



Chương 1: Giới thiệu tổng quan về Hệ Cơ Sở Dữ Liệu

2

Các định nghĩa và khái niệm cơ bản

Dữ liệu (*Data*)

- Các sự vật, đối tượng trong thế giới thực tạo thành những khối thông tin
 - Dữ liệu là thành phần chính của hệ cơ sở dữ liệu.
 - Dữ liệu không truyền đạt được thông tin có ích.
- **Các loại dữ liệu**
 - Dữ liệu có cấu trúc: số, ngày, chuỗi ký tự, ...
 - Dữ liệu không có cấu trúc: hình ảnh, âm thanh, đoạn phim, ...
 - Siêu dữ liệu (Meta data): dữ liệu dùng để mô tả các tính chất/đặc điểm của dữ liệu.

Thông tin (*Information*)

- ➡ **Dữ liệu** đã được xử lý và truyền tải những **thông tin** có ích



Thông tin (*Information*)

➤ Ví dụ:

➤ Dữ liệu

➤ Thông tin

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145

Class Roster			
Course:	MGT 500 Business Policy	Semester:	Spring 2010
Section:	2		
Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Cơ sở dữ liệu (*Database*)

- Là tập hợp các dữ liệu có quan hệ theo một cách có nghĩa mà nó có thể được truy cập theo những trình tự logic khác nhau
- Đặc điểm của dữ liệu trong cơ sở dữ liệu
 - Có cấu trúc.
 - Có quan hệ
 - Được truy cập theo những trình tự logic khác nhau

Hệ quản trị cơ sở dữ liệu

(*Database Management system-DBMS*)

- Là hệ thống các phần mềm hỗ trợ người dùng có thể phân tích, thiết kế, duy trì và khai thác CSDL
- **Các DBMS thông dụng:** SQLServer, DB2, Oracle ... hầu hết các DBMS hiện nay đều dựa trên **mô hình quan hệ**
- **Tính chất của các hệ quản trị CSDL (DBMS)**
 - Tính sẵn sàng
 - Toàn vẹn
 - Bảo mật
 - Độc lập

Hệ quản trị cơ sở dữ liệu

(*Database Management system-DBMS*)

➤ Chức năng của hệ QTCSDL

- **Định nghĩa cơ sở dữ liệu:** Xác định kiểu dữ liệu, cấu trúc dữ liệu và các ràng buộc dữ liệu lưu trữ trong cơ sở dữ liệu.
- **Xây dựng cơ sở dữ liệu:** việc lưu trữ trong cơ sở dữ liệu và được điều khiển bởi hệ quản trị cơ sở dữ liệu.
- **Thao tác cơ sở dữ liệu**
 - Truy vấn cơ sở dữ liệu (querying the database)
 - Cập nhật cơ sở dữ liệu (updating the database): phản ánh sự thay đổi dữ liệu từ thế giới thực, kết xuất dữ liệu

Đặc điểm của hệ cơ sở dữ liệu

➤ Tính tự mô tả

- Định nghĩa, mô tả cấu trúc và các ràng buộc của cơ sở dữ liệu được lưu trữ trong meta-data.
- Đặc điểm này giúp các phần mềm hệ quản trị cơ sở dữ liệu làm việc với các cơ sở dữ liệu khác nhau.

➤ Tính độc lập dữ liệu và chương trình

- Dữ liệu lưu trữ trong các hệ quản trị cơ sở dữ liệu là riêng biệt do đó, sự thay đổi cấu trúc lưu trữ dữ liệu không ảnh hưởng đến các chương trình truy cập vào CSDL.

Đặc điểm của hệ cơ sở dữ liệu

➤ Tính trừu tượng (Data Abstraction)

- Hệ quản trị cơ sở dữ liệu chỉ cung cấp việc biểu diễn dữ liệu ở mức khái niệm và che giấu chi tiết và cách thức lưu trữ dữ liệu.
- Tính trừu tượng hóa giúp bảo đảm tính độc lập chương trình và dữ liệu.

