Principles of Database Systems

By Xu Lizhen
School of Computer Science and Engineering, Southeast University
Nanjing, China

Main Contents

In this course, we will learn the basic concepts, principles and applications of database systems, especially the relational database systems. The contents mainly include :

- The data models, SQL language and user interfaces
- Key principles of DBMS (mainly architecture, query optimization, concurrency control, recovery, etc.)
- The security and integrity constrains of database
- Introduction of distributed database systems
- Some new research and application fields of database technology, such as data warehouse, data mining, XML data management, etc.

References

1) Wang Nengbin, “Textbook of Database Systems”
2) Raghu Ramakrishnan, Johannes Gehrke, “Database Management Systems” , 3rd Edition, McGraw-Hill Companies, 2002
3) Hector Garcia-Molina, Jeffrey.D.Ullman, “Database Systems: the Complete Book”
4) C.J.Date, “An Introduction to Database Systems”
5) Web Site of our course:

   http://cselab.seu.edu.cn/course/dbprinciple/

Table of Contents

1. **Introduction**

   The history, classification, and main research contents of database systems; The database system; the concepts of data model

2. **Data Model***

   Hierarchical and network model; Relational model; ER model; Object-Oriented model and other data models

3. **User Interfaces and SQL Language***

   User interface; SQL language, including QL, DDL, DCL, DML, view, embedded SQL and dynamic SQL, etc.

4. **Database Management Systems***

   The architecture of database systems, query optimization, file structure and index, transaction management, concurrency control, recovery mechanism

Table of Contents

5. **The Security and Integrity Constrain**

   The security model of database system; Integrity constrain and its expression, implementing method, assertion, trigger

6. **Database Design***

   Design procedure; ER graph; Normalization of Relational Schema

7. **Distributed Database Systems**

   What and Why DDBS, data distribution, distributed database design; Query optimization, distributed transaction management in DDBMS

8. **New Research and Application Fields**

   Data warehouse, OLAP; Data mining; XML data management

1. Introduction

What Is Database?
What Is DBMS?

- A very large, integrated collection of data.
- Models real-world *enterprise.*
  - Entities (e.g., students, courses)
  - Relationships (e.g., electives)
- A *Database Management System (DBMS)* is a software package designed to store and manage databases.

Files vs. Databases

- Application must stage large datasets between main memory and secondary storage (e.g., buffering, page-oriented access, 32-bit addressing, etc.)
- Special code for different queries
- Must protect data from inconsistency due to multiple concurrent users
- Crash recovery
- Security and access control

Why Use a DBMS?

- Data independence and efficient access.
- Reduced application development time.
- Data integrity and security.
- Uniform data administration.
- Concurrent access, recovery from crashes.

Why Study Databases ?

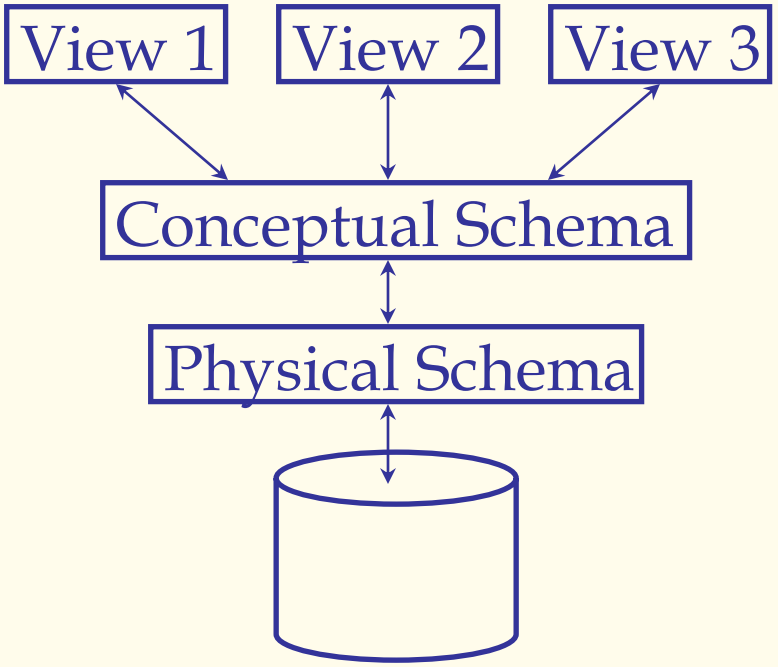
- Shift from *computation* to *information*
  - at the "low end": scramble to webspace (a mess!)
  - at the "high end": scientific applications
- Datasets increasing in diversity and volume.
  - Digital libraries, interactive video, Human Genome project, EOS project
  - ... need for DBMS exploding
- DBMS encompasses most of CS
  - OS, languages, theory, AI, multimedia, logic

Data, Data Model and Data Schema

- *Data* are symbols for describing the things of real world. They are existing form of information.
- A *data model* is a collection of concepts and definitions for describing data.
- A *schema* is a description of a particular collection of data, using a given data model.
- The *relational model of data* is the most widely used model today.
  - Main concept: *relation*, basically a table with rows and columns.
  - Every relation has a *schema*, which describes the columns, or fields.

