

# Chapter 11: Data Warehousing

## ***Modern Database Management*** ***7<sup>th</sup> Edition***

***Jeffrey A. Hoffer, Mary B. Prescott,  
Fred R. McFadden***

© 2005 by Prentice Hall

1

## Definition

- **Data Warehouse:**
  - A subject-oriented, integrated, time-variant, non-updatable collection of data used in support of management decision-making processes
  - ***Subject-oriented:*** e.g. customers, patients, students, products
  - ***Integrated:*** Consistent naming conventions, formats, encoding structures; from multiple data sources
  - ***Time-variant:*** Can study trends and changes
  - ***Non-updatable:*** Read-only, periodically refreshed
- **Data Mart:**
  - A data warehouse that is limited in scope

# Need for Data Warehousing

- Integrated, company-wide view of high-quality information (from disparate databases)
- Separation of *operational* (OLTP) and *informational* systems and data (for improved performance)

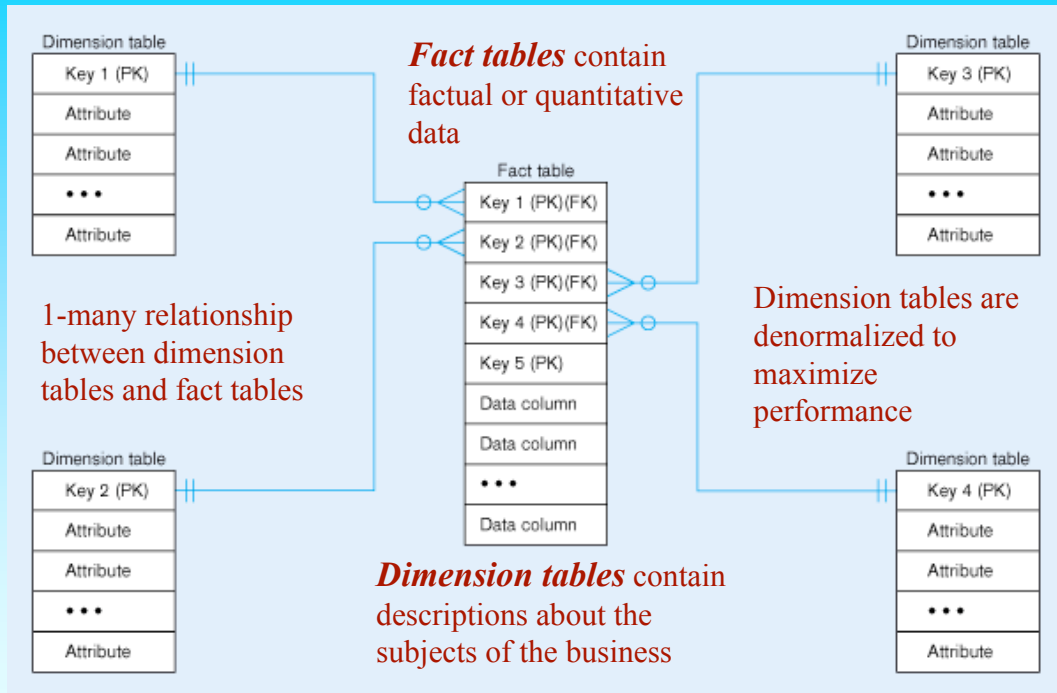
**Table 11-1 Comparison of Operational and Informational Systems**

Characteristic	Operational Systems	Informational Systems
Primary purpose	Run the business on a current basis	Support managerial decision making
Type of data	Current representation of state of the business	Historical point-in-time (snapshots) and predictions
Primary users	Clerks, salespersons, administrators	Managers, business analysts, customers
Scope of usage	Narrow, planned, and simple updates and queries	Broad, ad hoc, complex queries and analysis
Design goal	Performance throughput, availability	Ease of flexible access and use
Volume	Many, constant updates and queries on one or a few table rows	Periodic batch updates and queries requiring many or all rows

