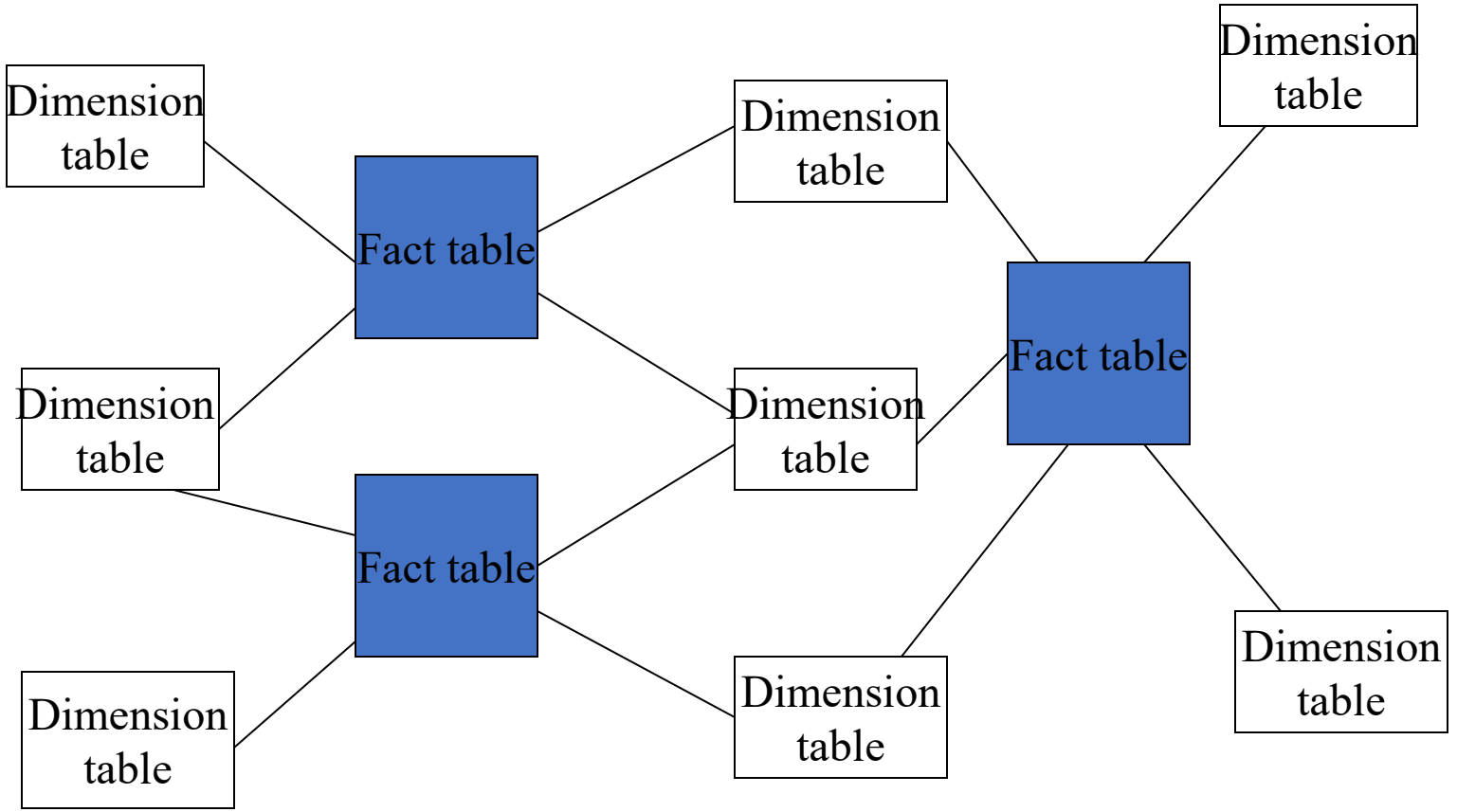


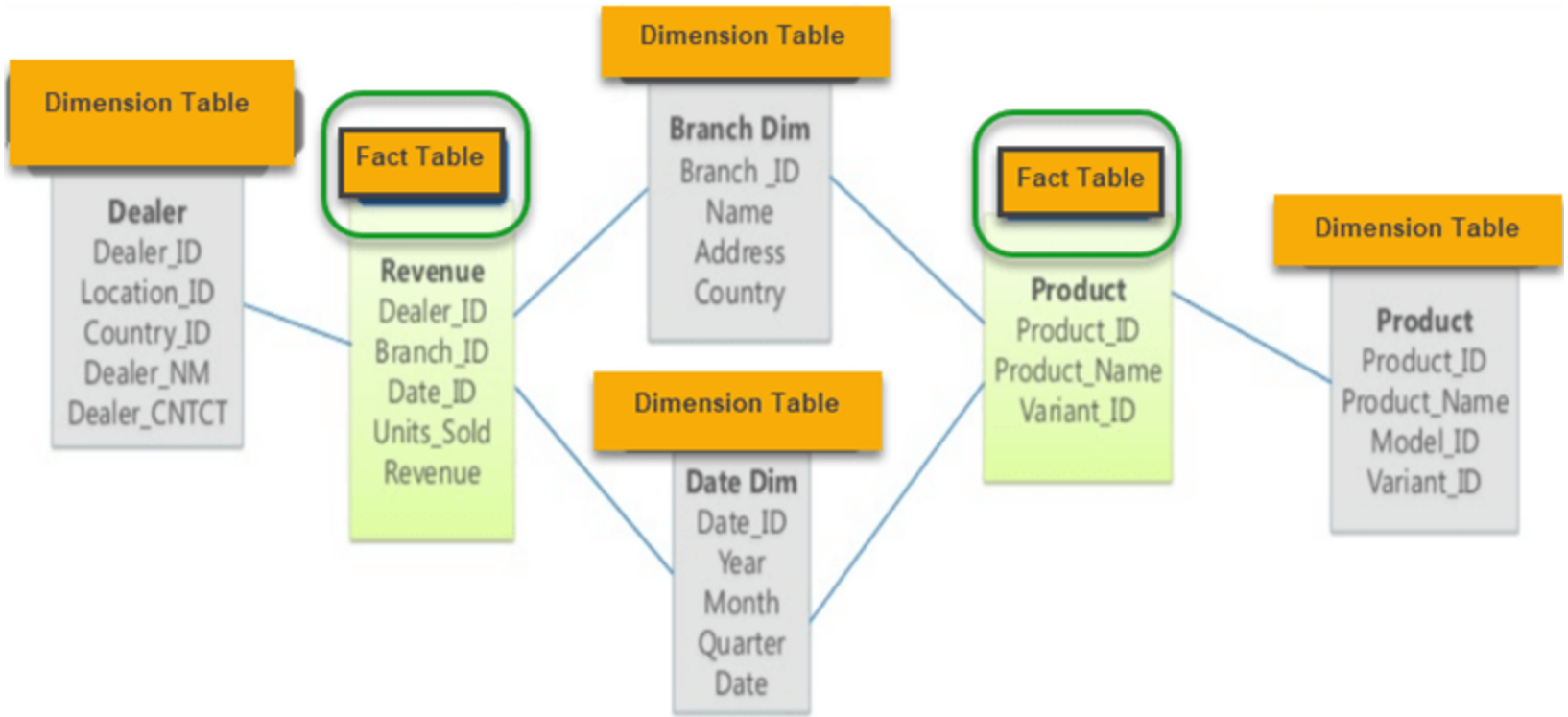
Fact constellation schema of a data warehouse for sales and shipping

