Chapter 1:
An Overview of Database Systems

Database Systems (CO2013)

Computer Science Program

Assoc. Prof. Dr. Võ Thị Ngọc Châu

(chauvtn@hcmut.edu.vn)

Semester 1 – 2022-2023

Main References

Text:

- [1] R. Elmasri, S. R. Navathe, Fundamentals of Database Systems- 6th Edition, Pearson- Addison Wesley, 2011.
    - ***R. Elmasri, S. R. Navathe, Fundamentals of Database Systems- 7th Edition, Pearson, 2016.***

References:

[1] S. Chittayasothorn, *Relational Database Systems: Language, Conceptual Modeling and Design for Engineers*, Nutcha Printing Co. Ltd, 2017.

[3] A. Silberschatz, H. F. Korth, S. Sudarshan, Database System Concepts – 7th Edition, McGraw-Hill, 2020.

[4] H. G. Molina, J. D. Ullman, J. Widom, *Database Systems: The Complete Book - 2nd Edition,* Prentice-Hall, 2009.

[5] R. Ramakrishnan, J. Gehrke, *Database Management Systems – 4th Edition,* McGraw-Hill, 2018.

[6] M. P. Papazoglou, S. Spaccapietra, Z. Tari, *Advances in Object-Oriented Data Modeling*, MIT Press, 2000.

[7]. G. Simsion, *Data Modeling: Theory and Practice*, Technics Publications, LLC, 2007.

Content

Chapter 1: An overview of database systems

- 1.1. Concepts
- 1.2. File processing systems
- 1.3. The database approach
- 1.4. Data models
- 1.5. Database management systems
- 1.6. Database systems
- 1.7. Applications of database systems

1.1. Concepts

- Data/ Information/ Knowledge/ Metadata

  - Data

  - Information

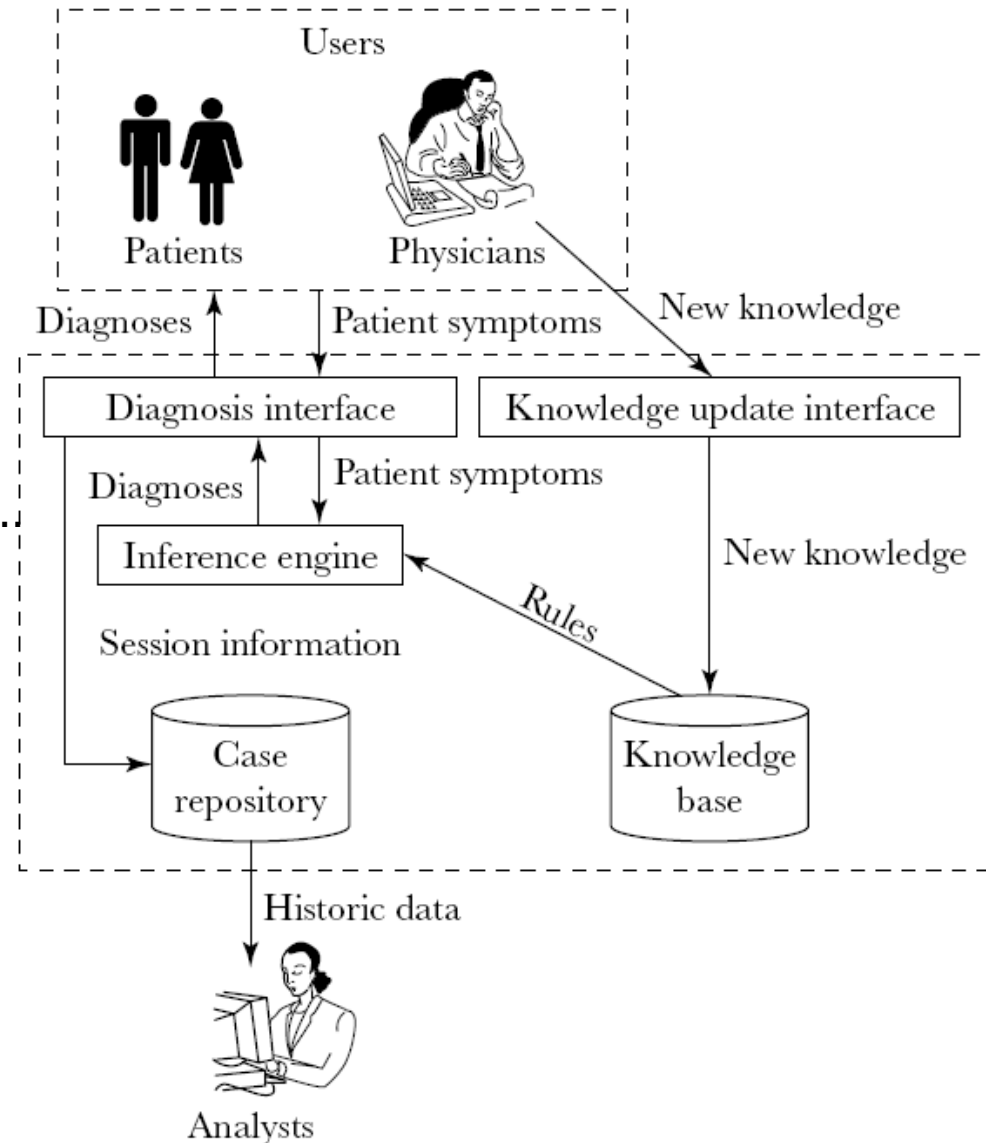 - Knowledge

  - Metadata

  → Relative

- Database

Data

- information, especially facts or numbers, collected for examination and consideration and used to help decision-making, or information in an electronic form that can be stored and processed by a computer

  - Cambridge Advanced Learner's Dictionary

- an elementary description of things, events, activities, and transactions that are recorded, classified, and stored but not organized to convey any specific meaning

  - R. K. Rainer, C. G. Cegielski, "Introduction to Information Systems", 3rd Edition, John Wiley & Sons, Inc, pp. 10, 2004.

