Database Systems

- Need for information
  - information age and information explosion
  - information: structured and interpreted data
  - one of the major driving force on the growing use of computers individually and organizationally
  - plays a critical role in almost all areas where computers are used: business, engineering, education

- Database
  - a collection of self-describing integrated data
  - represents some aspects of the real world
  - a logically coherent collection
  - designed, built, and populated with data for a specific purpose (users and applications)
A collection of self-describing integrated data

- self-describing: file layouts and structures are kept in data dictionary; user need not know them
- integrated: data aggregation
  bits > bytes > attributes > tuples > relations

Representation of a mini-world

- database objects represent entities of interest whose properties (attributes) are represented by data values
- database operations (transactions) represent events of interest that add, delete, update objects
Database Management System (DBMS)

- **Definition**
  - a collection of programs that enables users to create and maintain a database
  - a general purpose software system that facilitates the process of defining, constructing, and manipulating databases for various applications

- **Defining a database**
  - specifying the data types, structures, and constraints for the data to be stored in the database

- **Constructing a database**
  - process of storing the data on some storage medium controlled by the DBMS

- **Manipulating a database**
  - retrieve specific data and update to reflect changes
Database Management System

- Primary goal
  - to provide an environment that is convenient and efficient to use in storing and retrieving data
  - efficiency should not impair convenience: data must be available to users (applications) in a form that can be used immediately
  - one major reason for data abstraction
  - factors for efficiency: data structure, access method, and query optimization.

- Query optimization
  
  retrieve (name = Emily) AND (age > 20):
  
  search for the name first or age first
Requirements for Database Management System

● Maintain the integrity of the database
  - a set of assertions (integrity constraints) must hold

● Maintain the security of the database
  - access control (not everyone can access all data)
  - multi-level security

● Synchronization
  - control concurrent access for improved performance
  - anomalies not allowed: lost update, inconsistent read

● Replication control
  - reduce duplicated information
  - control divergence
• Crash protection and recovery
  - logging, checkpointing, recovery management
  - redo and undo rules

• Data independence
  - application programs should be immune to the change
    of physical database and/or access methods, etc.
Data Dependence

- Traditional file processing
  
  user --- application program --- files
  user --- application program --- files

  - data isolation and duplication
  
  - definitions of file structures are built into every application program that uses it

  - a change to that structure requires changes to all those application programs

  - expensive: ~70% programming effort goes modifying working software
Data Dependence

- Traditional file processing
  - other problem: two programs that need the same data in different formats
  - having two copies of the file in different format (space and potential inconsistency among copies)
  - reorganize files (expensive and limited on-line access)