Principles of Dimensional Modeling

Learning Outcomes

- Clearly understand how the requirements definition determines data design
- Introduce dimensional modeling and contrast it with entity-relationship modeling
- Review the basics of the STAR schema
- Find out what is inside the fact table and inside the dimension tables
- Determine the advantages of the STAR schema for data warehouses

Topics Covered

- FROM REQUIREMENTS TO DATA DESIGN
- THE STAR SCHEMA
- STAR SCHEMA KEYS
- ADVANTAGES OF THE STAR SCHEMA

FROM REQUIREMENTS TO DATA DESIGN

- The requirements definition completely drives the data design for the data warehouse.
- Data design consists of putting together the data structures.

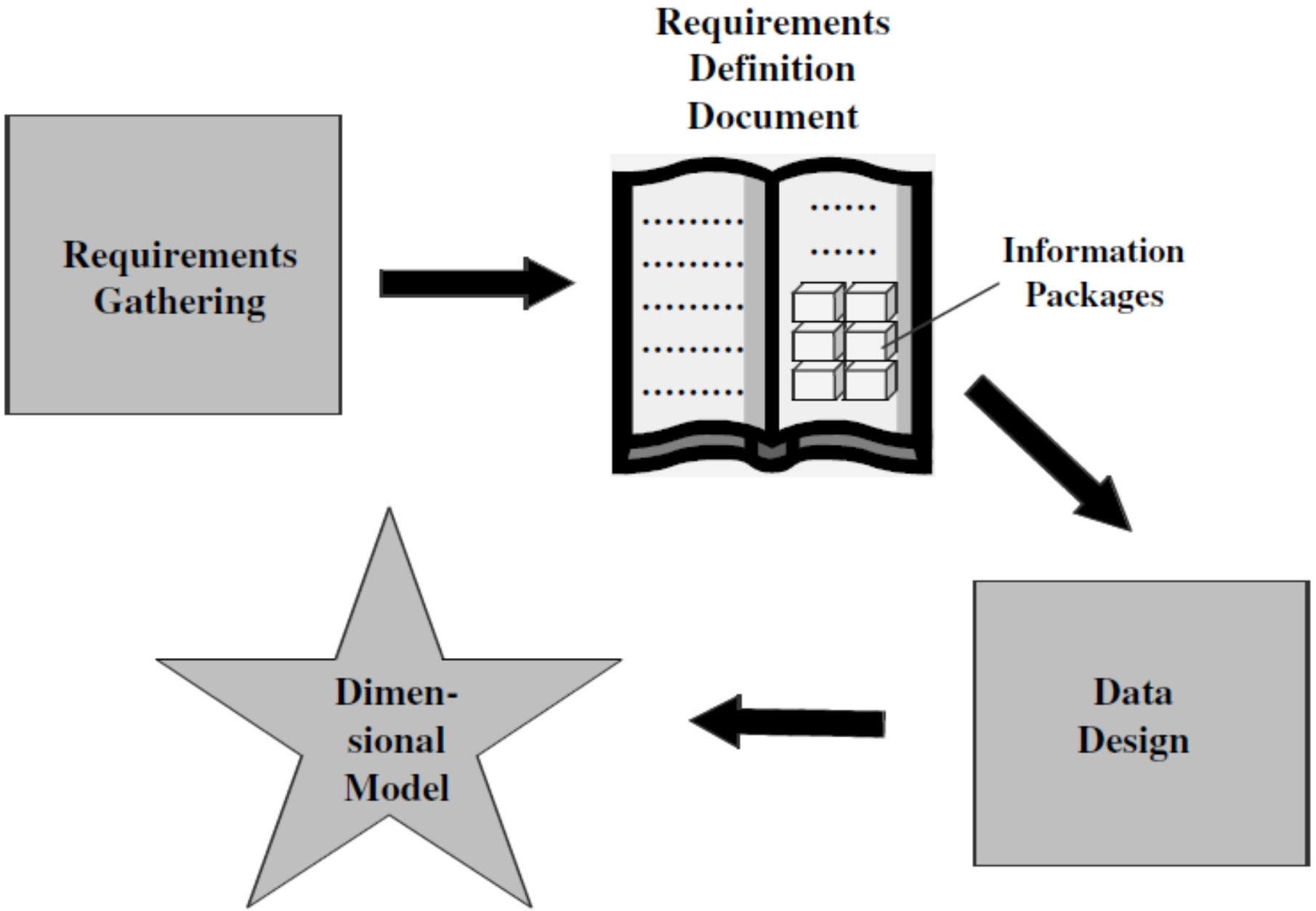


Figure 10-1 From requirements to data design.