Data

Factual data and clinical evidence provided by a clinician or patient

→Bệnh nhân A: tên, địa chỉ, thân

nhiệt, hình ảnh về bệnh nhân, …

→ Bác sĩ B: giờ khám, tên thuốc, …

Users

Patients   Physicians

Diagnoses   Patient symptoms   New knowledge

Diagnosis interface   Knowledge update interface

Diagnoses   Patient symptoms

Inference engine   New knowledge

Rules

Session information

Case repository   Knowledge base

Historic data

Analysts

Kiến trúc của hệ hỗ trợ chẩn đoán dựa trên Web (architecture of a Web-based diagnosis support system)

Information

- <mark>facts</mark> about a situation, person, event, etc
    - Cambridge Advanced Learner's Dictionary
- <mark>data</mark> that have been <mark>organized</mark> so that they have meaning and value to the recipient
- → the <mark>recipient interprets the meaning</mark> and draws conclusions and implications from the information
    - R. K. Rainer, C. G. Cegielski, "Introduction to Information Systems", 3rd Edition, John Wiley & Sons, Inc, pp. 10, 2004.
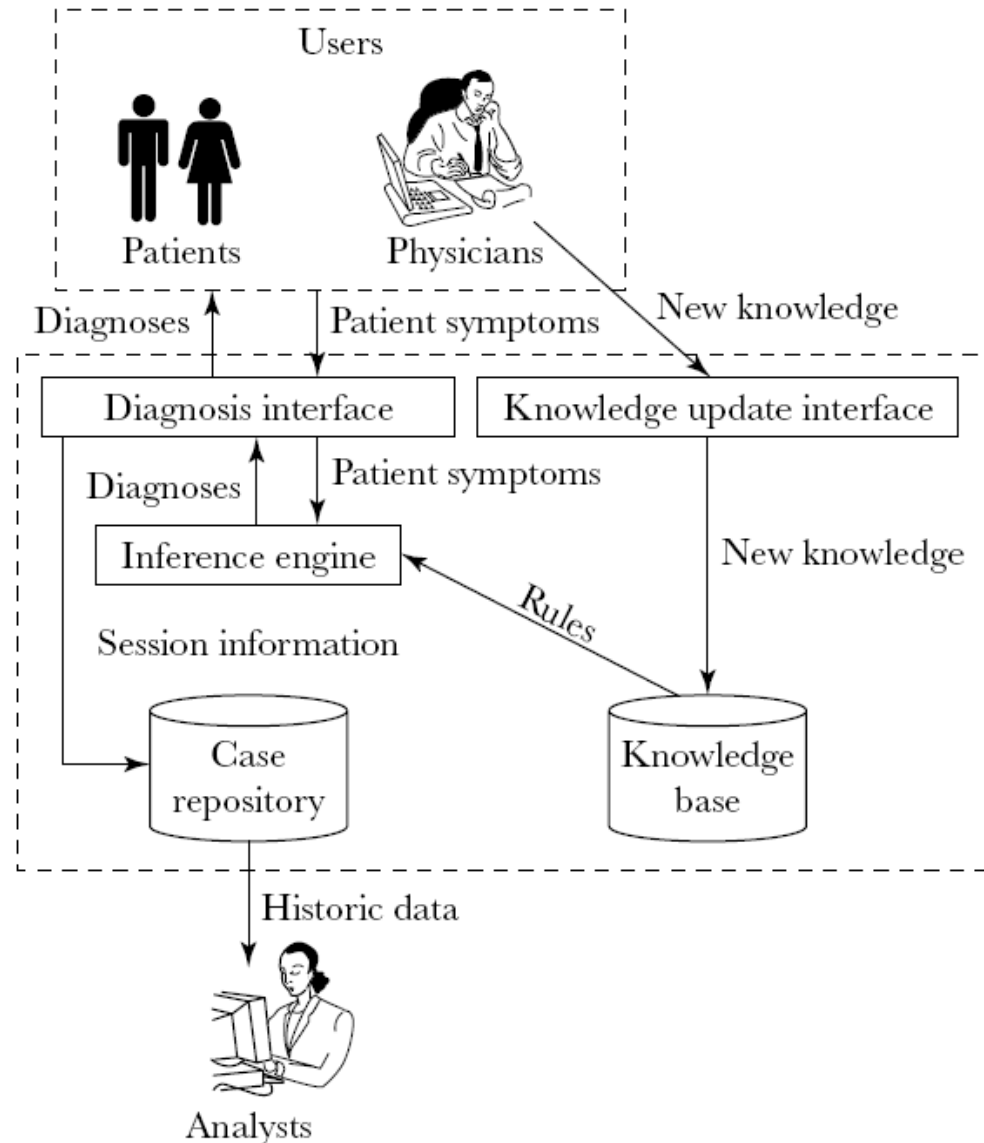
Information

Bệnh nhân A có thân nhiệt 37.5°.

Bác sĩ B chuyên chẩn đoán bệnh về tim mạch.

Mỗi tuần, trung bình 100 bệnh nhân tương tác với hệ thống.

…

Kiến trúc của hệ hỗ trợ chẩn đoán dựa trên Web (architecture of a Web-based diagnosis support system)

Knowledge

- Awareness; ==understanding== of or information about a ==subject== which has been obtained by experience or study, and which is either in a person's mind or possessed by people generally

  - Cambridge Advanced Learner's Dictionary

- ==data/information== that have been organized and ==processed to convey understanding==, experience, accumulated learning, and expertise as they apply to a current business problem

  - R. K. Rainer, C. G. Cegielski, "Introduction to Information Systems", 3rd Edition, John Wiley & Sons, Inc, pp. 10, 2004.

Knowledge

Nếu bệnh nhân có thân nhiệt cao trong vòng 3 ngày, có dấu hiệu mệt mỏi thì bệnh nhân đang có bệnh cúm.

Cho bệnh cúm nhẹ, bệnh nhân cần dùng thuốc …

Nếu thuốc được dùng trong vòng 5 ngày nhưng thân nhiệt không giảm thì bệnh nhân cần nhập viện thực hiện các xét nghiệm về máu, …