Example of **Fact Constellation**

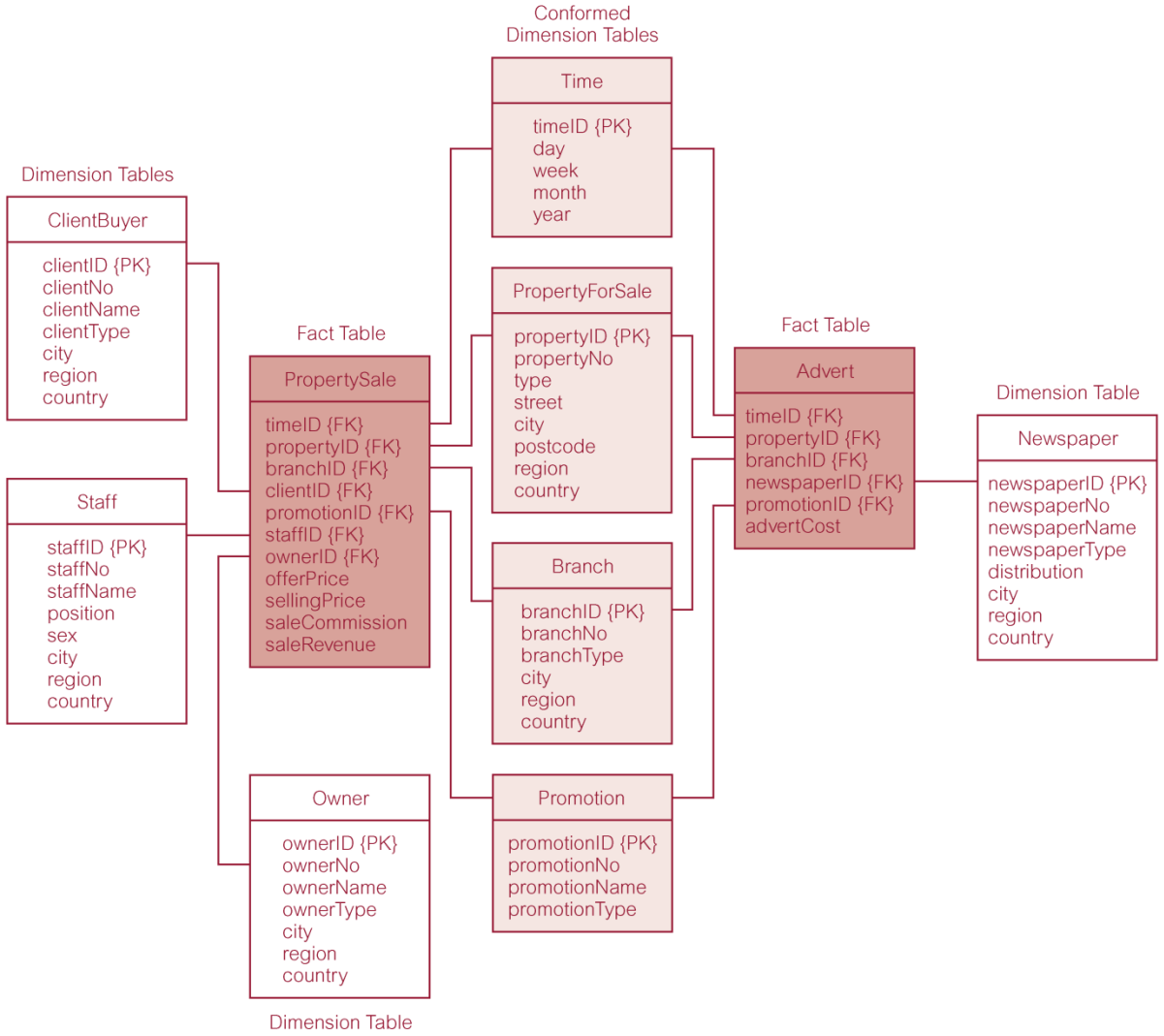


Families of Stars




Dimension Table
Dealer
Dealer_ID
Location_ID
Country_ID
Dealer_NM
Dealer_CNTCT
Fact Table
Revenue
Dealer_ID
Branch_ID
Date_ID
Units_Sold
Revenue
Dimension Table
Branch Dim
Branch _ID
Name
Address
Country
Dimension Table
Date Dim
Date_ID
Year
Month
Quarter
Date
Fact Table
Product
Product_ID
Product_Name
Variant_ID
Dimension Table
Product
Product_ID
Product_Name
Model_ID
Variant_ID

Star schemas for property sales and property advertising



Blog

- https://bimanu.de/blog/galaxy-schema/

Fragen?