Design Decisions

Some of the design decisions to make:

- **Choosing the process.** Selecting the subjects.
- **Choosing the grain.** Level of detail for the data.
- **Identifying and conforming the dimensions.** Choosing the business dimensions(such as product, market, time, etc.)
- **Choosing the facts.** Selecting the metrics (e.g. product sale units, dollar sales, dollar revenue)
- **Choosing the duration of the database.** Determining the range of historical data.

Dimensional Modeling Basics

- Dimensional modeling gets its name from the business dimensions we need to incorporate into the logical data model.
- It is a logical design technique to structure the business dimensions and the metrics that are analyzed along these dimensions.

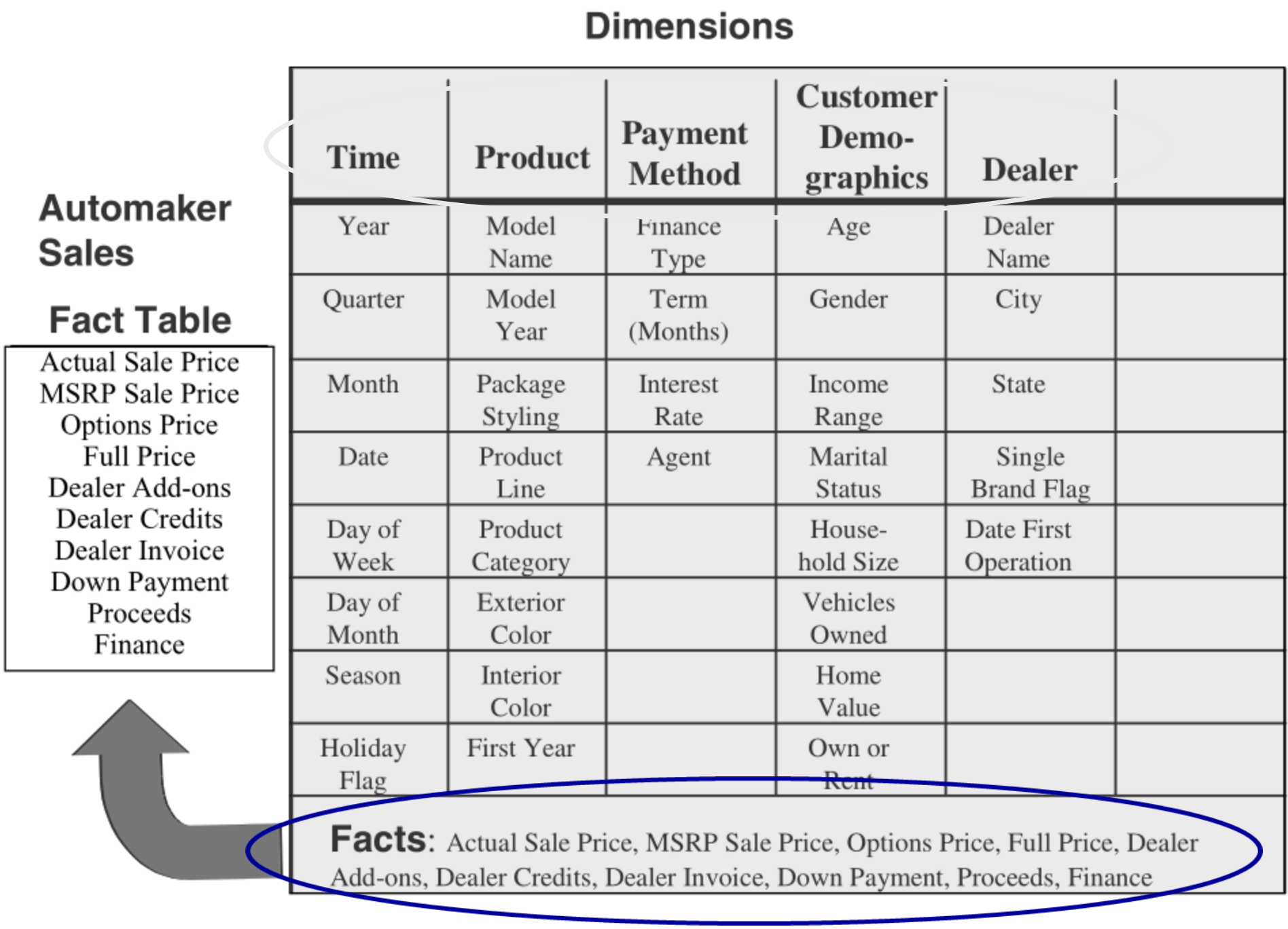


Figure 10-2 Formation of the automaker sales fact table.