…

Kiến trúc của hệ hỗ trợ chẩn đoán dựa trên Web (architecture of a Web-based diagnosis support system)
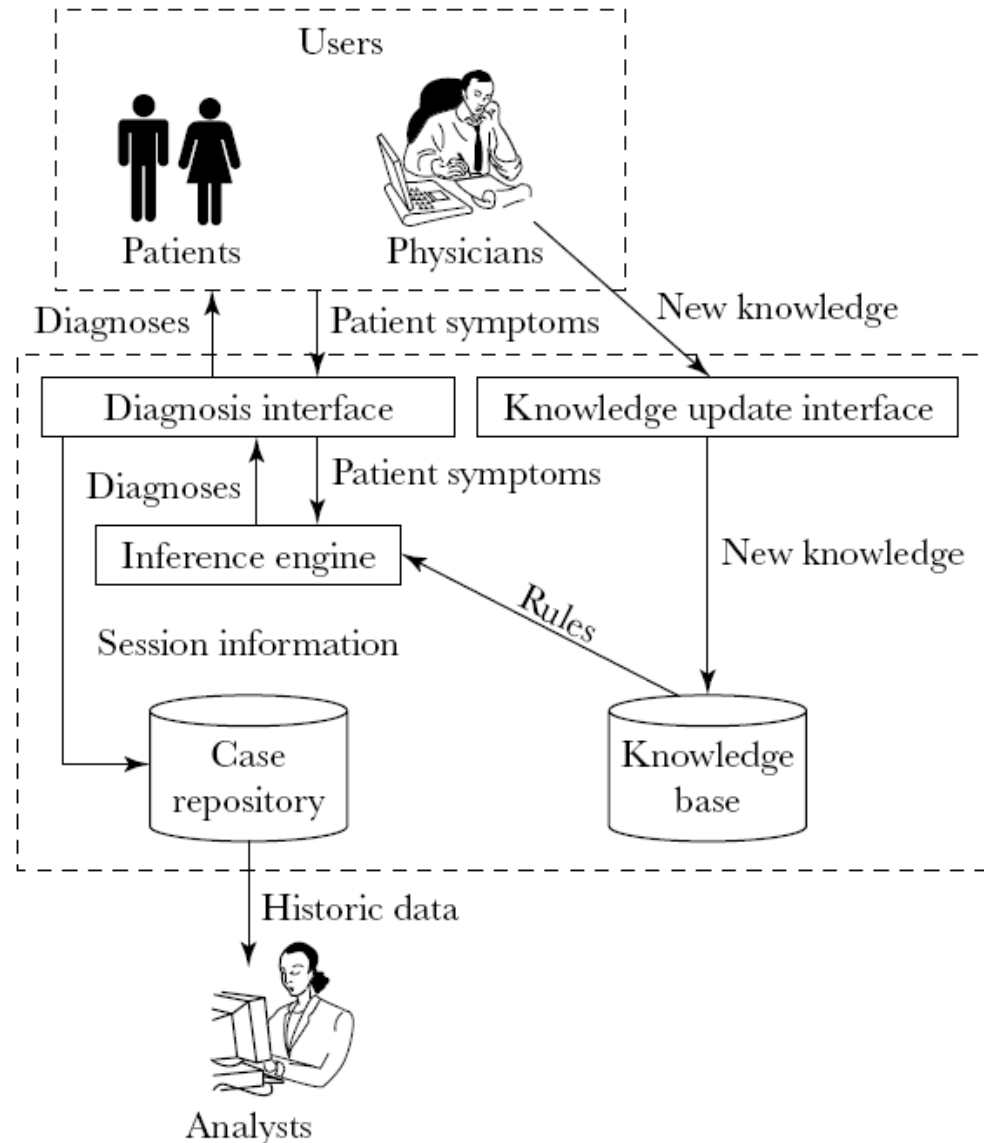
Metadata

- ==Data about data==

- Ví dụ: thông tin ==mô tả kỹ thuật== của 1 word document: title, subject, author, manager, company, …

  - Data: content của word document

  - Metadata: data values của title, subject, author, manager, company, …

Database

- A **collection of related data** with an implicit meaning

- **Implicit** properties

  - A database represents some **aspect of the real world**, called the miniworld or the **universe of discourse (UoD)**. subset

    - Changes to the miniworld are reflected in the database.

  - A database is a **logically coherent collection** of data with some inherent meaning.

    - A random assortment of data cannot correctly be referred to as a database.

  - A database is designed, built, and populated with data for a **specific purpose**.

    - It has an **intended group of users** and some preconceived applications in which these users are interested.

  - A database can be of any size and of varying complexity.