➤ Hỗ trợ nhiều cách nhìn cơ sở dữ liệu

- Đáp ứng yêu cầu đa người dùng, giúp người có thể xem dữ liệu theo những yêu cầu khác nhau

Đặc điểm của hệ cơ sở dữ liệu

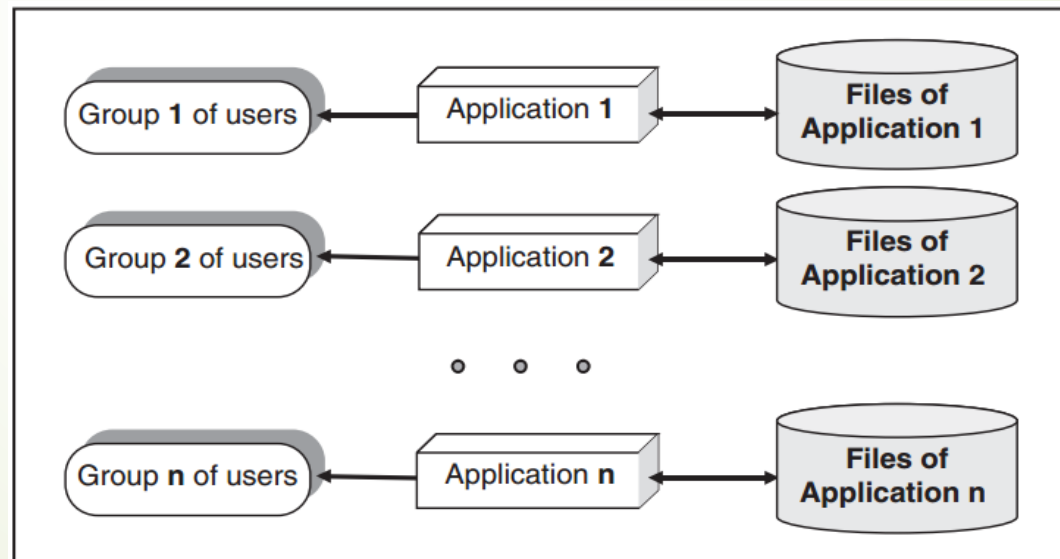
➤ Chia sẻ dữ liệu và xử lý đa tiến trình

- Nhiều người dùng cùng truy cập cơ sở dữ liệu tại cùng một thời điểm
- Cho phép các tiến trình đồng thời bên trong DBMS

Cách tiếp cận hệ thống tập tin

➤ Hệ thống tập tin (file system)

- Trong phát triển phần mềm truyền thống, dùng cách xử lý tập tin, mỗi nhóm người dùng duy trì các tập tin riêng để xử lý các ứng dụng của họ



Cách tiếp cận hệ thống tập tin

➤ Hệ thống tập tin (file system)

➤ Ưu điểm

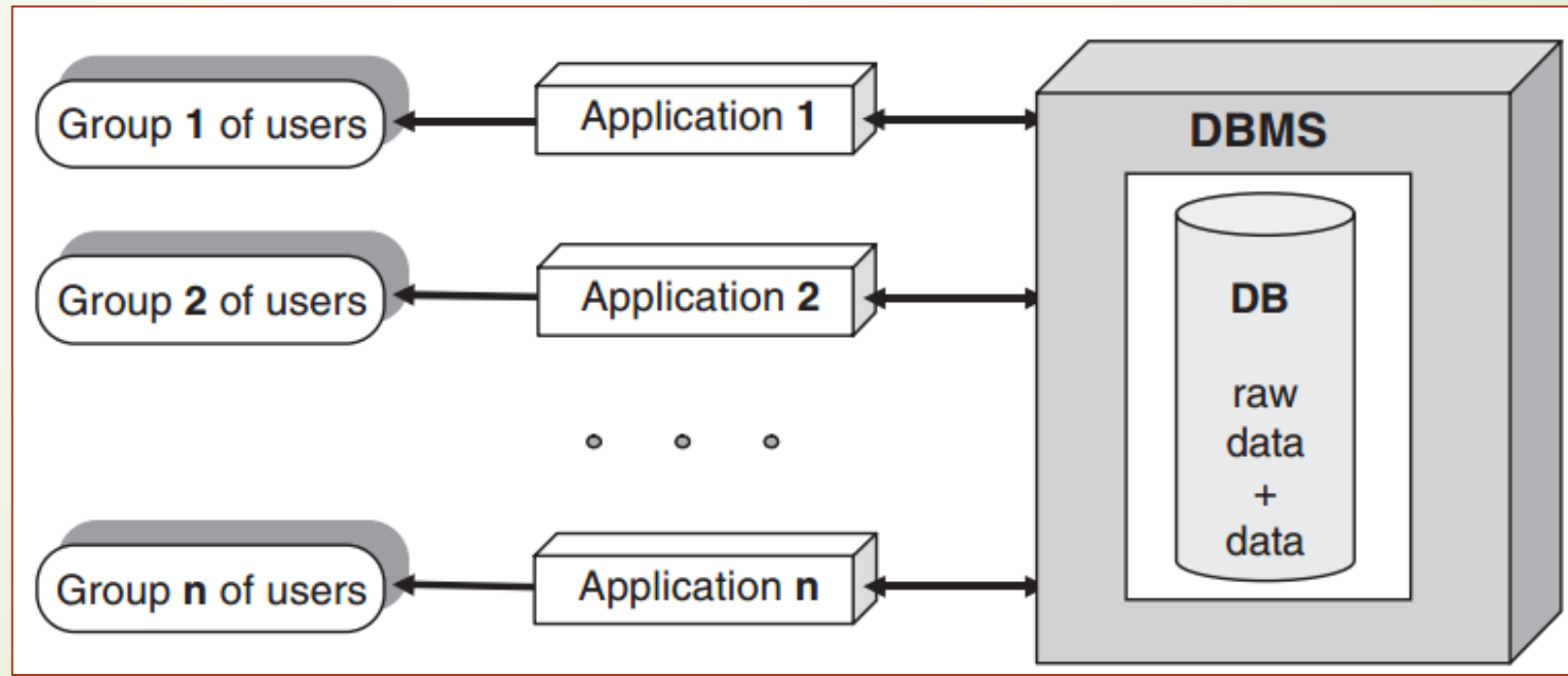
- Triển khai nhanh
- Khả năng đáp ứng nhanh, kịp thời