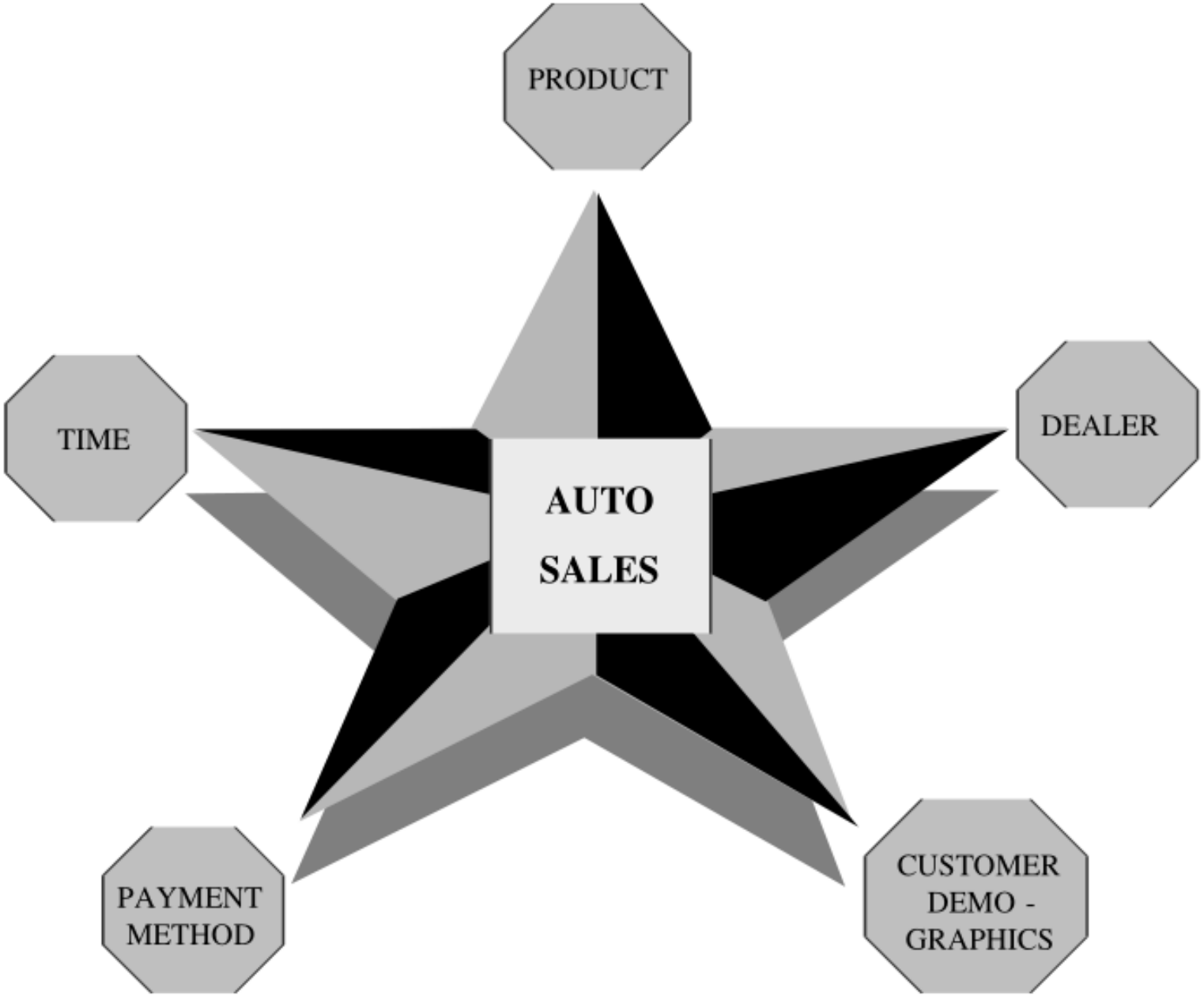


Figure 10-4 STAR schema for automaker sales.

