COURSE NAME:
DATA WAREHOUSING & DATA MINING

LECTURE 3
TOPICS TO BE COVERED:

- Data warehouse Process
- Architecture
- Schemas for Multidimensional Database:
  - stars
  - snowflakes and
  - fact constellations.

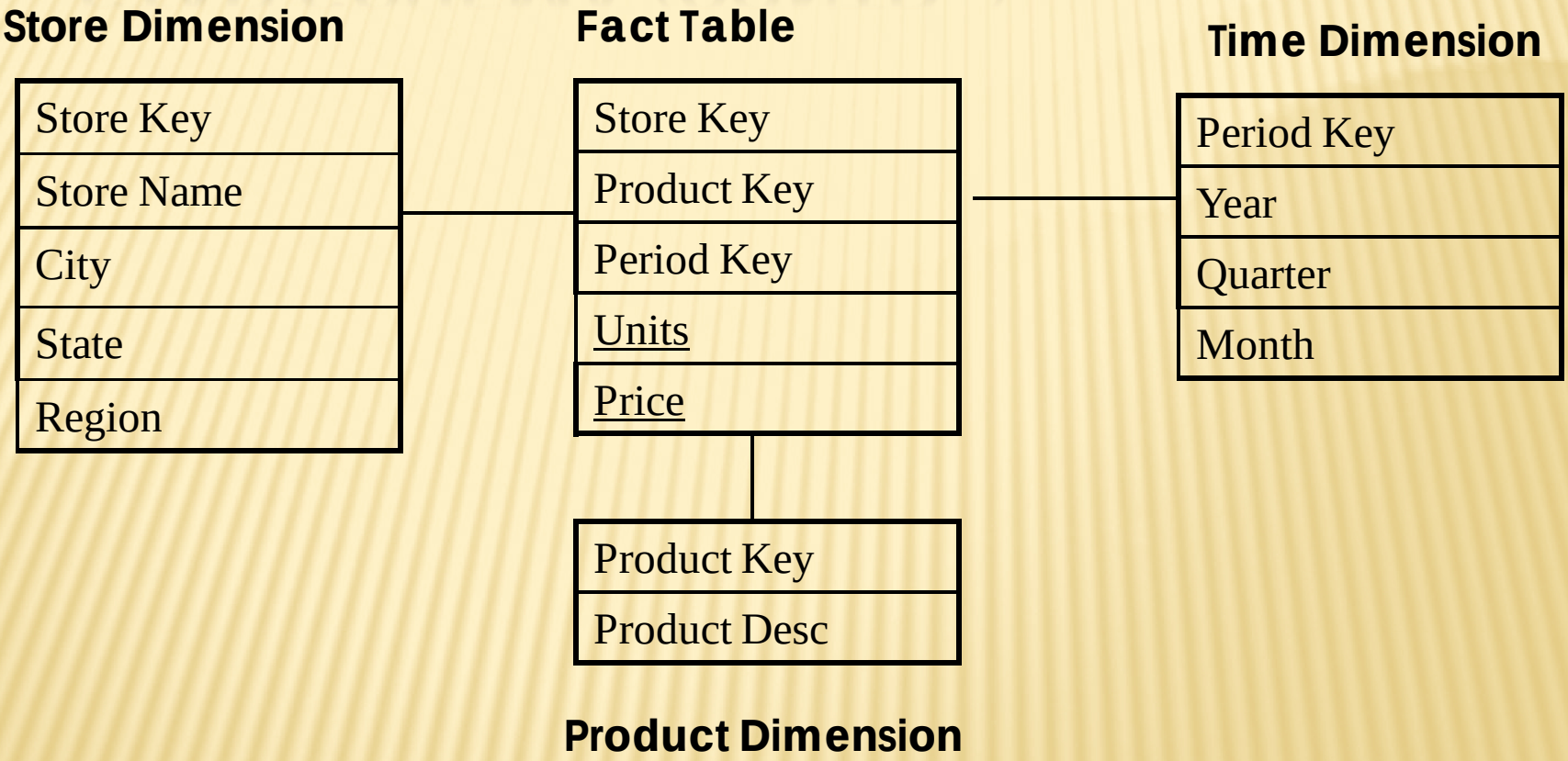
DATA WAREHOUSE SCHEMA

- Star Schema
- Snowflake Schema
- Fact Constellation Schema

STAR SCHEMA

- A single,large and central fact table and one table for each dimension.
- Every fact points to one tuple in each of the dimensions and has additional attributes.
- Does not capture hierarchies directly.

STAR SCHEMA (CONTD..)