Levels of Abstraction

- Many *views*, single *conceptual (logical) schema* and *physical schema*.
  - Views describe how users see the data.
  - Conceptual schema defines logical structure
  - Physical schema describes the files and indexes used.



* *Schemas are defined using DDL; data is modified/queried using DML.*

Example: University Database

- Conceptual schema:
  - *Students(sid: string, name: string, login: string, age: integer, gpa:real)*
  - *Courses(cid: string, cname:string, credits:integer)*
  - *Enrolled(sid:string, cid:string, grade:integer)*
- Physical schema:
  - Relations stored as unordered files.
  - Index on first column of Students.
- External Schema (View):
  - *Course_info(cid:string,enrollment:integer)*

Data Independence *

- Applications insulated from how data is structured and stored.
- *Logical data independence*: Protection from changes in *logical* structure of data.
- *Physical data independence*: Protection from changes in *physical* structure of data.

* *One of the most important benefits of using a DBMS!*

The History of Database Technology and its Classification

(1) According to the development of data model

- No management(before 1960’): Scientific computing
- File system: Simple data management
- Demand of data management growing continuously, DBMS emerged.

  - 1964, the first DBMS (American): IDS, network
  - 1969, the first commercial DBMS of IBM, hierarchical
  - 1970, E.F.Codd(IBM) bring forward relational data model
  - Other data model: Object Oriented, deductive, ER, ...

The History of Database Technology and its Classification

(1) According to the development of data model

- No management(before 1960’): Scientific computing
- File system: Simple data management
- Demand of data management growing continuously, DBMS emerged.

  - 1964, the first DBMS (American): IDS, network
  - 1969, the first commercial DBMS of IBM, hierarchical
  - 1970, E.F.Codd(IBM) bring forward relational data model
  - Other data model: Object Oriented, deductive, ER, ...

(2) According to the development of DBMS architectures

- Centralized database systems
- Parallel database systems
- Distributed database systems (and Federated database systems)
- Mobile database systems

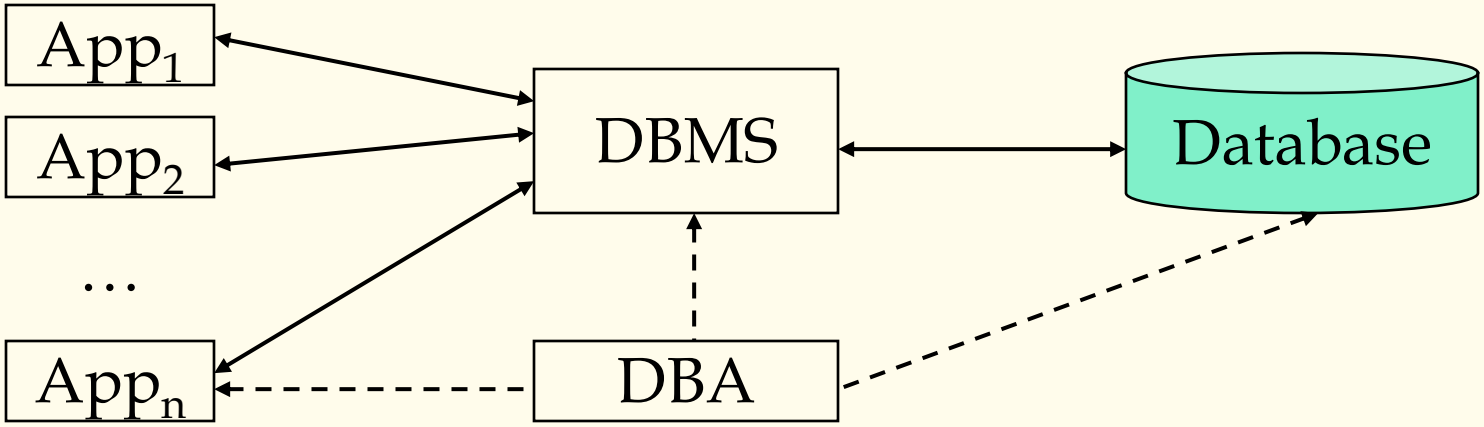
(3) According to the development of architectures of application systems based on databases

- Centralized structure : Host + Terminal
- Distributed structure
- Client/Server structure
- Three tier/multi-tier structure
- Mobile computing
- Grid computing / Cloud computing

(4) According to the expanding of application fields

- OLTP
- Engineering Database
- Deductive Database
- Multimedia Database
- Temporal Database
- Spatial Database
- Data Warehouse, OLAP, Data Mining
- Knowledge Management
- … …

Database System



- Applications + DBMS + Database + DBA
- DBMS is the core of database system
  - High level user interfaces
  - Query processing and optimization
  - Catalog management
  - Concurrency control and Recovery
  - Integrity constraints checking
  - Access control

Life cycle of database systems

- Database system planning
- Database designing
- Database establishing and loading
- Database running, managing and maintaining
- Database extending and restructuring