ER Modeling vs. Dimensional Modeling

- OLTP systems capture details of events or transactions
- OLTP systems focus on individual events
- An OLTP system is a window into micro-level transactions
- Picture at detail level necessary to run the business
- Suitable only for questions at transaction level
- Data consistency, non-redundancy, and efficient data storage critical

Entity-Relationship Modeling

Removes data redundancy
Ensures data consistency
Expresses microscopic relationships

ER Modeling vs. Dimensional Modeling

- DW meant to answer questions on overall process
- DW focus is on how managers view the business
- DW reveals business trends
- Information is centered around a business process
- Answers show how the business measures the process
- The measures to be studied in many ways along several business dimensions

Dimensional Modeling

Captures critical measures
Views along dimensions
Intuitive to business users

STAR Schema: Example

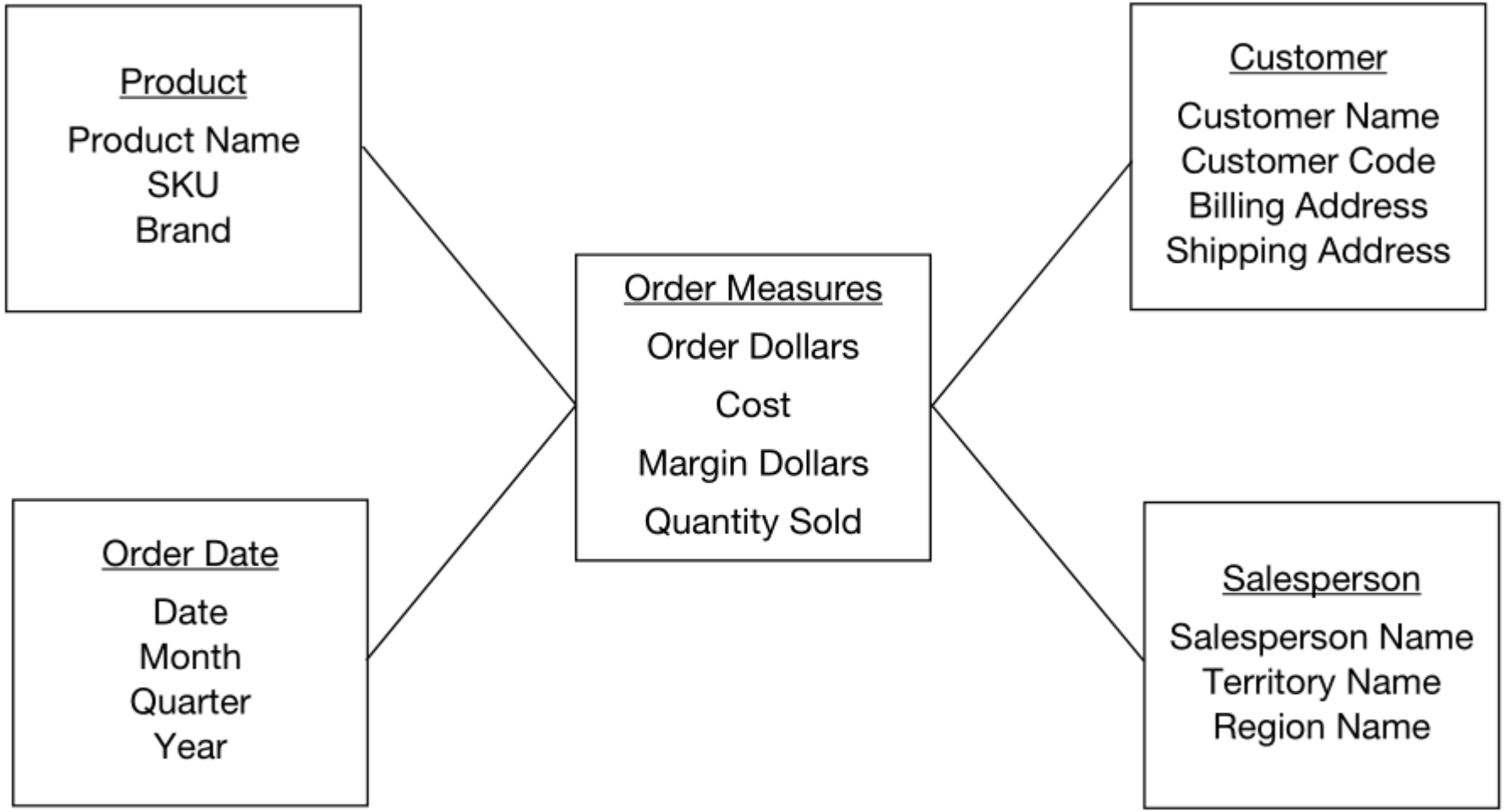
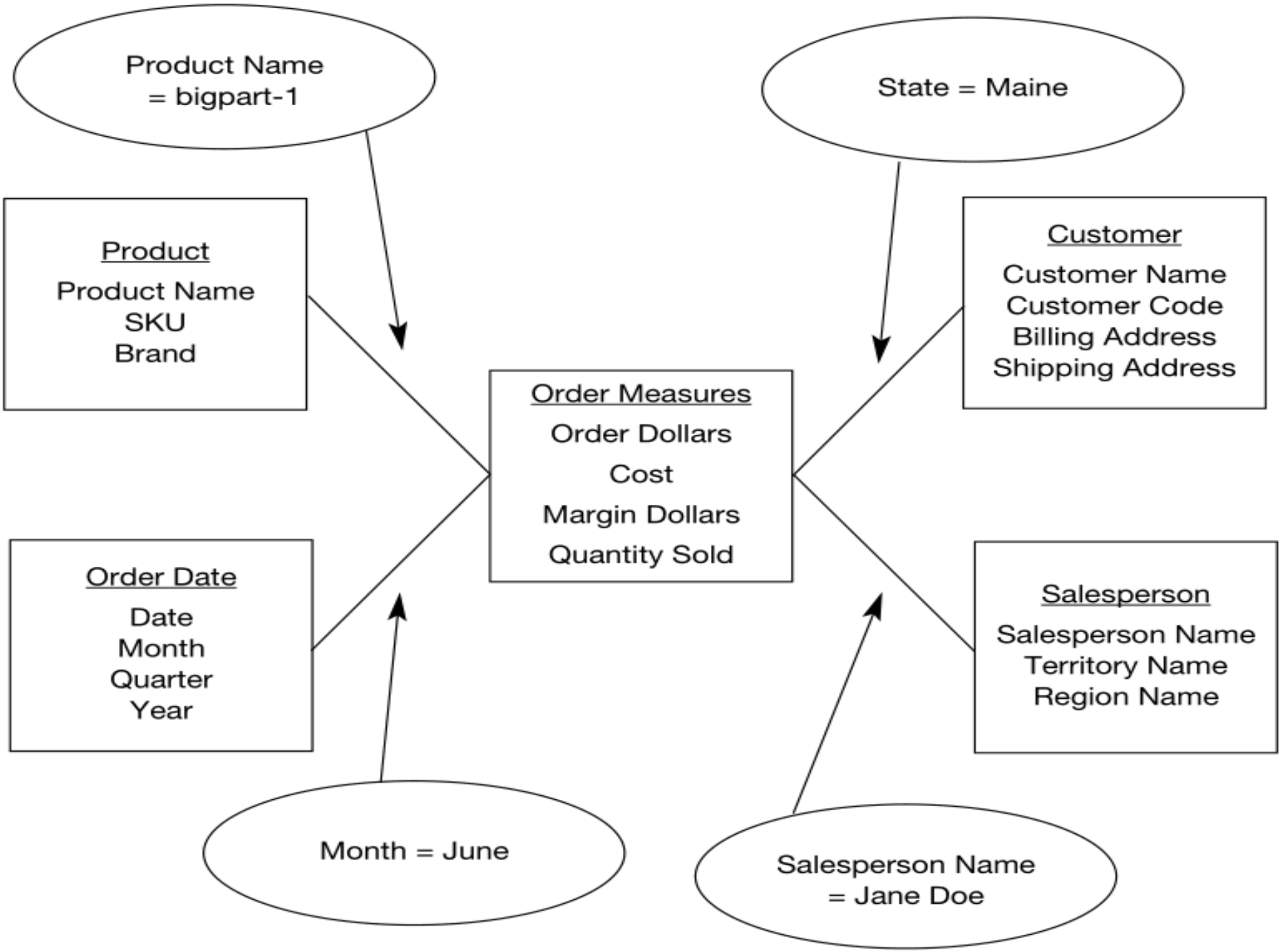
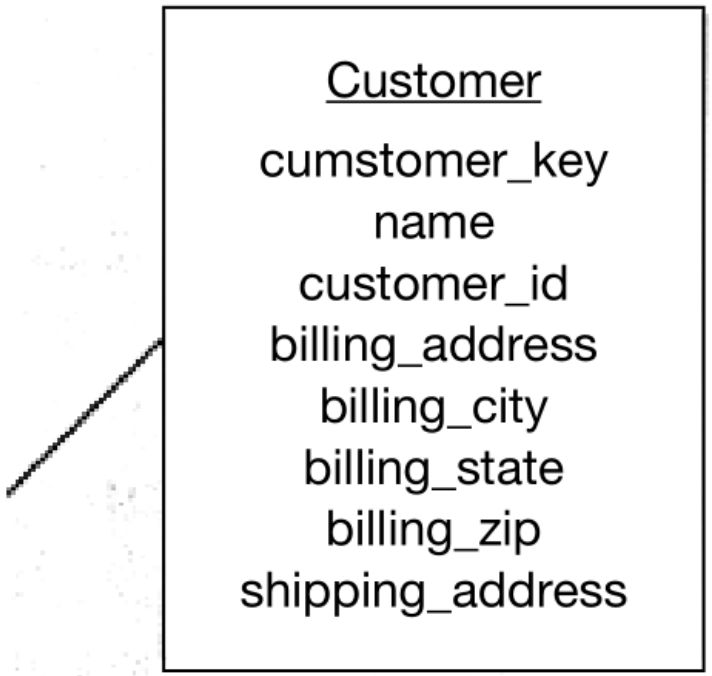