➤ Nhược điểm

- Trùng lặp dữ liệu → lãng phí, dữ liệu không nhất quán
- Chi phí cao
- Chia sẻ dữ liệu kém
- Khó định dạng
- Có vấn đề về tính toàn vẹn dữ liệu

Cách tiếp cận CSDL

- Các view của các nhóm người dùng khác nhau được tích hợp trong khi thiết kế cơ sở dữ liệu



Thuận lợi của cách tiếp cận CSDL

- **Quản lý dữ liệu tập trung (Centralized data management)**
 - Giảm dư thừa (Reducing redundancies)
 - Quản lý dữ liệu hiệu quả hơn.
 - Cung cấp duy trì việc lưu trữ của các cấu trúc dữ liệu
 - Cung cấp cơ chế sao lưu và phục hồi dữ liệu.
- **Tính độc lập dữ liệu (Data Independence)**
 - Các chương trình độc lập với sự thay đổi cấu trúc lưu trữ dữ liệu.
 - DBMS cung cấp giao diện giữa chương trình ứng dụng và dữ liệu.

Mô tả và lưu trữ dữ liệu trong Hệ QTCSDL

Mô hình dữ liệu - Data model

- Một mô hình dữ liệu là một tập hợp các mô tả cấu trúc dữ liệu mức cao và che giấu nhiều chi tiết lưu trữ chi tiết.
- Một DBMS cho phép người dùng xác định các dữ liệu được lưu trữ trong điều kiện của một mô hình dữ liệu.
- Hầu hết các DBMS hiện nay đều dựa trên mô hình dữ liệu quan hệ (*the relational data model*)

Mô hình quan hệ - The Relational Model

- Mô hình quan hệ là mô hình mà cấu trúc chính dùng mô tả dữ liệu trong mô hình là các **quan hệ** (*a relation*)
 - **Lược đồ quan hệ (relation schema)**
 - Mô tả dữ liệu trong một quan hệ
 - Một lược đồ quan hệ được xác định bởi tên lược đồ và danh sách các thuộc tính
- Ví dụ: Lược đồ quan hệ ***sinhvien***(***MaSV, Hoten, Ngaysinh***)

Mô hình quan hệ - The Relational Model

- **Thể hiện của lược đồ quan hệ (instances):** một bảng giá trị bao gồm các cột và các dòng