Database

Part of the Company database

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

1.2. File processing systems
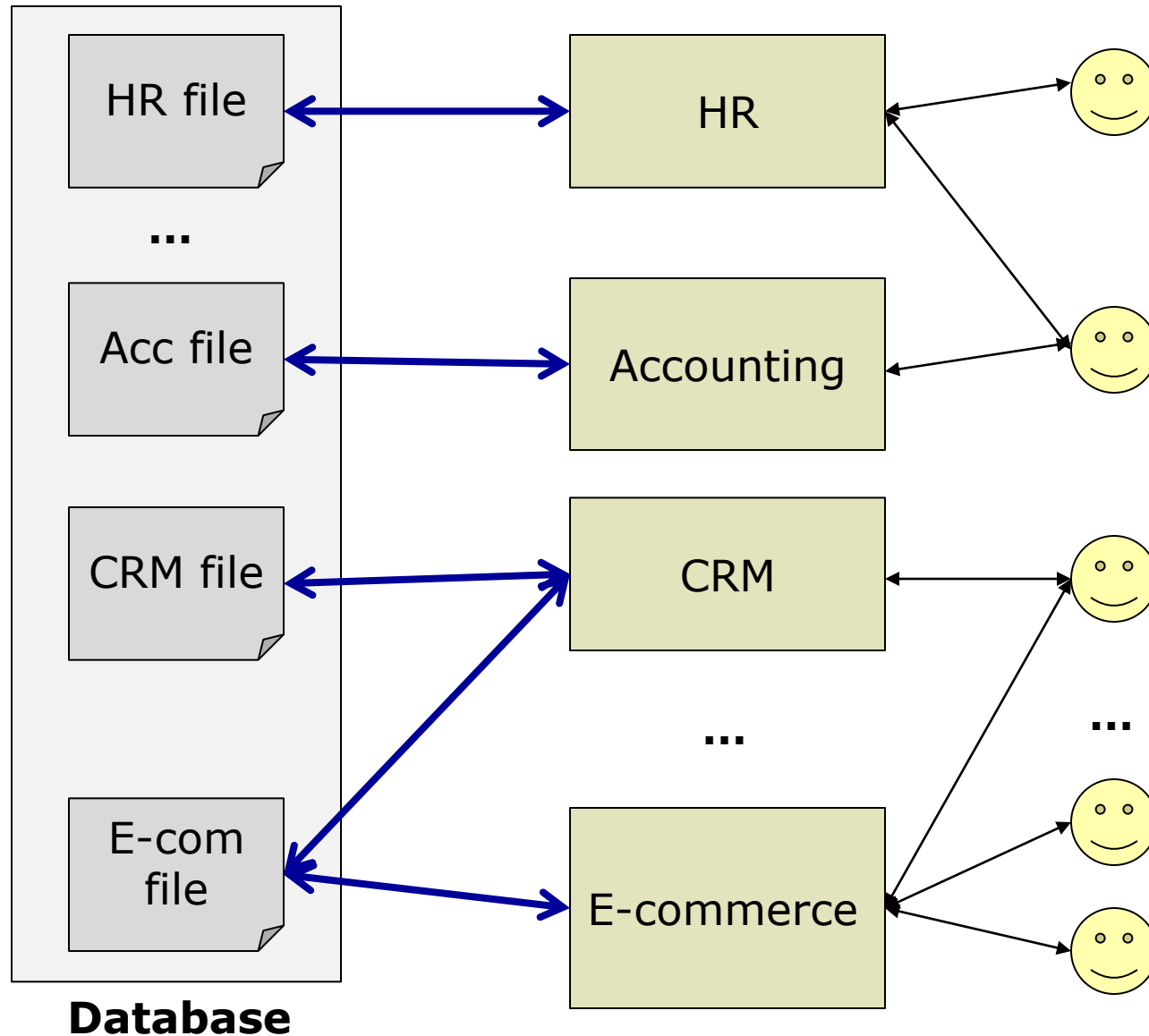
Database

1.3. The database approach
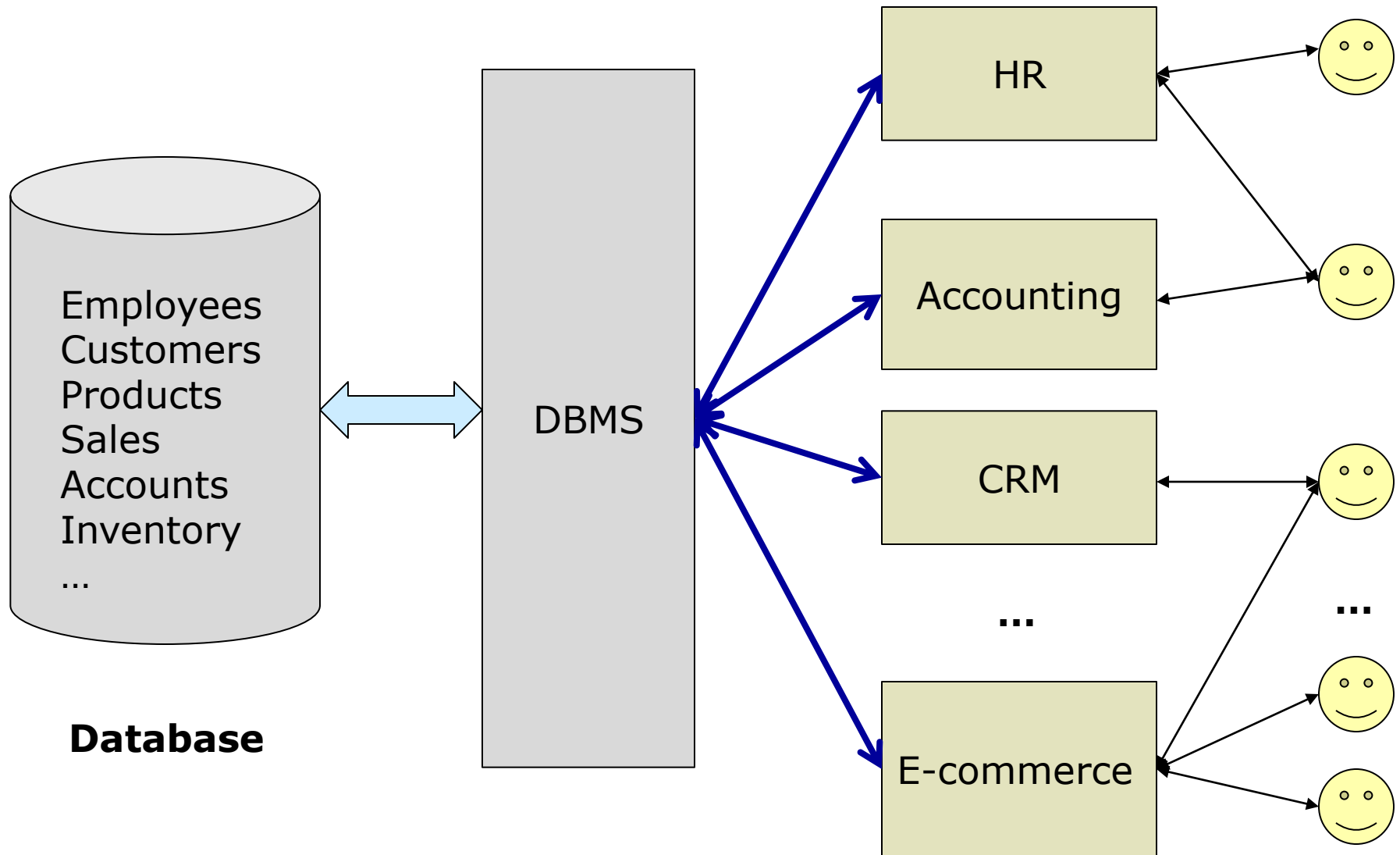
Database

File systems vs. Database systems

- **File**
  - Specifically define and implement the data files for *each user*'s needs
  - *Uncontrolled* data redundancy
  - *No* program-data independence
  - *Hard* maintenance
  - No overhead cost of a DBMS software
  - …

- **Database**
  - Define and implement the repository for various users' needs
  - Controlled data redundancy
  - Program-data independence
  - Easy maintenance
  - *Overhead cost* of a DBMS software
  - …

1.4. Data models

- Informally, a data model is a *type of data abstraction* that is used to provide this *conceptual representation*.

- The data model uses *logical concepts*, such as objects, their properties, and their interrelationships, that may be easier *for most users to understand* than computer storage concepts.

- The data model *hides storage and implementation details* that are not of interest to most database users.

Data model
E. F. Codd. Data models in database management, ACM, 1980.

- A combination of three following components

  - (1). A collection of *data structure types* (the building blocks of any database that conforms to the model);

  - (2). A collection of *operators or inferencing rules*, which can be applied to any valid instances of the data types listed in (1), to retrieve or derive data from any parts of those structures in any combinations desired;

  - (3). A collection of *general integrity rules*, which implicitly or explicitly define the set of consistent database states or changes of state or both ––– these rules may sometimes be expressed as insert-update-delete rules

Data model

- A collection of concepts that can be used to describe *the structure of a database*

  - the data types, relationships, and constraints that should hold for the data

  - a set of basic operations for specifying retrievals and updates on the database

- To provide the necessary means to achieve some *level of abstraction* by hiding details of data storage that are not needed by most database users

Purposes of a data model
E. F. Codd. Data models in database management, ACM, 1980.

- 1. As a tool for specifying the kinds of data and data organization that are permissible in a specific database;

- 2. As a basis for developing a general design methodology for databases;

- 3. As a basis for coping with evolution of databases so as to have minimal logical impact on existing application programs and terminal activities;

- 4. As a basis for the development of families of very high level languages for query and data manipulation;

- 5. As a focus for DBMS architecture;

- 6. As a vehicle for research into the behavioral properties of alternative organizations of data.