Figure 10-7 Simple STAR schema for orders analysis.


Product Name
= bigpart-1
State = Maine
Product
Product Name
SKU
Brand
Customer
Customer Name
Customer Code
Billing Address
Shipping Address
Order Measures
Order Dollars
Cost
Margin Dollars
Quantity Sold
Order Date
Date
Month
Quarter
Year
Salesperson
Salesperson Name
Territory Name
Region Name
Month = June
Salesperson Name
= Jane Doe

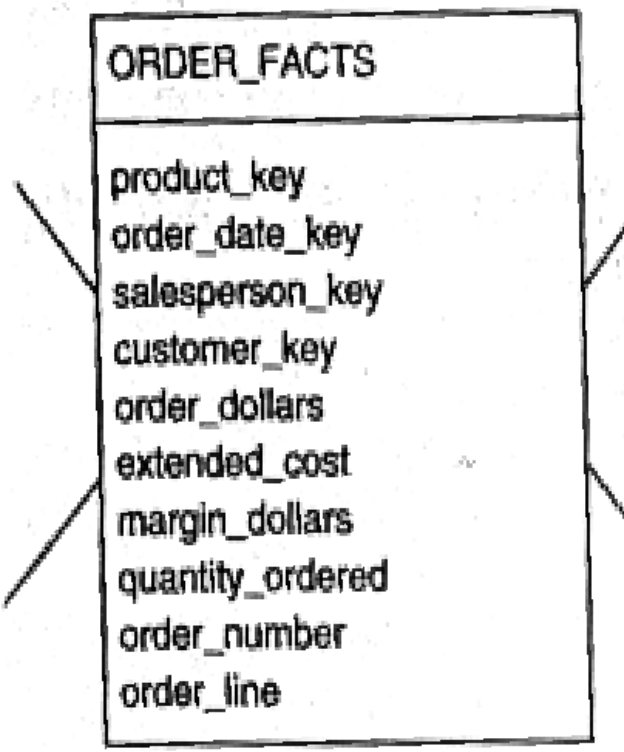
Inside a Dimension Table

- Dimension table key
- Large number of attributes (wide)
- Textual attributes
- Attributes not directly related
- Flattened out, not normalized
- Ability to drill down / roll up
- Multiple hierarchies
- Less number of records