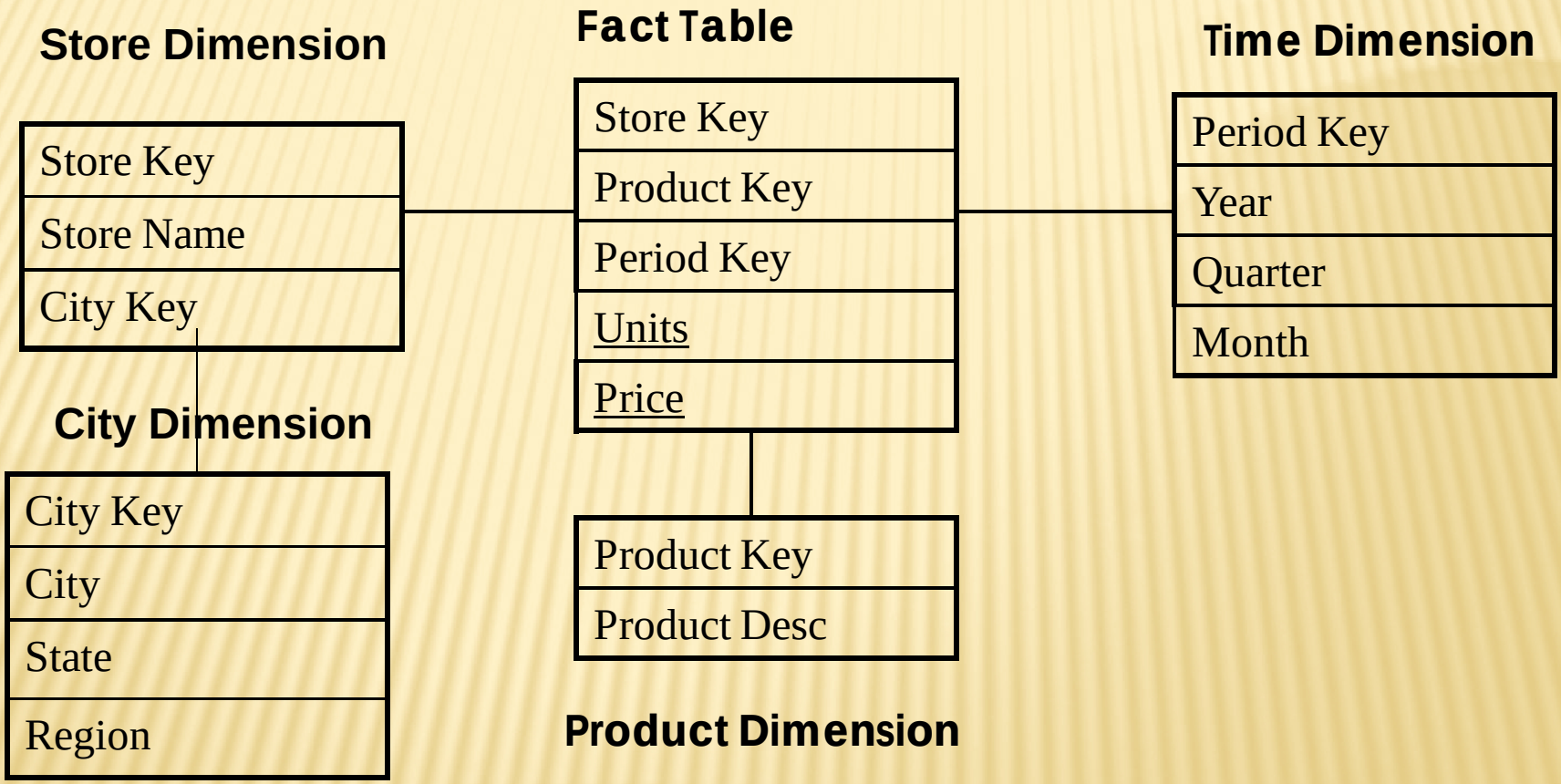


Benefits: **Easy to understand, easy to define hierarchies, reduces no. of physical joins.**

SNOWFLAKE SCHEMA

- Variant of star schema model.
- A single,large and central fact table and one or more tables for each dimension.
- Dimension tables are normalized i.e. split dimension table data into additional tables

SNOWFLAKE SCHEMA (CONTD..)