Categories of data models

- *High-level* or *conceptual* data models
  - provide concepts that are close to the way many users perceive data
  - e.g. entity relationship model

- *Representational* or *implementation* data models
  - provide concepts that may be understood by end users but that are not too far removed from the way data is organized within the computer
  - hide some details of data storage
  - able to be implemented on a computer system in a direct way
  - e.g. relational data model, object-oriented data model

- *Low-level* or *physical* data models
  - provide concepts that describe the details of how data is stored in the computer

1.5. Database management systems

- a collection of programs that enables users to <mark>create and maintain a database</mark>

- a general-purpose software system that facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications
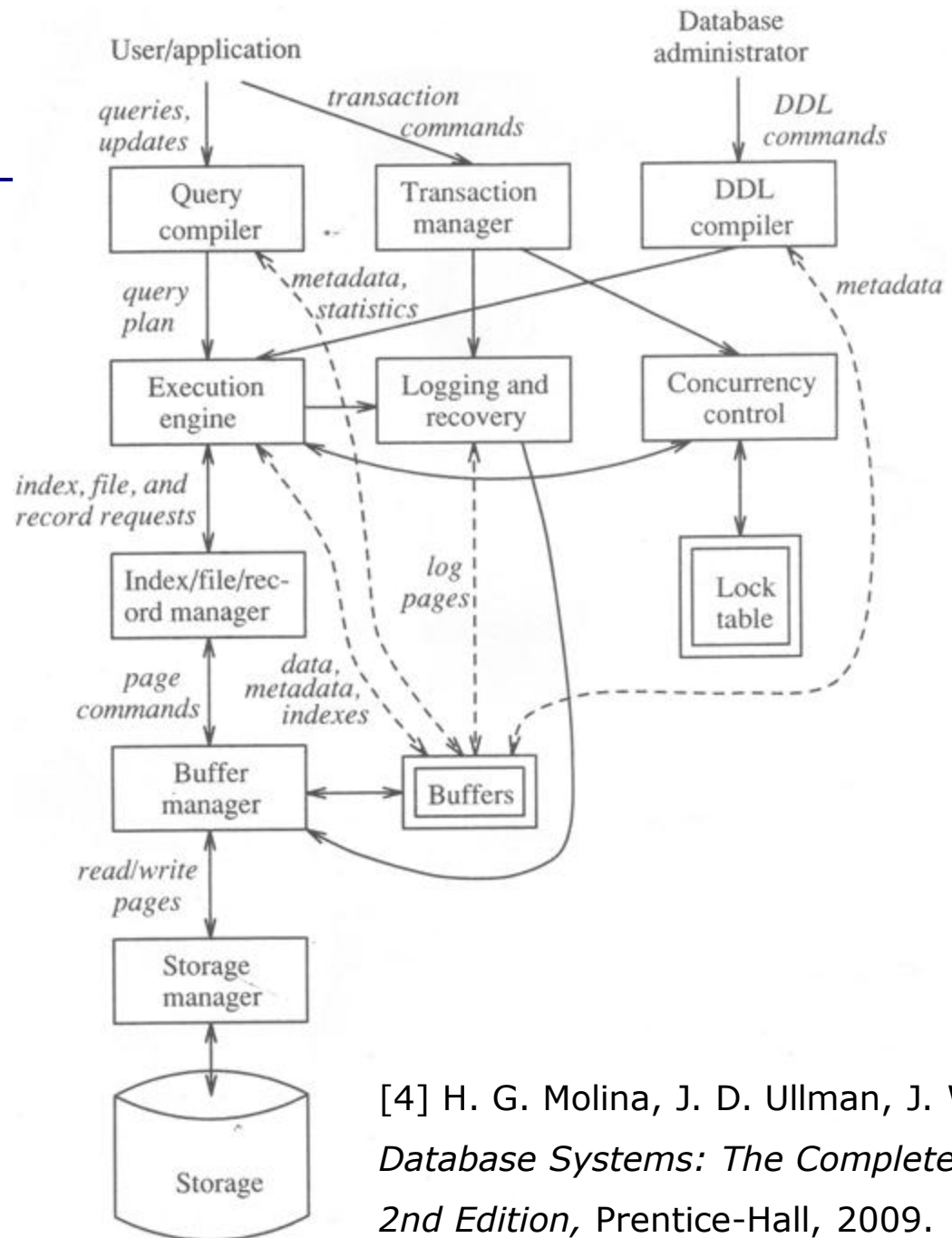
Database management systems

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

- *Constructing* the database is the process of storing the data itself on some storage medium that is controlled by the DBMS.

- *Manipulating* a database includes such functions as querying the database to retrieve specific data, updating the database to reflect changes in the miniworld, and generating reports from the data.

- *Sharing* a database allows multiple users and programs to access the database concurrently.

- *Protection* includes both system protection against hardware or software malfunction (or crashes), and security protection against unauthorized or malicious access.

- A typical large database may have a life cycle of many years, so the DBMS must be able to *maintain* the database system by allowing the system to evolve as requirements change over time.

DBMS components

User/application

queries, updates

transaction commands

Database administrator

DDL commands

Query compiler

Transaction manager

DDL compiler

query plan

metadata, statistics

metadata

Execution engine

Logging and recovery

Concurrency control

index, file, and record requests

log pages

Lock table

Index/file/rec-ord manager

page commands

data, metadata, indexes

Buffer manager

Buffers

read/write pages

Storage manager

Storage

[ ] system component

[[ ]] in-memory structure

——— control/data flow

------ data flow

[4] H. G. Molina, J. D. Ullman, J. Widom, *Database Systems: The Complete Book - 2nd Edition,* Prentice-Hall, 2009.

25

History of DBMS development

- 1960s, navigational DBMSs
  - IBM's IMS with the hierarchical model,
  - IDMS with the CODASYL network model, …
- 1970s-late 1980s, relational DBMSs with SQL
  - Oracle,
  - MS SQL Server,
  - IBM's DB2,
  - MySQL, …
- 1990s, object-oriented DBMSs (object, object-relational)
  - Oracle,
  - PostgreSQL,
  - Informix, …
- 2000s, NoSQL and NewSQL
  - XML DBMSs: Oracle Berkely DB XML, …
 - NoSQL DBMSs: MongoDB, Hbase, Cassandra, …
  - NewSQL DBMSs: ScaleBase, VoltDB, …

Database management system

- When *not* to use
  - Unnecessary overhead costs of using a DBMS
    - High initial investment in hardware, software, and training
    - The generality that a DBMS provides for defining and processing data
    - Overhead for providing security, concurrency control, recovery, and integrity functions
  - The database and applications are simple, well defined, and not expected to change.
  - There are stringent real-time requirements for some programs that may not be met because of DBMS overhead.
  - Multiple-user access to data is not required.