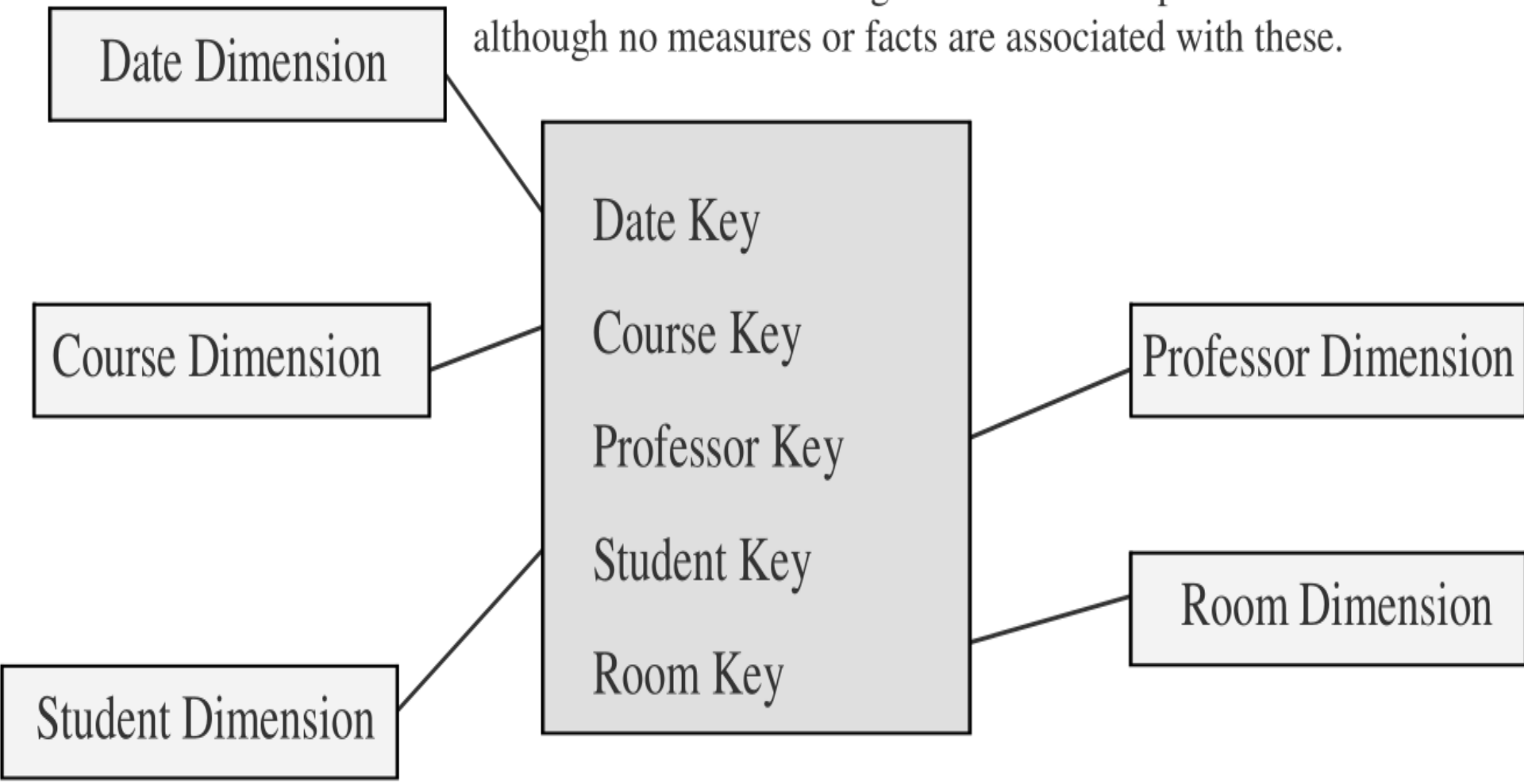
Inside a Fact Table

- Concatenated fact table key
- Grain or level of data identified
- Fully additive measures
- Semi-additive measures
- Large number of records
- Only a few attributes
- Sparsity of data
- Degenerate dimensions