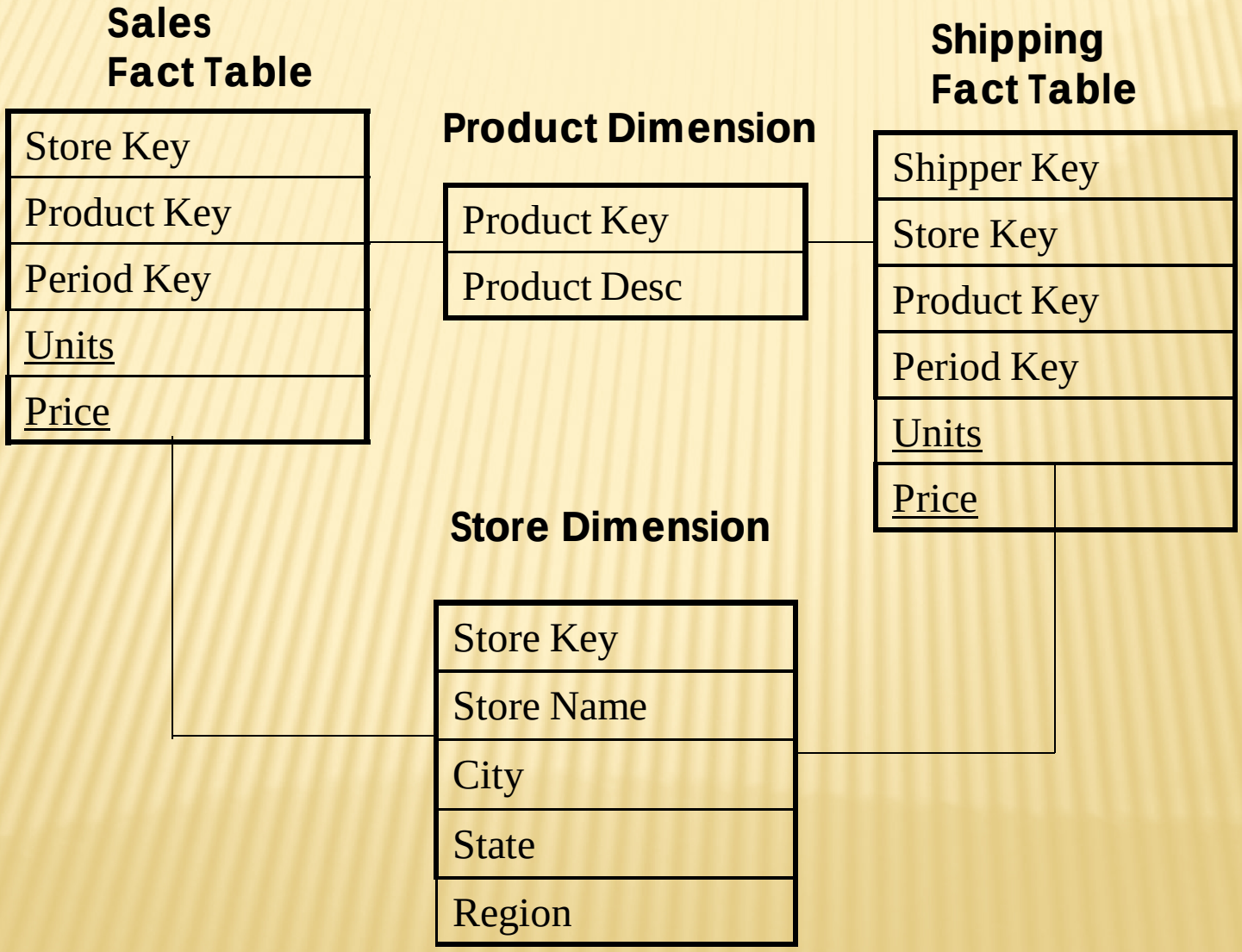


Drawbacks: Time consuming joins,report generation slow

FACT CONSTELLATION

- Multiple fact tables share dimension tables.
- This schema is viewed as collection of stars hence called galaxy schema or fact constellation.
- Sophisticated application requires such schema.

FACT CONSTELLATION (CONTD..)



DIMENSIONAL DATA MODELING

- E-R model
  - Symmetric
  - Divides data into many entities
  - Describes entities and relationships
  - Seeks to eliminate data redundancy
  - Good for high transaction performance
- Dimensional model
  - Asymmetric
  - Divides data into dimensions and facts
  - Describes dimensions and measures
  - Encourages data redundancy
  - Good for high query performance

FACTS/DIMENSIONS

- Fact
  - Central, dominant table
  - Multi-part primary key
  - Holds millions & billions of records
  - Links directly to dimensions
  - Stores business measures
  - Constantly varying data

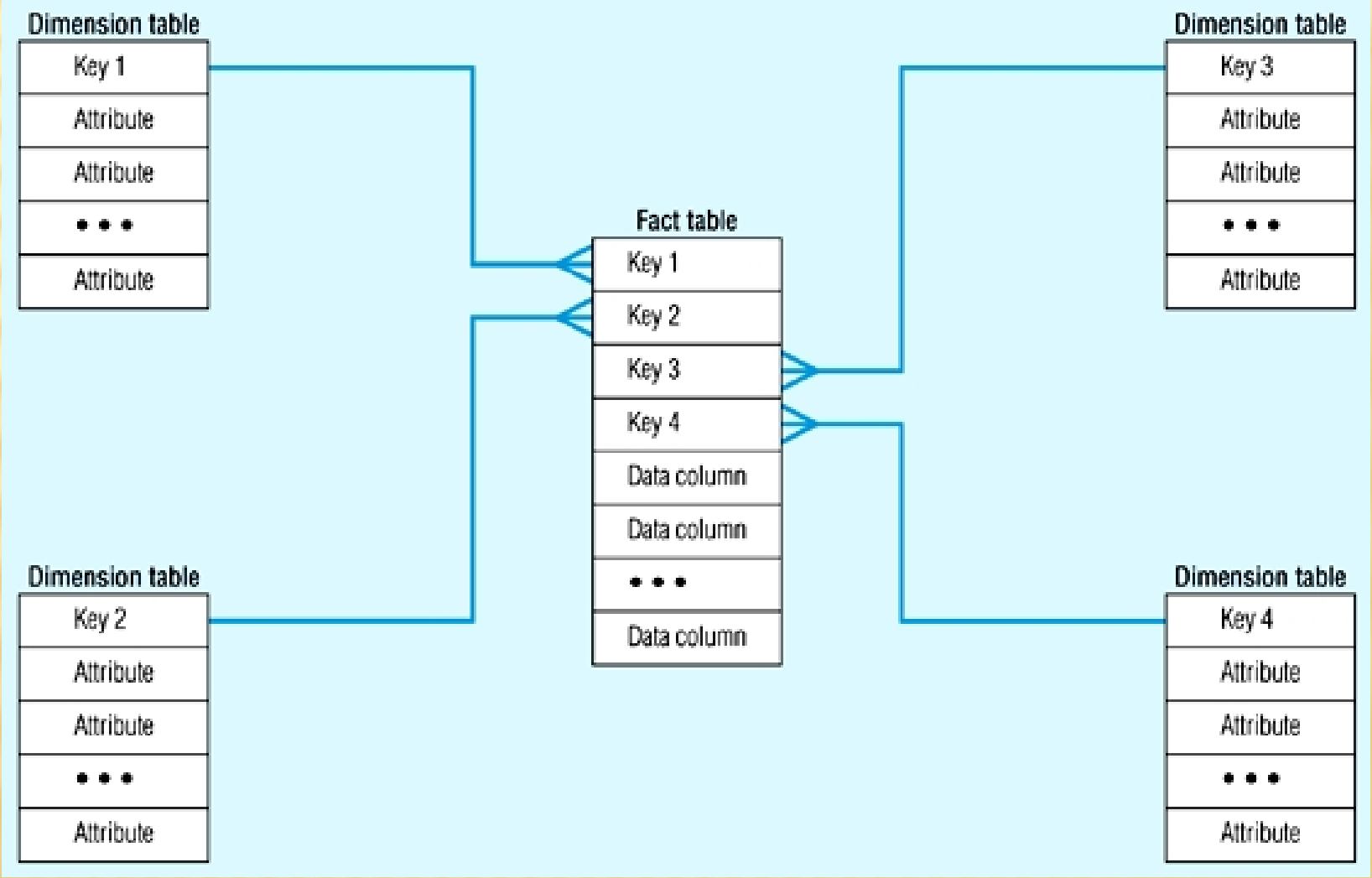
FACTS/DIMENSIONS (CONTD.)

- Dimensions
  - Single join to the fact table (single primary key)
  - Stores business attributes
  - Attributes are textual in nature
  - Organized into hierarchies
  - More or less constant data
  - E.g. Time, Product, Customer, Store, etc.