A simplified database system environment
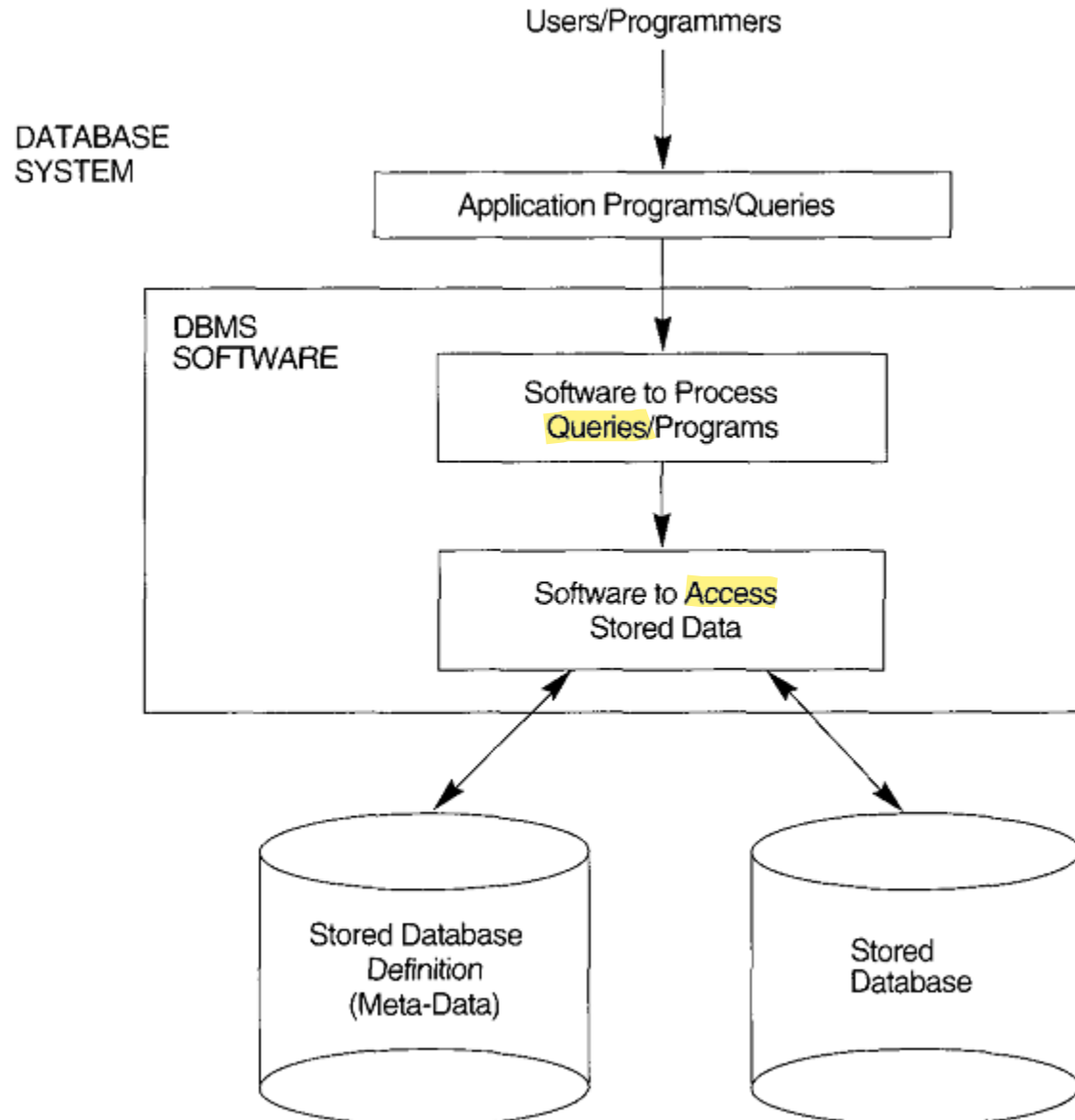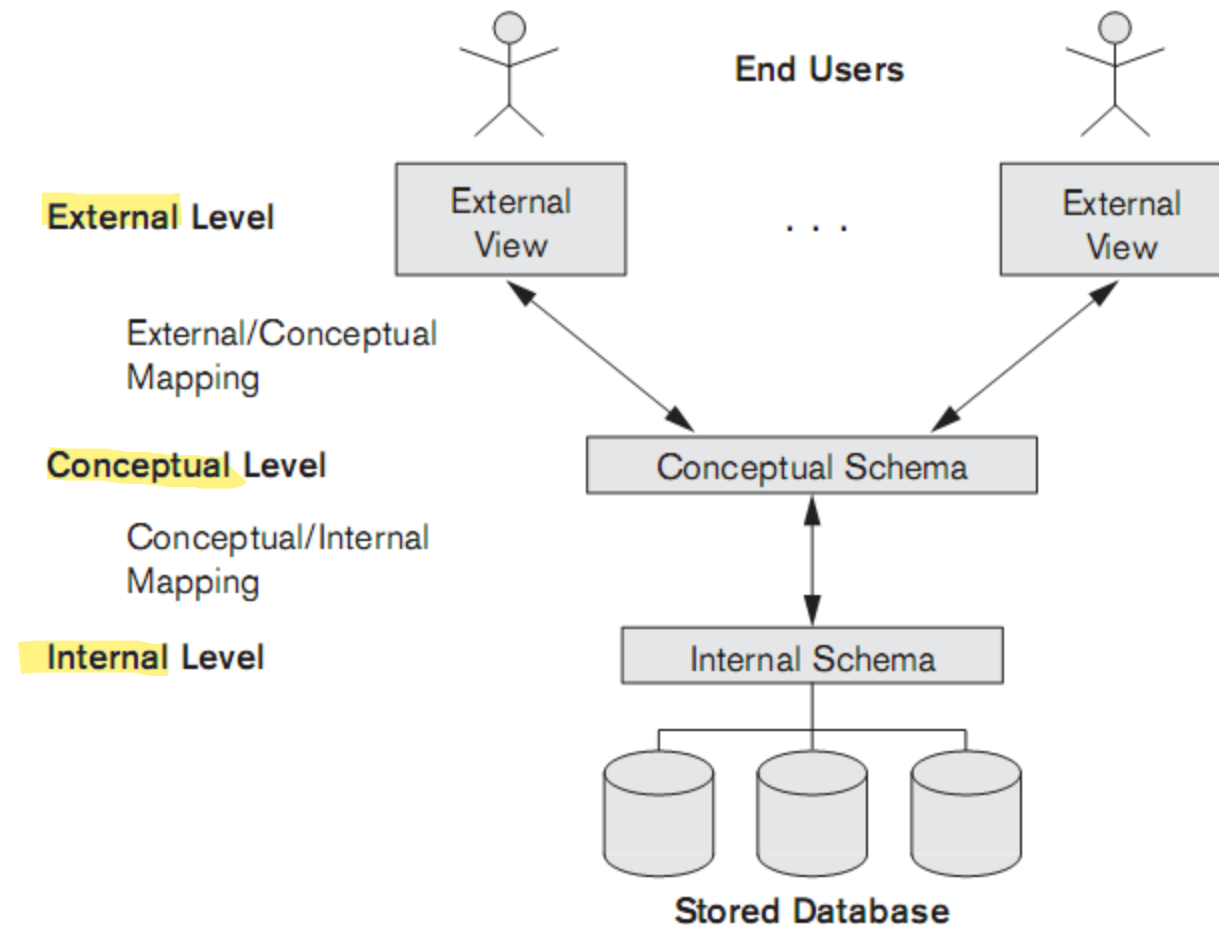
A simplified database system environment
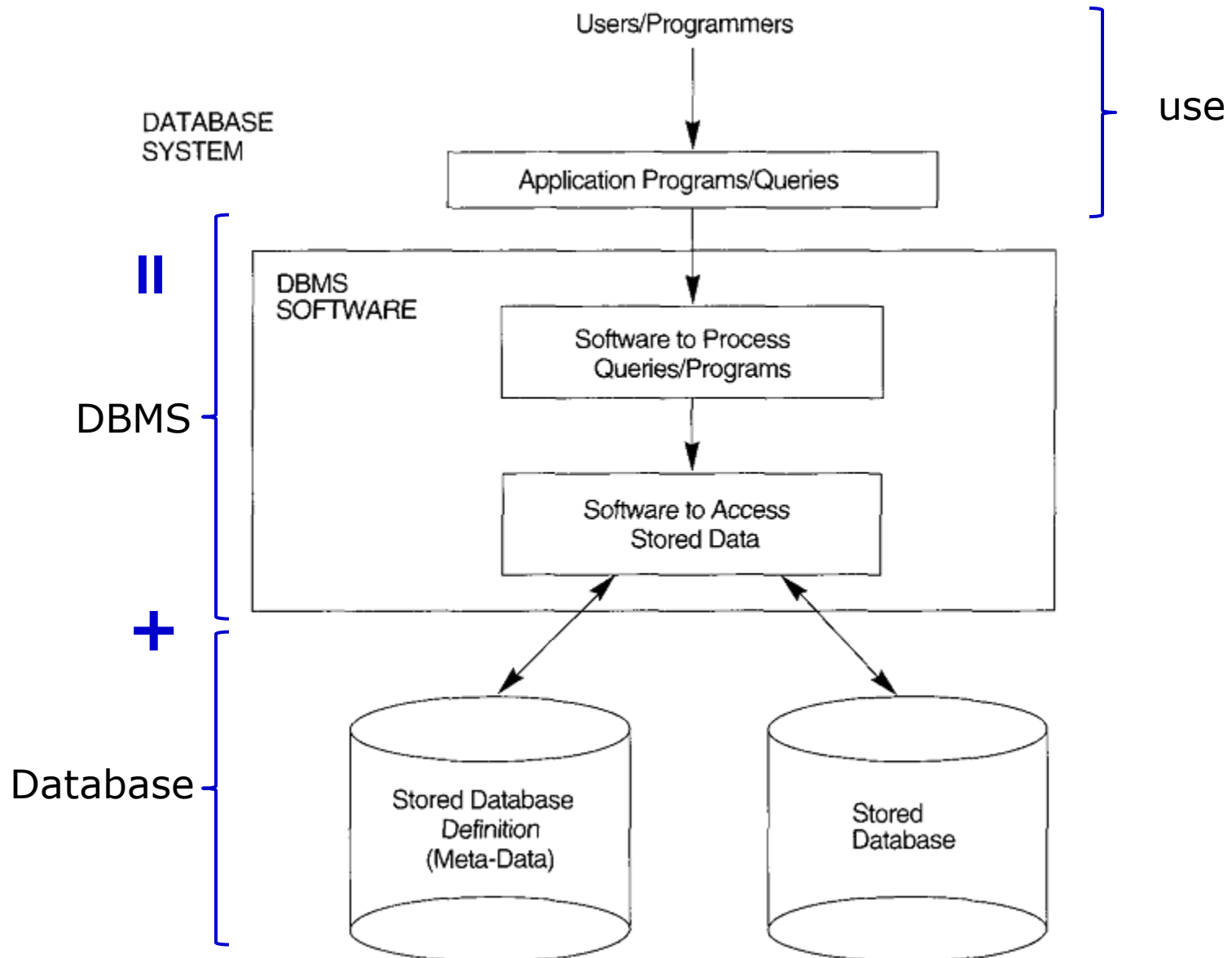
1.6. Database systems

