

Database Management Systems

Data Models

What is data?

- Data are known facts that can be recorded and that have implicit meaning.
- Data are different from information (interpreted data are information and they have explicit meaning).
- Data can be viewed as a true proposition (a proposition is a sentence that could be true or false).
- For example: “Oracle developed DBMS Oracle” is a true proposition, while “Oracle developed DBMS SQL Server” is a false proposition.
- A database is really a collection of true propositions.

Company	DBMS_Product
Oracle	Oracle
Microsoft	SQL Server

Data Model

- A data model is a notation for describing the information of the real world objects and the relationships among the objects, and mapping these information into data items.
- Any data model generally consists of three parts:
 - Structure of the data or mathematical representation of data
e.g., files in a file system data model
 - Operations on the data
e.g., append and truncate operations in the file system data model
 - Constraints on the data
e.g., max size constraints in the file system data model
- A database system implements a data model.

Relational Data Model

- A single way to represent data: as a two-dimensional table called a relation.
- structure of the data: relations (as arrays of structs/records)
- operations: selection, projection, join, etc, or any operations from “relational algebra” that we will discuss later
- constraints: primary key and foreign key constraints, etc, that we will discuss later
- Advantages of relational data model: VERY simple (just rows and columns)
- Disadvantages of relational data model:
 - TOO simple sometimes
 - Not flexible with fixed schema

Other Data Models (I)

- network, hierarchy data model (outdated)
- object oriented data model
 - allows richer data model for complex data and relationships, instead of just relational rows and columns
- semi structured data model
 - Key-Value structure integrates schema with data
 - extremely flexible
 - used in most NoSQL database systems

Other Data Models (II)

- logic-based deductive systems
 - used in AI fields
 - sometimes called “knowledge base” systems
- multi-database systems
 - useful in the integration of multiple heterogeneous legacy database systems
- data warehousing
 - useful in integrating data from different sources
 - widely used in data mining
- the web

Advantage of Relational Data Model

- simple and solid conceptual model
- logical view and physical implementation strictly separated
- powerful, bag-oriented and declarative query languages (SQL)
- clean transaction processing

Requirements for a DBMS

- Provide data definition facilities
 - provide a data definition language (DDL)
 - provide a user accessible catalog (Data Dictionary)
- provide methods for storing, retrieving and updating data (Data Manipulation Language - DML)
- support multiple views (DDL)
- be able to specify integrity constraints (Data Control Language - DCL)
- provide security control over the data (DCL)
- Internal Processes
 - provide concurrency control mechanism (lock or timestamp based)
 - support logical transactions (ACID)
 - protect data while system crashes (recovery)
 - provide other maintenance utilities

Technologies involved in DBMS

- programming languages
- software engineering
- compilation/optimization
- operating systems
- concurrent programming
- data structure and algorithms
- distributed systems
- AI (Information retrieval)
- statistical techniques

People who work with database systems

- database system developers
- database administrators (DBA)
 - Manages conceptual schema
 - Assists with application view integration
 - Monitors overall performance of DBMS
 - Defines internal schema
 - Loads and reformats database
 - responsible for security and reliability

People who work with database systems (cont.)

- database application developers
 - programmer: implements applications to access the database
 - analyst: develops application specifications
- end users
 - naive user: access database through applications
 - sophisticated users: access database directly using ad-hoc DML queries