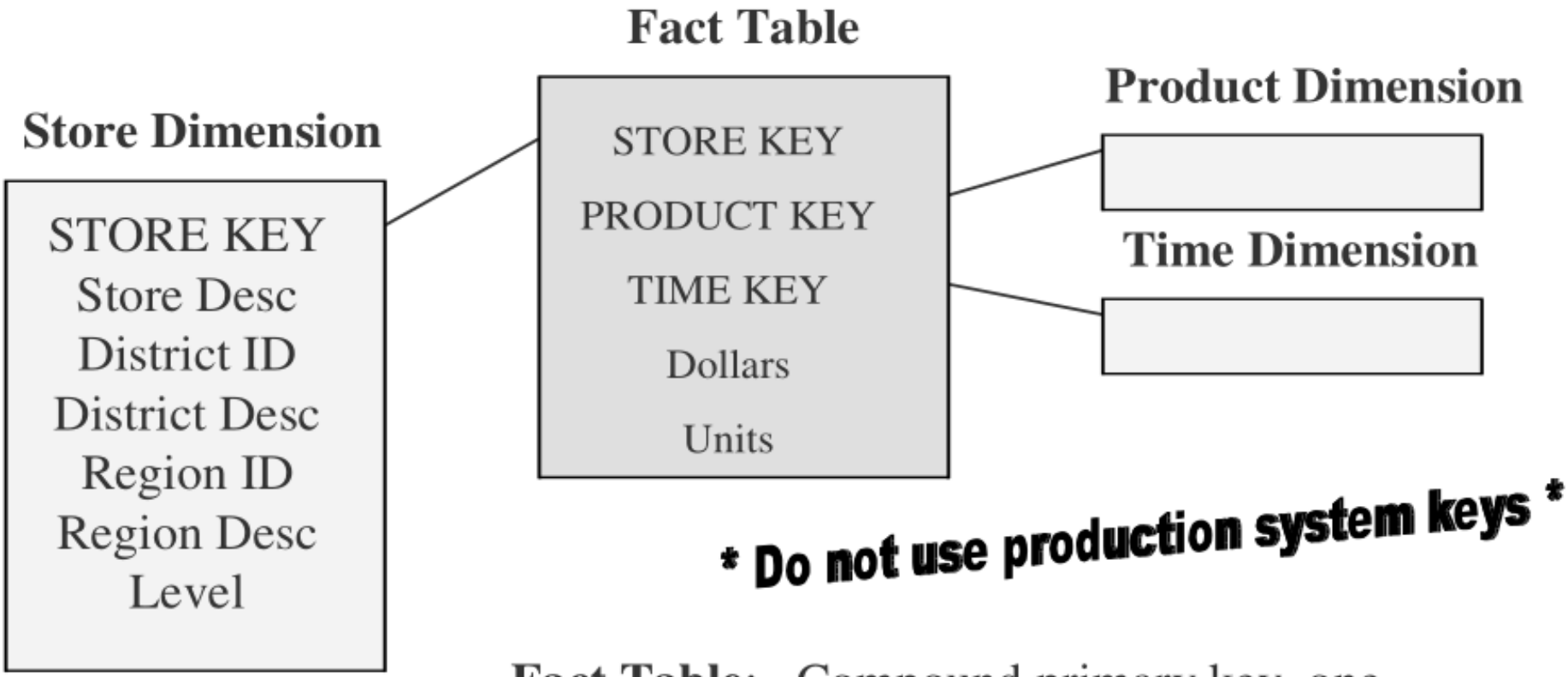


The Factless Fact Table

Measures or facts are represented in a fact table. However, there are business events or coverage that could be represented in a fact table although no measures or facts are associated with these.



The STAR Schema Keys



*** Do not use production system keys ***

Fact Table: Compound primary key, one segment for each dimension

Dimension Table: Generated primary key

Advantages of the STAR Schema

A STAR schema is simply a relational model with a one-to-many relationship between each dimension and the fact table.

What is so special about the arrangement of the STAR schema?

Why is it suitable for the data warehouse?

CHAPTER SUMMARY

- The components of the dimensional model are derived from the information packages in the requirements definition.
- The entity-relationship modeling technique is not suitable for data warehouses; the dimensional modeling technique is appropriate.
- The STAR schema used for data design is a relational model consisting of fact and dimension tables.

CHAPTER SUMMARY

- The fact table contains the business metrics or measurements; the dimensional tables contain the business dimensions. Hierarchies within each dimension table are used for drilling down to lower levels of data.
- STAR schema advantages are: easy for users to understand, optimizes navigation, most suitable for query processing, and enables specific performance schemes.

REVIEW QUESTIONS

1. Why is the entity-relationship modeling technique not suitable for the data warehouse? How is dimensional modeling different?
2. What is the STAR schema? What are the component tables?
3. A dimension table is wide; the fact table is deep. Explain.

REFERENCES

- Ponniah, P. (2001) Data Warehousing Fundamentals. New York: John Wiley & Sons.
- Golfarelli, M. & Rizzi, S. (2009) Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill.