### 13.7 Star Schema

- Data modeling technique
  - Maps multidimensional decision support data into relational database
- Creates near equivalent of multidimensional database schema from relational data
- Easily implemented model for multidimensional data analysis
  - Preserves relational structures on which operational database is built
- Four components: facts, dimensions, attributes, and attribute hierarchies

#### Facts
- Numeric measurements that represent specific business aspect or activity
  - Normally stored in fact table that is center of star schema
- Fact table contains facts linked through their dimensions
- Metrics are facts computed at run time

#### Dimensions
- Qualifying characteristics provide additional perspectives to a given fact
- Decision support data almost always viewed in relation to other data
- Study facts via dimensions
- Dimensions stored in dimension tables

#### Attributes
- Each dimension table contains attributes
- Use to search, filter, and classify facts
- Dimensions provide descriptive characteristics of facts through their attributes
- No mathematical limit to the number of dimensions
- The data warehouse designer must define common business attributes, to be used by data analyst to narrow a search, group information, or describe dimension
### Attribute Hierarchies

- Attributes within dimensions can be *ordered* in a well-defined attribute hierarchy
- Provide top-down data organization
- Two purposes:
  - Aggregation
  - Drill-down/Roll-up data analysis
- Determine how the data are extracted and represented
- Stored in the DBMS's data dictionary
- Used by OLAP tool to access warehouse properly
Star Schema Representation

- Facts and dimensions represented in physical tables in data warehouse database
- Many fact rows related to each dimension row
  - Primary key of fact table is a composite primary key
  - Fact table primary key formed by combining foreign keys pointing to dimension tables
- Dimension tables smaller than fact tables
- Each dimension record related to thousands of fact records

13.9 Data Mining

- Data-mining tools do the following:
  - Analyze data
  - Uncover problems or opportunities hidden in data relationships
  - Form computer models based on their findings
  - Use models to predict business behavior
- Requires minimal end-user intervention
13.10 SQL Extensions for OLAP

- Proliferation of OLAP tools fostered development of SQL extensions
- Many innovations have become part of standard SQL
- All SQL commands will work in data warehouse as expected
- Most queries include many data groupings and aggregations over multiple columns
The ROLLUP Extension

- Used with GROUP BY clause to generate aggregates by different dimensions
- GROUP BY generates only one aggregate for each new value combination of attributes
- ROLLUP extension enables subtotal for each column listed except for the last one
  - Last column gets grand total
- Order of column list important

The CUBE Extension

- CUBE extension used with GROUP BY clause to generate aggregates by listed columns
  - Includes the last column
- Enables subtotal for each column in addition to grand total for last column
- Useful when you want to compute all possible subtotals within groupings
- Cross-tabulations good application of CUBE extension
Materialized Views

- A dynamic table that contains SQL query command to generate rows
  - Also contains the actual rows
- Created the first time query is run and summary rows are stored in table
- Automatically updated when base tables are updated

Summary

- Business intelligence generates information used to support decision making
- BI covers a range of technologies, applications, and functionalities
- Decision support systems were the precursor of current generation BI systems
- Operational data not suited for decision support
Summary (continued)

• Four categories of requirements for decision support DBMS:
  – Database schema
  – Data extraction and loading
  – End-user analytical interface
  – Database size requirements
• Data warehouse provides support for decision making
  – Usually read-only
  – Optimized for data analysis, query processing

Summary (continued)

• OLAP systems have four main characteristics:
  – Use of multidimensional data analysis
  – Advanced database support
  – Easy-to-use end-user interfaces
  – Client/server architecture
• ROLAP provides OLAP functionality with relational databases
• MOLAP provides OLAP functionality with MDBMSs

Summary (continued)

• Star schema is a data-modeling technique
  – Maps multidimensional decision support data into a relational database
• Star schema has four components:
  – Facts
  – Dimensions
  – Attributes
  – Attribute hierarchies
• Data mining automates analysis of operational data
• SQL extensions support OLAP-type processing and data generation