Database system = database + DBMS

- Database: data modeling
- Database management system (DBMS): functionalities
  - File organization & indexing
  - Query processing & optimization
  - Database security
  - Transaction processing & concurrency control
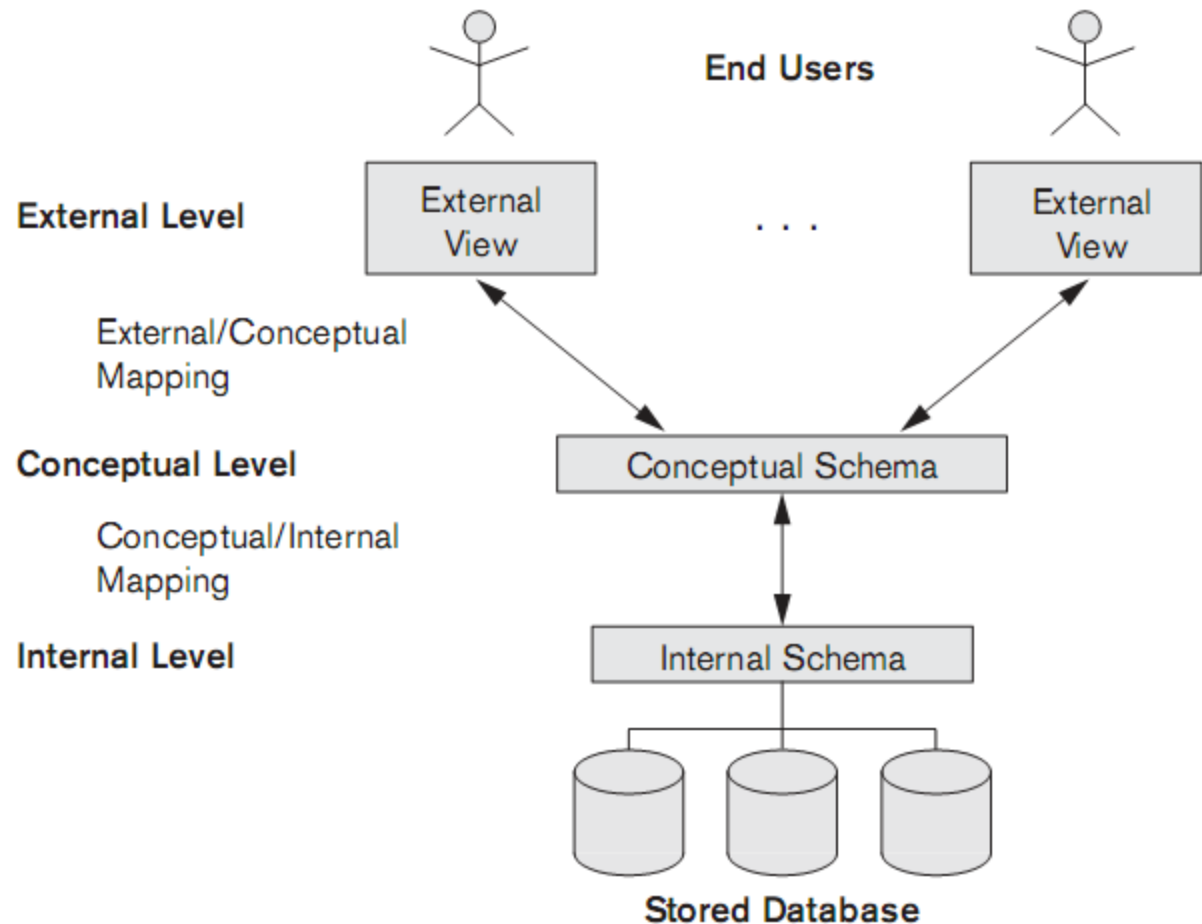 - Backup & recovery

The Three-Schema Architecture

The three-schema architecture

The Three-Schema Architecture

the part of the database that a particular user group is interested in and hides the rest of the database from that user group

the structure of the whole database for a community of users

the physical storage structure of the database

End Users

External Level

External View   . . .   External View

External/Conceptual Mapping

Conceptual Level

Conceptual Schema

Conceptual/Internal Mapping

Internal Level

Internal Schema

Stored Database

The three-schema architecture

The Three-Schema Architecture

- The three-schema architecture

  - An *internal schema* describes the physical storage structure of the database.

