

BUSINESS ANALYTICS

CHAPTER 29

LECTURE OUTLINE

- Data warehouses
- Comparison with operational databases
- Multi-dimensional schemas
- Functionality of a data warehouse

DATA WAREHOUSES

- **Data warehouse**

- “A subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management’s decisions.” [W.H.Inmon]
- Data comes from multiple databases
- Tools to make business decisions quickly and reliably based on historical data

- **Supported applications**

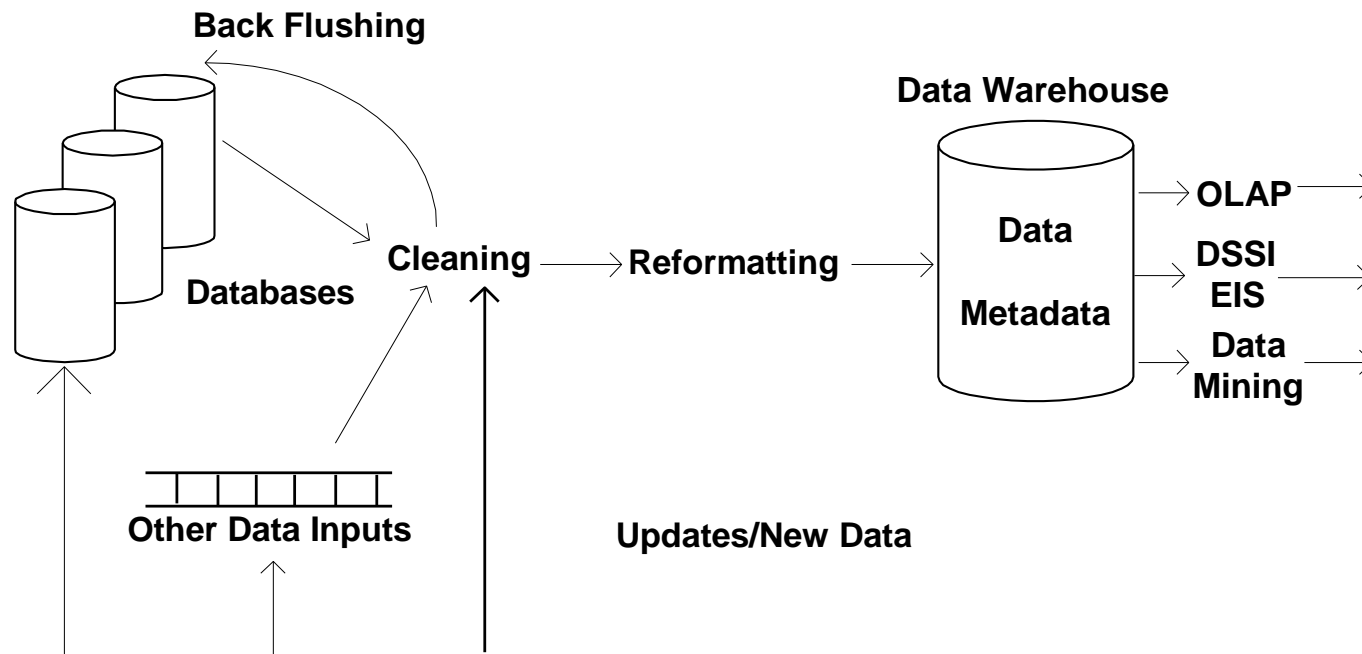
- **OLAP** (Online Analytical Processing)
 - Analysis of complex data from data warehouse
- **DSS** (Decision Support Systems)
 - Also known as EIS (Executive Information Systems)
 - Provides data and tools for complex decision-making

- **Data mining**

- Knowledge discovery: searching data for unanticipated new knowledge

INTEGRATED WAREHOUSE AND DB

- **Extract, Transform, and Load (ETL)**
 - Extracted from multiple, heterogeneous sources.
 - Includes data cleaning to ensure validity and consistency
 - **Back flushing:** upgrading the data with cleaned data
- Analyzed data fed back into operating DB and data management



DATA WAREHOUSES VS. DBS

- Operations
 - Data warehouses optimized to find data correlations and to support trend analyses
 - Traditional databases are transactional: optimized for access, update, and integrity assurance
 - Data warehouses are less volatile than operational DBs.
- Data currency
 - Operational DBs required to maintain up-to-date, detailed data
 - Data warehouses characterized by historical data
 - Information in data warehouse is relatively coarse grained (“view from 10,000 ft.”) and refresh policy is carefully chosen, usually incremental.
- Data volume
 - Data warehouses may be exceptionally large (7 years of records)
- Data warehouse can be interpreted as a (special) view of the data.

MULTI-DIMENSIONAL SCHEMAS

- Multi-dimensional schemas specified using:
 - **Dimension table**
 - Consists of tuples of attributes of the dimension.
 - **Fact table**
 - Each tuple is a recorded fact.
 - Some measured or observed variable(s) and references to dimension tables.

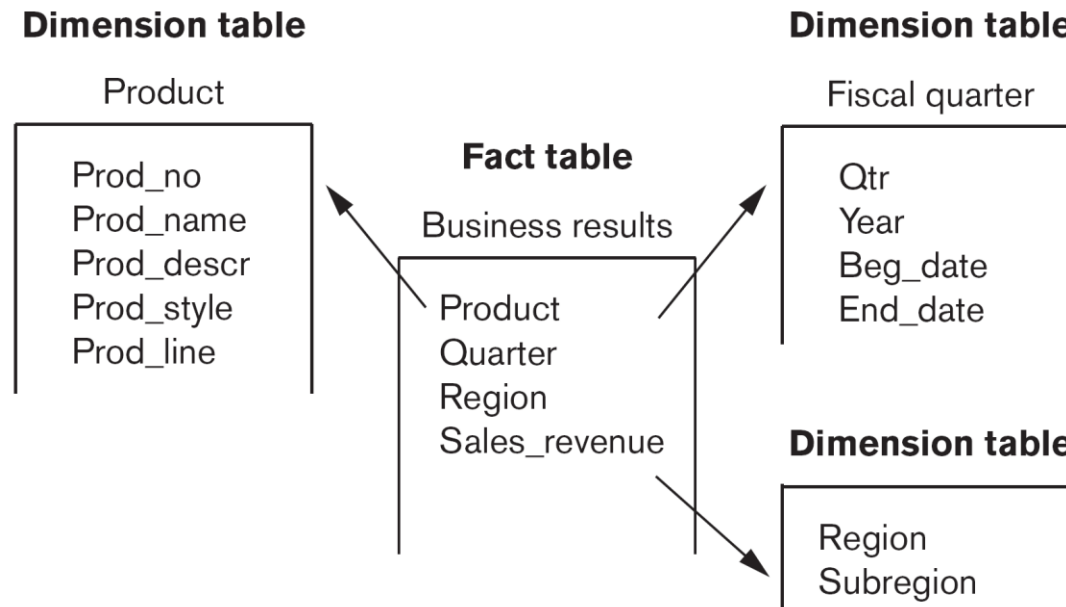


Figure 29.7

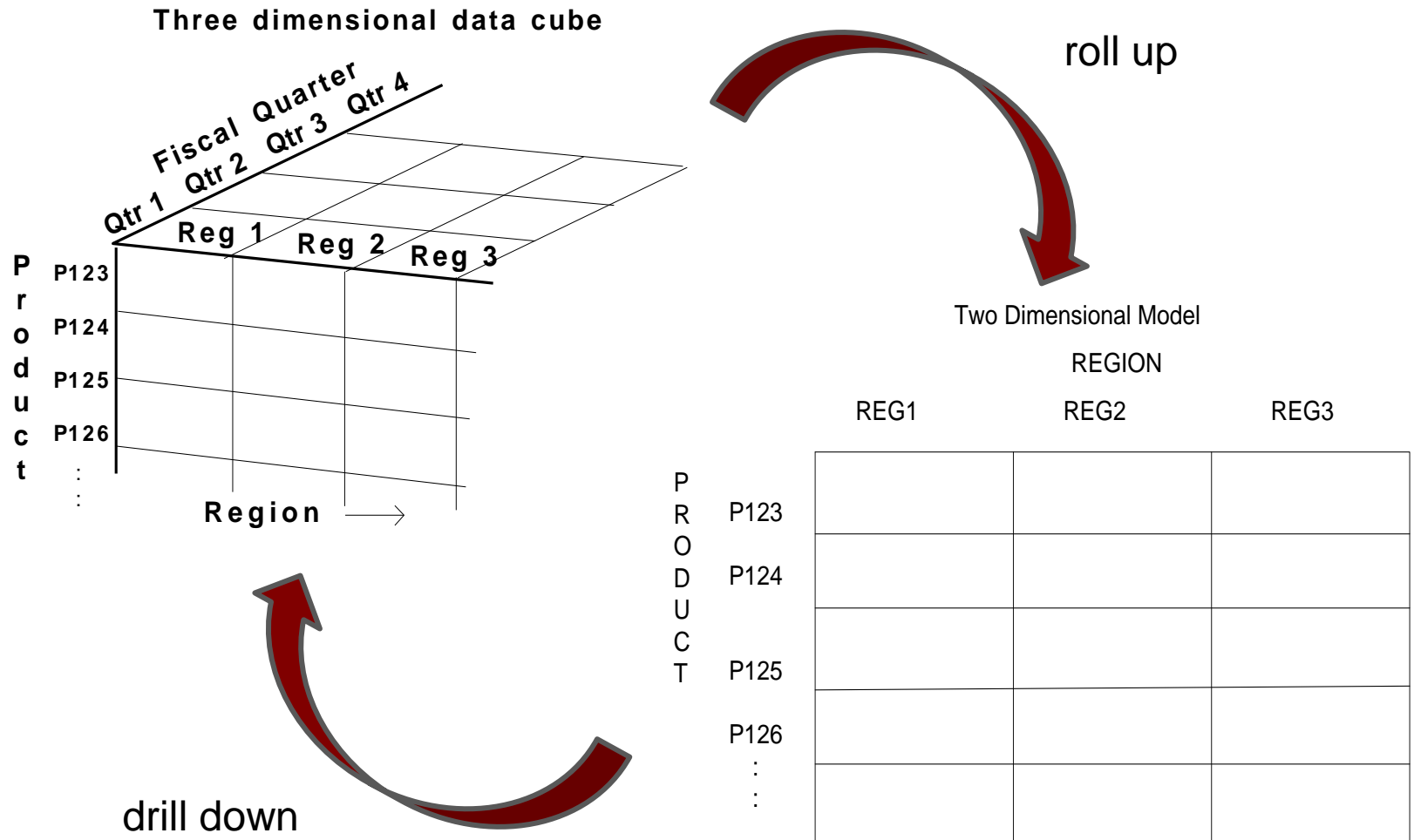
A star schema with fact and dimensional tables.

WAREHOUSE FUNCTIONALITY

- **Roll-up:** Data is summarized with increasing generalization
 - E.g., going from daily or weekly reports to annual aggregations
- **Drill-Down:** Increasing levels of detail are revealed
 - E.g., going from national sales to sales from a particular region
- **Pivot:** Cross tabulation is performed from given perspective
- **Slice and dice:** Select and project data wrt some dimensions
- Plus traditional operations
 - Sorting by ordinal value.
 - Selection by value or range.
- **Derived attributes:** Attributes are computed by operations on stored derived values.

DATA MODELING FOR WAREHOUSES

- OLAP data cube



LECTURE SUMMARY

- Glimpse at DBMS support for business analytics
 - Role of data warehouses
- OLAP cube
 - Model and operations