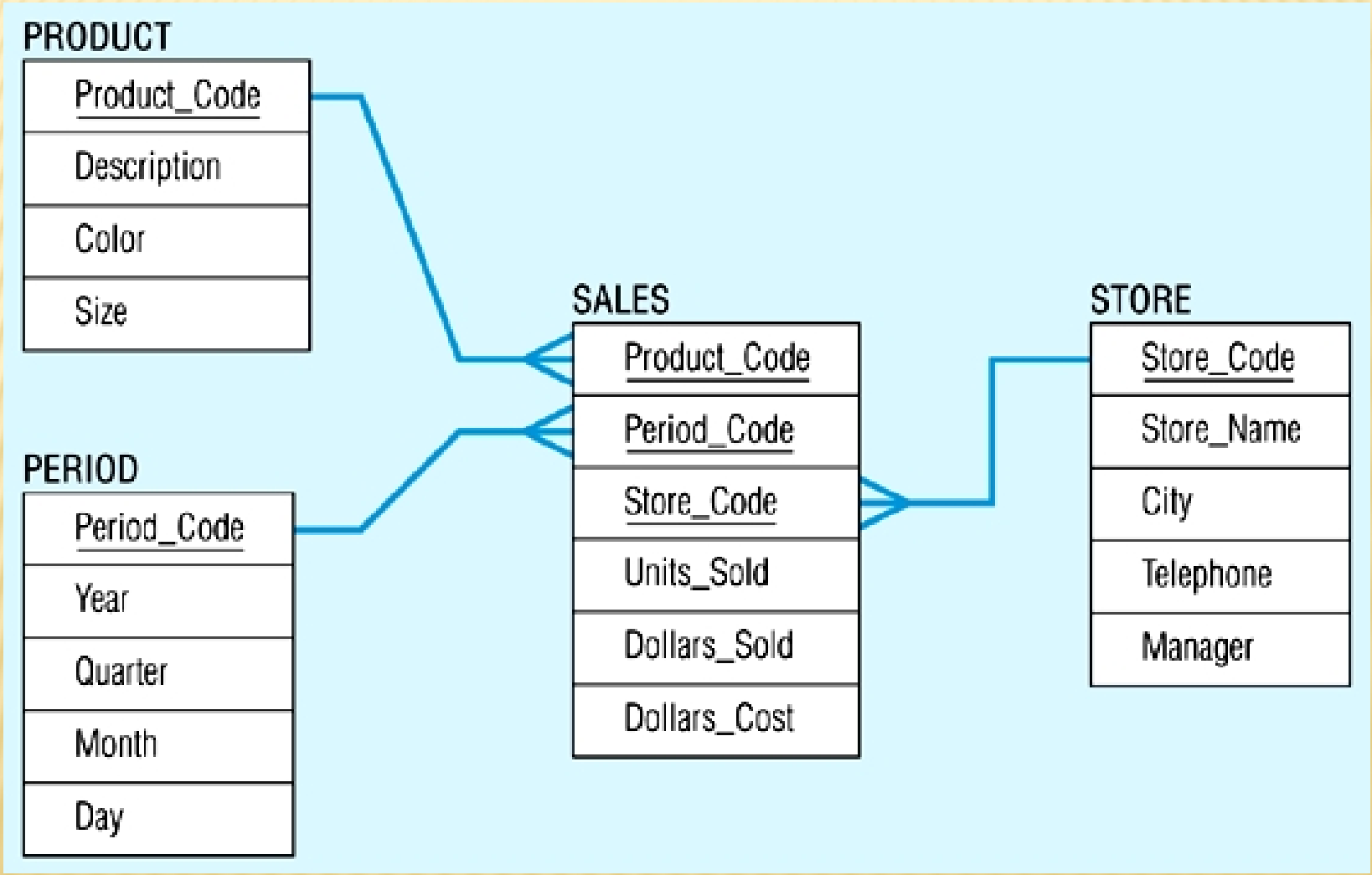
STAR/SNOWFLAKE SCHEMA

- Star schema
  - Fact surrounded by 4-15 dimensions
  - Dimensions are de-normalized
- Snowflake schema
  - Star schema with secondary dimensions
  - Don’t snowflake for saving space
  - Snowflake if secondary dimensions have many attributes

STAR SCHEMA


Dimension table
Key 1
Attribute
Attribute
• • •
Attribute
Dimension table
Key 3
Attribute
Attribute
• • •
Attribute
Fact table
Key 1
Key 2
Key 3
Key 4
Data column
Data column
• • •
Data column
Dimension table
Key 2
Attribute
Attribute
• • •
Attribute
Dimension table
Key 4
Attribute
Attribute
• • •
Attribute