Ví dụ: thể hiện của lược đồ quan hệ

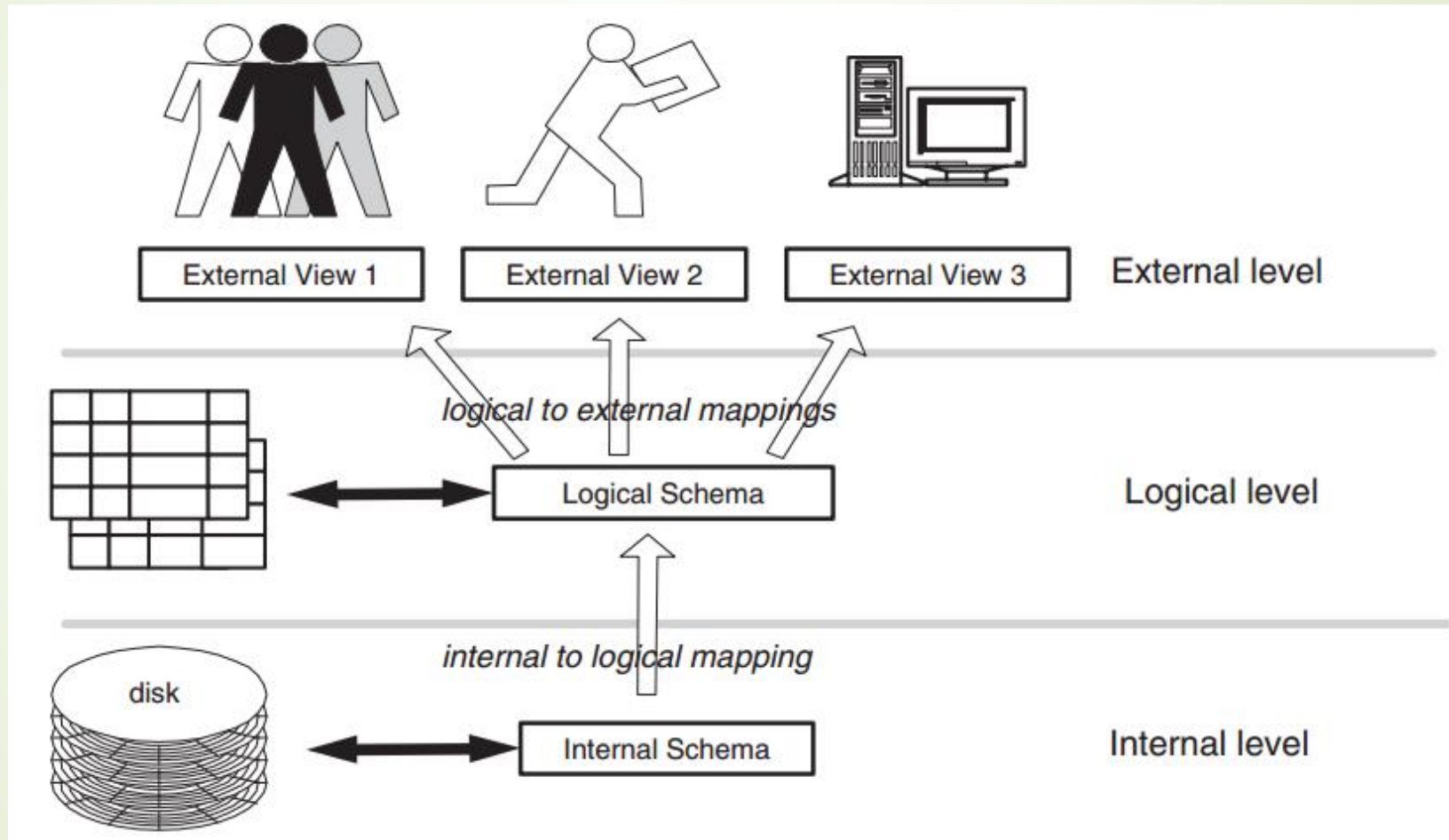
sinhvien(MaSV, Hoten, Ngaysinh)

MaSV	Hoten	Ngaysinh
001	Nguyễn Nhật Minh	10/12/1999
002	Lê Thanh Phúc	20/11/1998
003	Trần Hoài Nam	23/12/1999

Các mức trừu tượng của dữ liệu

- Dữ liệu trong DBMS được mô tả ở 3 mức.
 - Mức vật lý (internal level)
 - Mức logic (conceptual level)
 - Mức người dùng (external level)

Các mức trừu tượng của dữ liệu



Các mức trừu tượng của dữ liệu

➤ Mức vật lý:

- Lưu trữ vật lý của các thông tin
- Hiện thực các mô tả logic của các lược đồ
- Cấu trúc dữ liệu mức chi tiết
- Dữ liệu được xây dựng theo 5 loại mô hình: network, hierarchical, relational, object-oriented, object-relational

Các mức trừu tượng của dữ liệu

➤ Mức logic

- Mô tả cách mà dữ liệu được lưu trữ trong CSDL và mối quan hệ giữa chúng
- Trong hệ quản trị CSDL quan hệ, **lược đồ khái niệm** mô tả tất cả các **quan hệ** được lưu trữ trong CSDL, các quan hệ này chứa thông tin về các **thực thể** và mối **quan hệ**.

➤ Mức người dùng

- Đây là mức trừu tượng cao nhất, có nhiều view khác nhau cho các nhóm người dùng khác nhau trên cùng một loại dữ liệu.

Các mức trừu tượng của dữ liệu

- Mỗi mức trừu tượng có một lược đồ mô tả cơ sở dữ liệu tương ứng
 - Lược đồ khái niệm (conceptual schemas)
 - Lược đồ vật lý (physical schemas)
 - Lược đồ ngoài (external schemas)

Lược đồ khái niệm

- