Database Systems

Codd's 12 Relational Database Rules (*See Table 13.8 in Section 3.9*)

Foundation: A system must manage DBs entirely through its relational capabilities.

- Information
- Guaranteed Access
- Systematic Treatment of Null Values
- Dynamic Online Catalog
- Comprehensive Data Sublanguage
- View Updating

- High-Level Insert, Update, and Delete
- Physical Data Independence
- Logical Data Independence
- Integrity Independence
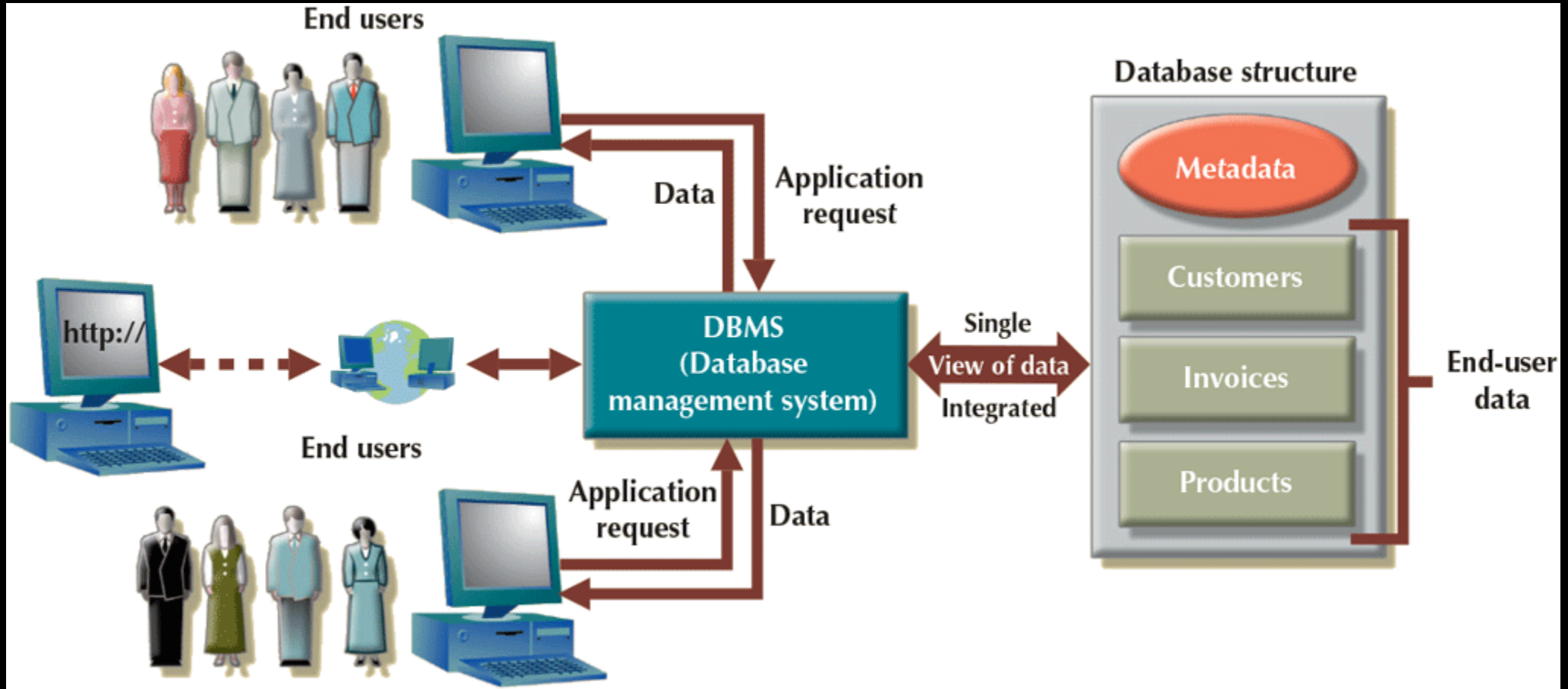- Distribution Independence
- Non-Subversion

Data vs. Information

- Data – raw facts; not yet processed
  - Building blocks of information
  - Need to be formatted for storage, processing, and presentation
- Information – processed data to reveal it's meaning
  - Produced by processing data
  - From organization to data analysis
  - Requires context
  - <u>Accurate</u>, <u>relevant</u>, <u>valid</u>, and <u>timely</u> information is key to good decisions

Databases

- Data management
  - Discipline focused on the proper generation, storage, and retrieval of data
- Database – shared, integrated computer structure that stores:
  - End-user data – raw facts of interest to the user
  - Metadata – data about data used to integrate and manage end-user data
    - Describes data characteristics and the set of relationships
    - Examples: data name, data type, data constraints
- DBMS – Database Management System
  - Collections of programs to manage the database structure and access

DBMS Role

DBMS Advantages

- Improved data sharing
- Improved data security
- Better data integration
- Minimized data inconsistency
- Improved data access
- Improved decision making
- Increased end-user productivity

Important Terminology

- Data Inconsistency – different versions of the same data exists in different places and yield different (inconsistent) results

- Data quality – comprehensive approach to ensuring the <u>accuracy</u>, <u>validity</u>, and <u>timeliness</u> of data

- Data integrity – data in the DB complies with all entity and referential integrity constraints (important concept this semester)

- Query – question or task for the DB (in SQL Code)
  - A specific request for data retrieval or manipulation

- Ad hoc query – spur-of-the-moment query

Important Terminology

- Data dictionary (Section 3.5)
  - Stores metadata including data definition, characteristics, and relationships
  - May include information about data external to the database
- System catalog (Section 3.5)
  - Detailed system data dictionary that describes all objects of the database

DBMS Functions

- Data dictionary management
- Data storage management
- Data transformation and presentation
- Security management
- Multiuser access control
- Backup and recovery management
- Data integrity management
- Database access languages and application development interfaces
- Database communication interfaces

Why Not Just Use Files?

- Structural dependence
  - Dependence – access to data is dependent on its structure
  - Independence – changes in the schema (structure) doesn't affect data access
- Data dependence
  - Physical data format – how the computer organizes and stores the data
  - Logical data format – how we view the data
  - Dependence – data representation and manipulation are dependent on physical data storage characteristics
  - Independence – changes in the physical data storage does not affect representation and manipulation

Data Redundancy

- Same data stored unnecessarily at different places
- Uncontrolled data redundancy may lead to
  - Poor data security
  - Data inconsistency
  - Data entry errors
  - Data integrity problems
  - Data anomalies – not all changes in redundant data are made successfully
    - Update – change in existing redundant data is not consistent
    - Insertion – adding data into redundant locations is not consistent
    - Deletion – deleting existing redundant data is not consistent