## Data Reconciliation

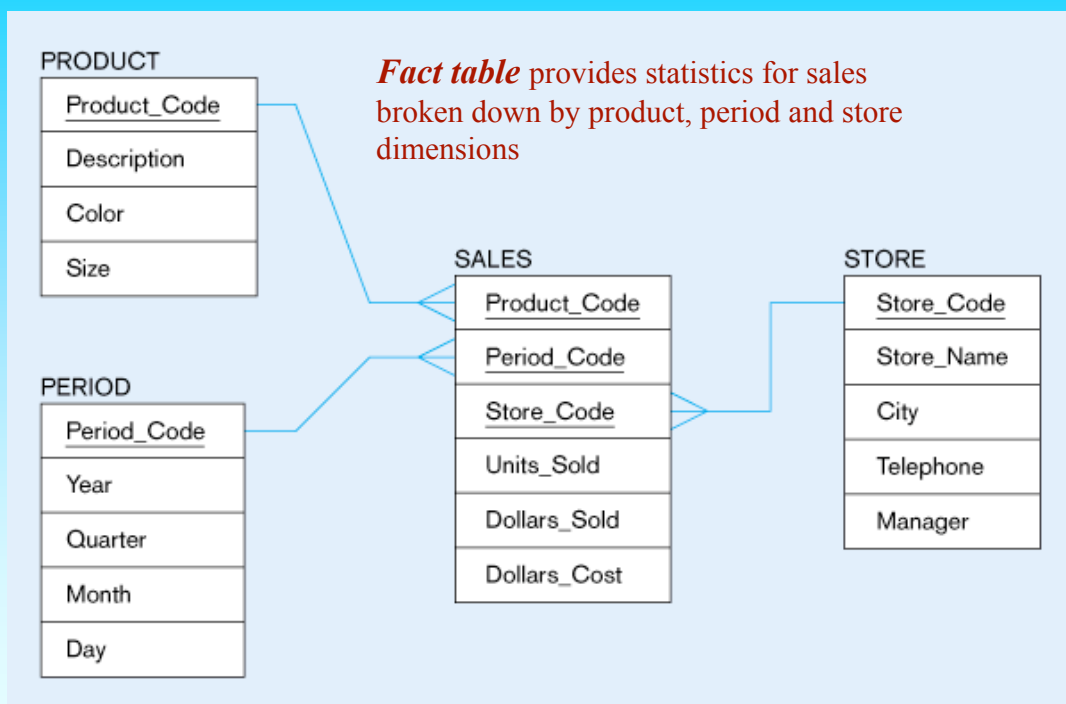
- Typical operational data is:
  - Transient – not historical
  - Not normalized (perhaps due to denormalization for performance)
  - Restricted in scope – not comprehensive
  - Sometimes poor quality – inconsistencies and errors
- After ETL (Extract-Transform-Load), data should be:
  - Detailed – not summarized yet
  - Historical – periodic
  - Normalized – 3<sup>rd</sup> normal form or higher
  - Comprehensive – enterprise-wide perspective
  - Timely – data should be current enough to assist decision-making
  - Quality controlled – accurate with full integrity

Figure 11-13: Components of a **star schema**

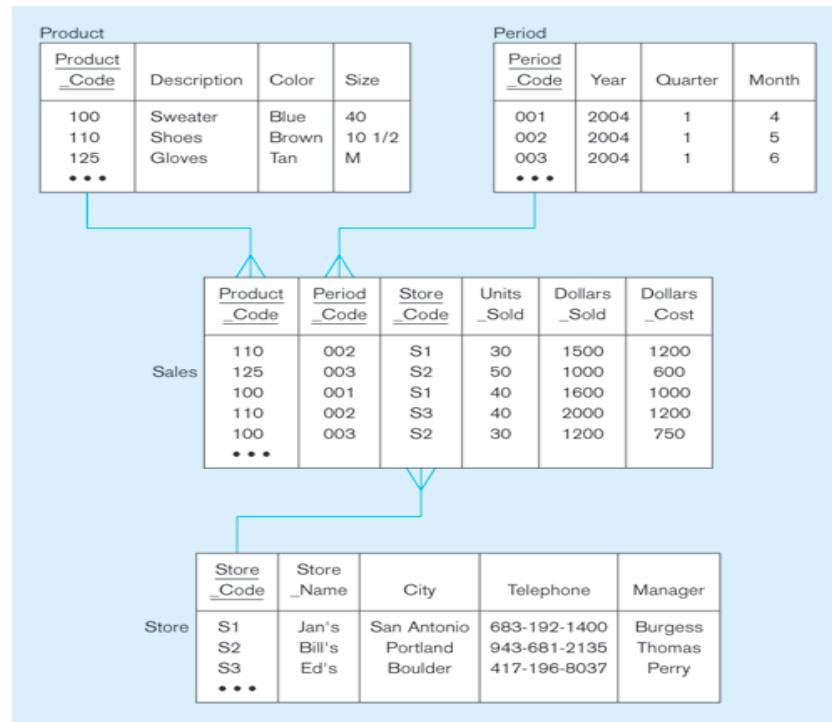


Data in fact table are called **measures** (or **dependent attributes**)

Figure 11-14: Star schema example



**Figure 11-15** Star schema with sample data



## Issues Regarding Star Schema

- Dimension table keys must be **surrogate** (non-intelligent and non-business related), because:
  - Keys may change over time
  - Length/format consistency
- Granularity of Fact Table – what level of detail do you want?
  - Transactional grain – finest level
  - Aggregated grain – more summarized
  - Finer grains → better **market basket analysis** capability
  - Finer grain → more dimension tables, more rows in fact table
- Duration of the database – how much history should be kept?
  - General guideline: As much as the technology employed allows.

# On-Line Analytical Processing (OLAP)

- The use of a set of (graphical) tools that provides users with multidimensional views of their data and allows them to analyze the data using simple windowing techniques
- OLAP Operations
  - **Dicing** - aggregating "dices" of the cube.
  - **Cube slicing** – come up with (2-D) view of part of data based on restricting the dimensions
  - **Drill-down** – going from summary to more detailed views

Figure 11-22: Dicing and slicing a data cube

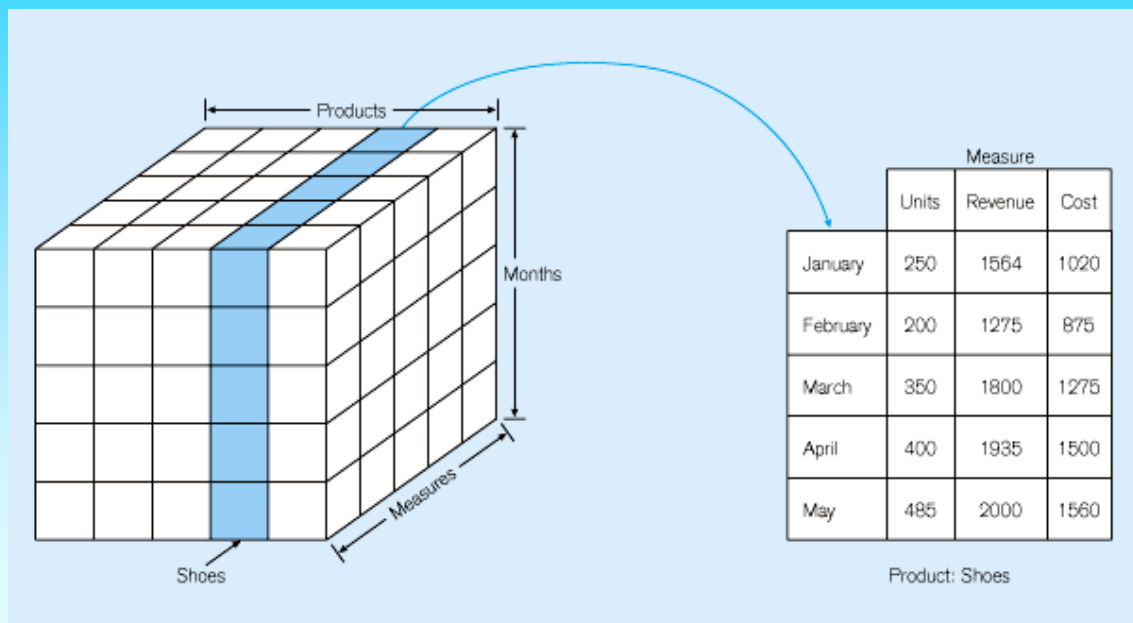


Figure 11-23:  
Example of drill-down

Starting with summary  
data, users can obtain  
details for particular  
cells

#### Summary report

Brand	Package size	Sales
SofTowel	2-pack	\$75
SofTowel	3-pack	\$100
SofTowel	6-pack	\$50

#### Drill-down with color added

Brand	Package size	Color	Sales
SofTowel	2-pack	White	\$30
SofTowel	2-pack	Yellow	\$25
SofTowel	2-pack	Pink	\$20
SofTowel	3-pack	White	\$50
SofTowel	3-pack	Green	\$25
SofTowel	3-pack	Yellow	\$25
SofTowel	6-pack	White	\$30
SofTowel	6-pack	Yellow	\$20