STAR SCHEMA EXAMPLE

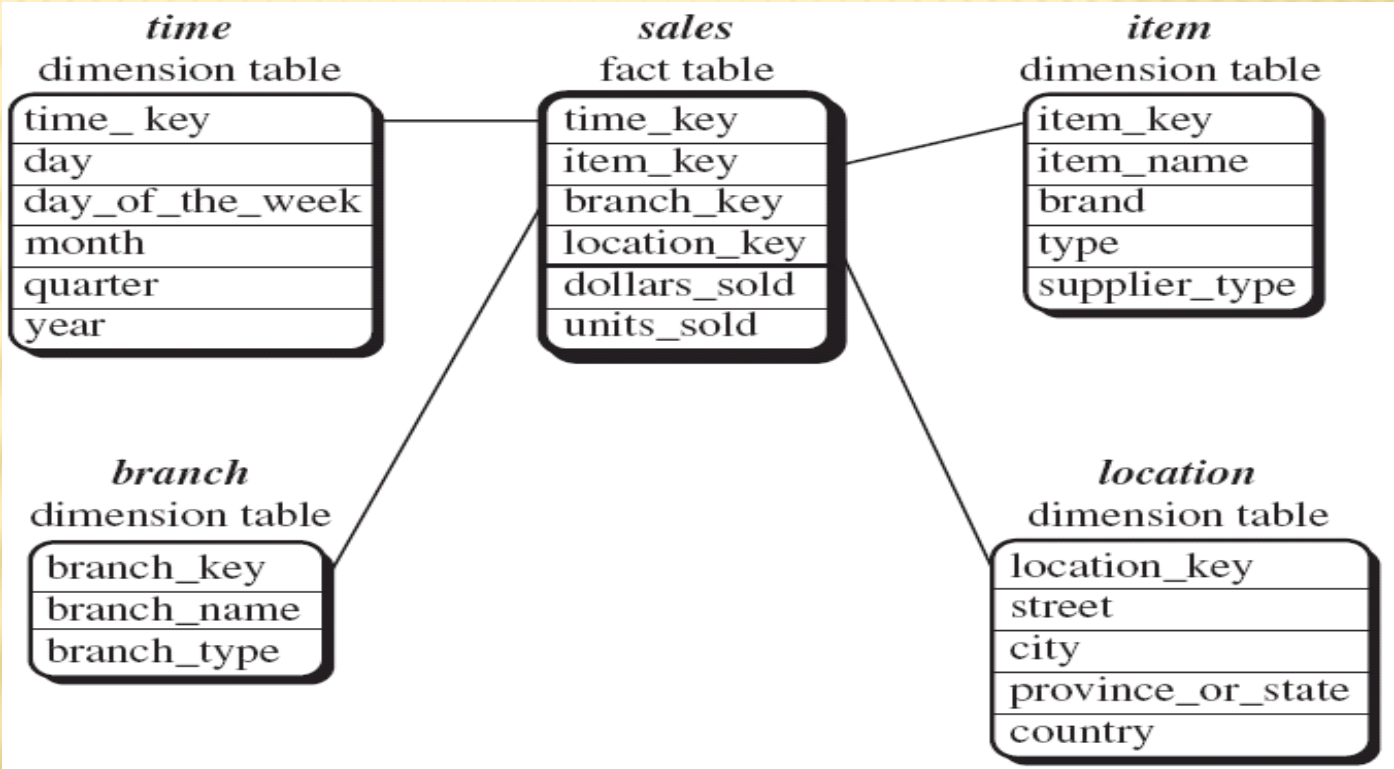


CONCEPTUAL MODELING OF DATA WAREHOUSES

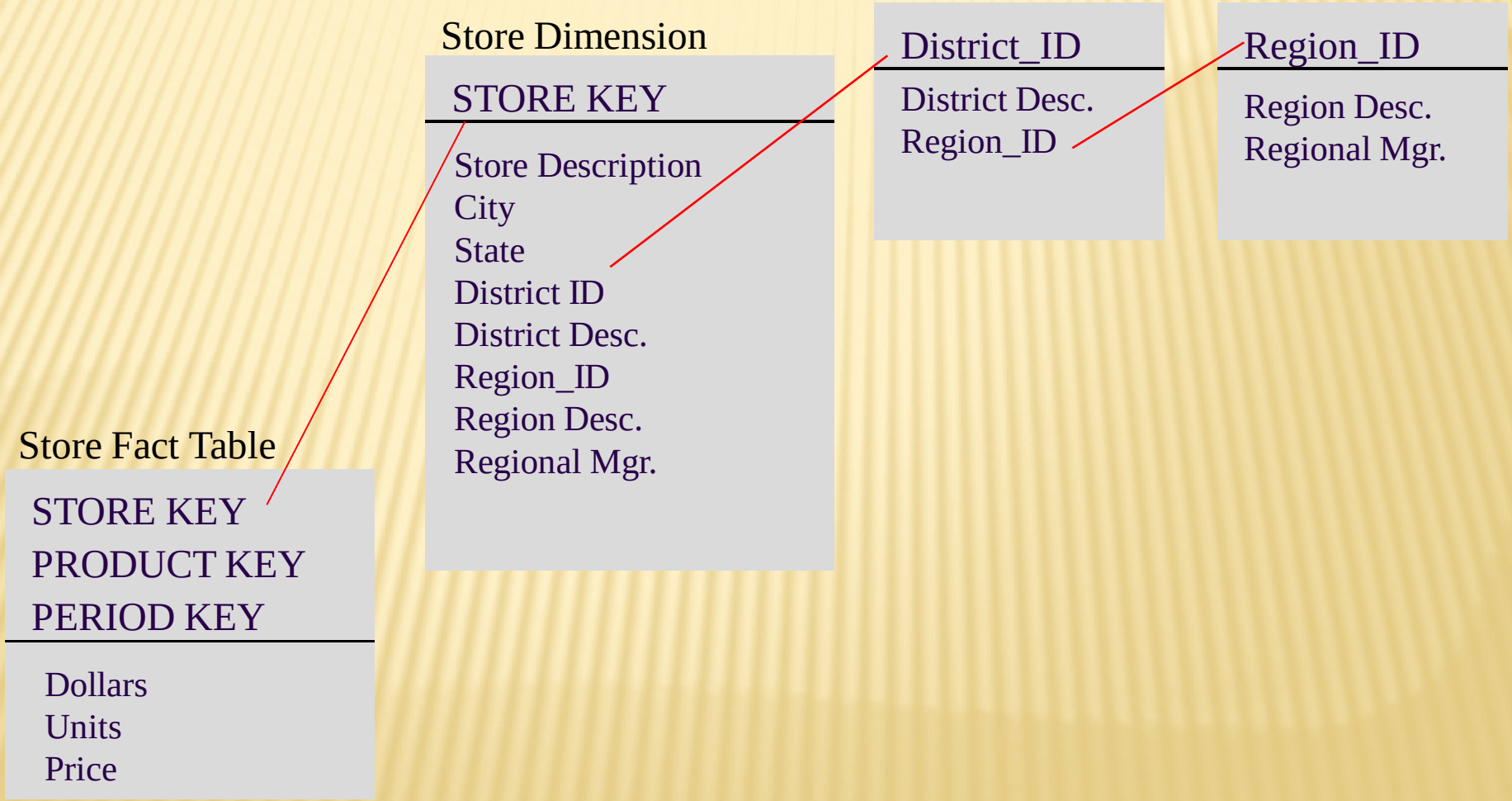
- Star schema: A fact table in the middle connected to a set of dimension tables

- It contains:
  - A large central table (fact table)
  - A set of smaller attendant tables (dimension table), one for each dimension

STAR SCHEMA


time
dimension table
time_ key
day
day_of_the_week
month
quarter
year
sales
fact table
time_key
item_key
branch_key
location_key
dollars_sold
units_sold
item
dimension table
item_key
item_name
brand
type
supplier_type
branch
dimension table
branch_key
branch_name
branch_type
location
dimension table
location_key
street
city
province_or_state
country

SNOWFLAKE SCHEMA EXAMPLE

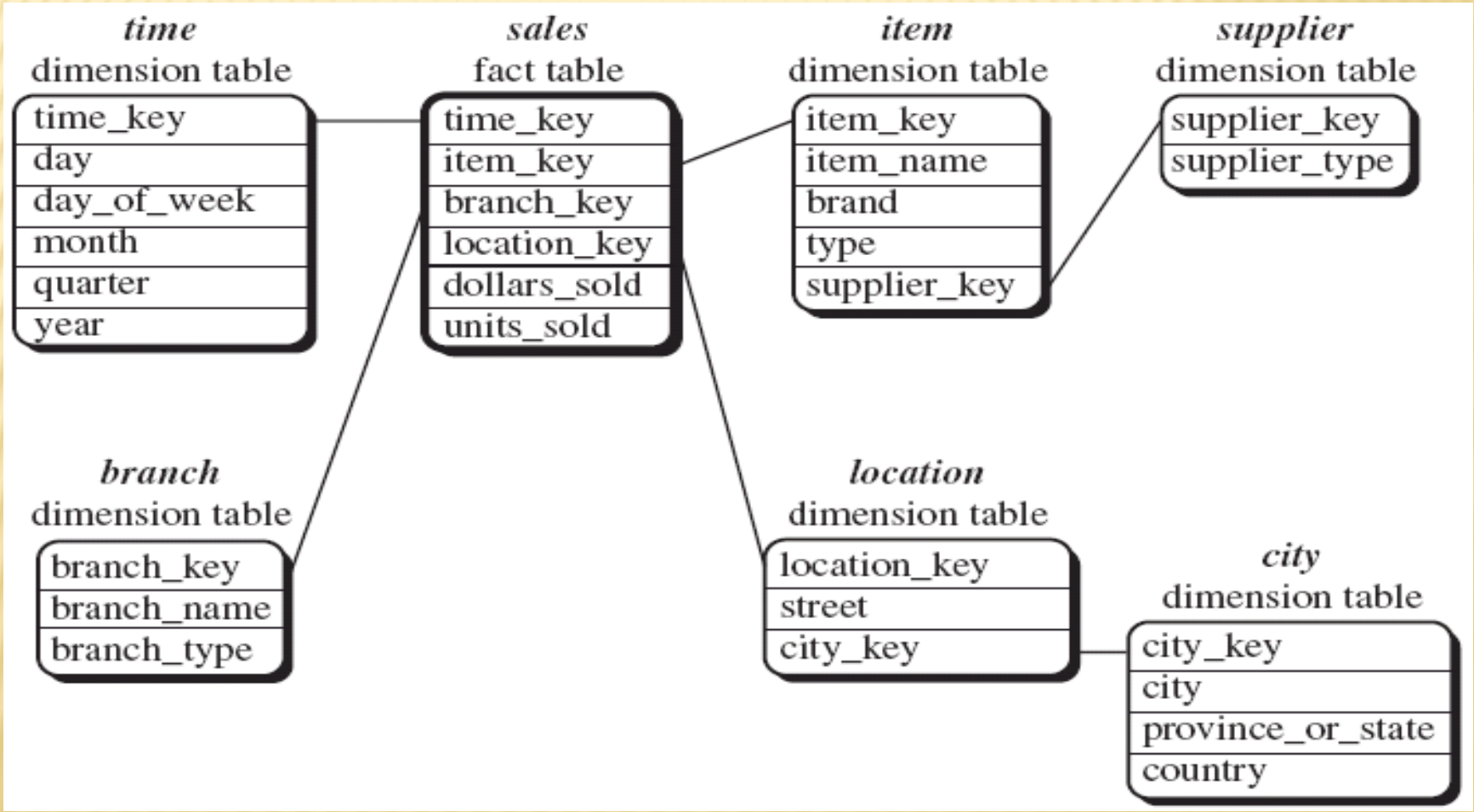


CONCEPTUAL MODELING OF DATA WAREHOUSES

- Snowflake schema: A refinement of star schema where some dimensional hierarchy is further splitting (normalized) into a set of smaller dimension tables, forming a shape similar to snowflake

- However, the snowflake structure can reduce the effectiveness of browsing, since more joins will be needed

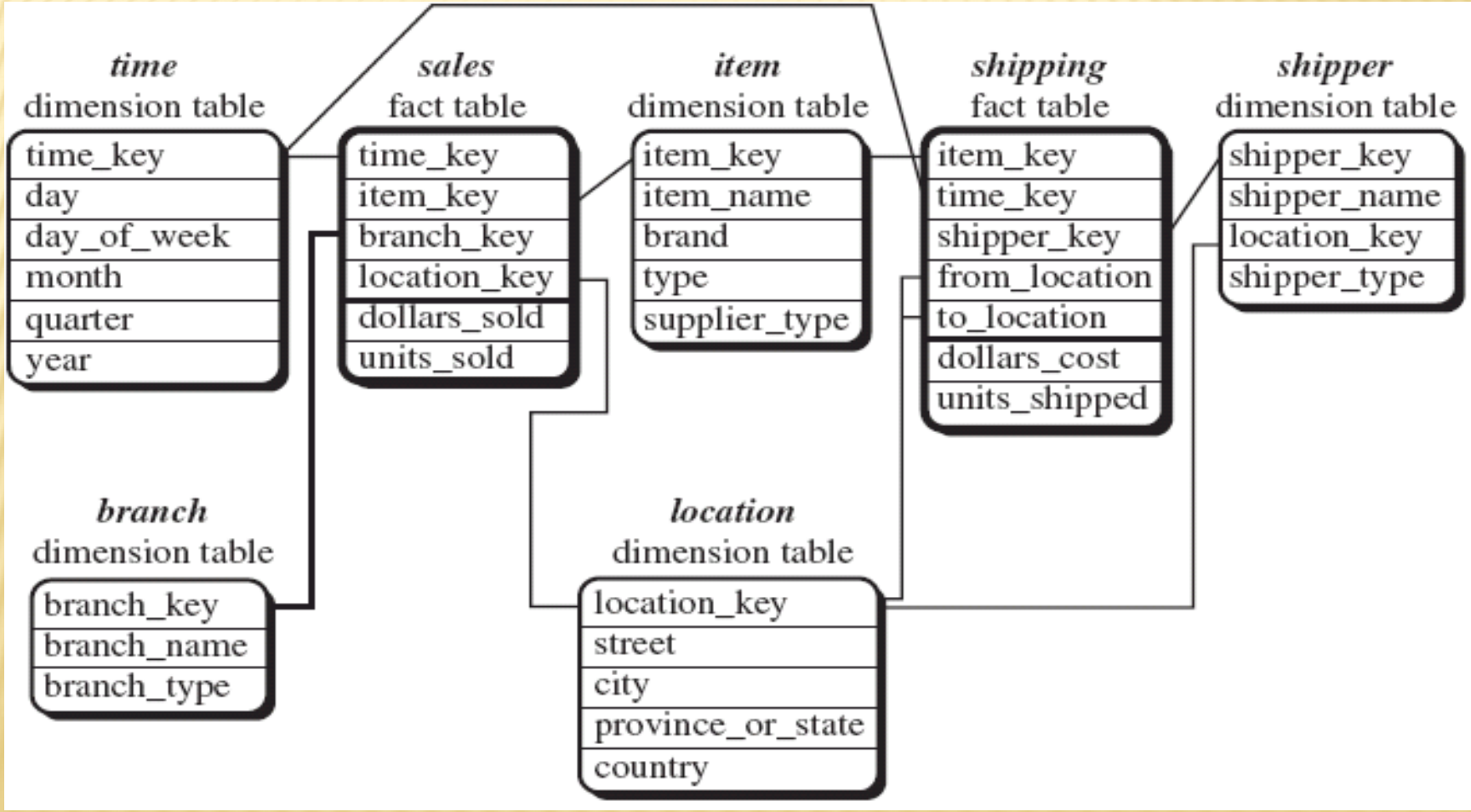
SNOWFLAKE SCHEMA



CONCEPTUAL MODELING OF DATA WAREHOUSES

- Fact constellations: Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called galaxy schema or fact constellation

FACT CONSTELLATIONS



CONCEPT HIERARCHIES

- A Concept Hierarchy defines a sequence of mappings from a set of low-level concepts to high-level
- Consider a concept hierarchy for the dimension “Location”

CONCEPT HIERARCHIES