  - A *conceptual schema* is a high-level description of the whole database.

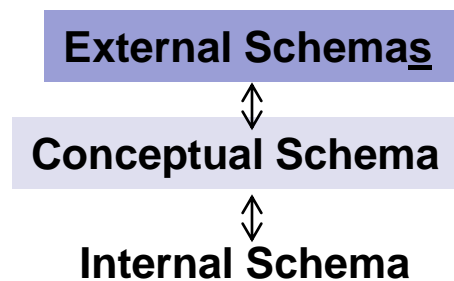 - *External schemas* describe the views of different user groups.

- Data independence

  - Data Independence is the capacity to change the schema at one level of a database system without having to change the schema at the next higher level.

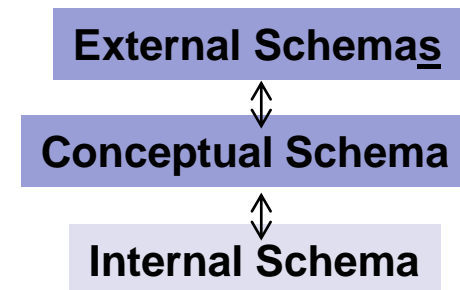 - *Logical* data independence & *Physical* data independence

The Three-Schema Architecture

- Data independence

    - *Logical* data independence: the capacity to change the *conceptual* schema without having to change *external* schemas or application programs

    - *Physical* data independence: the capacity to change the *internal* schema without having to change the *conceptual* schema

External Schemas

↕

Conceptual Schema

↕

Internal Schema

Logical Data Independence

External Schemas

↕

Conceptual Schema

↕

Internal Schema
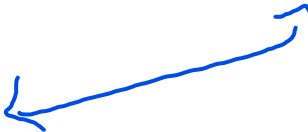
Physical Data Independence

Characteristics of database systems

- **Self-describing** nature of a database system
- **Insulation** between programs and data, and data abstraction
- Support of **multiple views** of the data
- **Sharing** of data and multiuser transaction processing
- Controlling **redundancy**
- Restricting unauthorized **access**
- Providing **persistent storage** for program objects

Characteristics of database systems

- Providing storage structures for efficient query processing

- Providing backup and recovery

- Providing multiple user interfaces

- Representing complex relationships among data

- Enforcing integrity constraints

- Permitting inferencing and actions using rules

Characteristics of database systems

- Potential for enforcing standards

- Reduced application development time

- Flexibility

- Availability of up-to-date information

- Economies of scale

Classification of database systems

- Based on data models (widely-used)

- Based on kinds of data

- Based on data storage and organization

- Based on architectures

- Based on the number of users

Classification of database systems

- Based on data models (widely-used)

  - Relational database systems

  - Object-oriented database systems

  - Object relational database systems

  - XML-enabled database systems

  - XML native database systems

  - Graph database systems

  - …

Classification of database systems

- Based on kinds of data

  - Traditional database systems (simple data)

  - Multimedia database systems

  - Spatial database systems

  - Temporal database systems

  - Spatiotemporal database systems

  - Inductive database systems

  - Deductive database systems

Classification of database systems

❑ Based on data storage and

organization

  ■ Traditional database systems

  ■ In-memory database systems

  ■ Columnar database systems

Classification of database systems

❑ Based on architectures

  ■ Centralized database systems

  ■ Distributed database systems

  ■ Parallel database systems

Classification of database systems

□ Based on the number of users

- Single-user database systems

- Multi-user database systems

→ The number of users who can use the system concurrently – that is, at the same time

1.7. Applications of database systems

- In any organization, in any application domain where there is a need:

  - A large database

  - A multiuser environment

→ Providing application flexibility with relational databases

→ Object-oriented applications and the need for more complex databases

→ Interchanging data on the Web for e-commerce

→ Extending database capabilities for new applications

1.7. Applications of database systems

- *Scientific applications* that store large amounts of data resulting from scientific experiments in areas such as high-energy physics or the mapping of the human genome.

- Storage and retrieval of *images*, from scanned news or personal photographs to satellite photograph images and images from medical procedures such as X-rays or MRI (magnetic resonance imaging).

- Storage and retrieval of *videos*, such as movies, or video clips from news or personal digital cameras.

- *Data mining applications* that analyze large amounts of data searching for the occurrences of specific patterns or relationships.

- *Spatial applications* that store spatial locations of data such as weather information or maps used in geographical information systems.

- *Time series applications* that store information such as economic data at regular points in time, for example, daily sales or monthly gross national product figures.

→ *NEED*: more complex data structures, new data types, new operations and query language constructs, new storage and indexing structures

→ *New general/special purpose functionalities* added to a database system

Summary

- Database system = database + database management system

  - Database

    - Data/ metadata → information/ knowledge

    - Data model (conceptual, logical)

  - Database management system

    - Three-schema architecture & data independence

    - Functionalities

- Characteristics, classification, and applications of database systems

- File processing systems vs. Database systems

Chapter 1: Overall Introduction to Database Systems

Review

- 1.1. Define the following terms: data, database, data model, DBMS, database system, program-data independence, metadata, transaction-processing application.

- 1.3. Discuss the main characteristics of the database approach and how it differs from traditional file systems.

- 1.6. Discuss the capabilities that should be provided by a DBMS.

- 1.8. What is the difference between controlled and uncontrolled redundancy?

Review

- 2.2. Discuss the main categories of data models.

- 2.3. What is the difference between a database schema and a database state?

- 2.4. Describe the three-schema architecture. Why do we need mappings between schema levels?

- 2.5. What is the difference between logical data independence and physical data independence?

- 2.10. Discuss some types of database utilities and tools and their functions.

Next
Chapter 2: The Entity-Relationship Model

- 2.1. Database design process from conceptual modeling

- 2.2. Conceptual data modeling

- 2.3. The entity-relationship model

- 2.4. The extended entity-relationship model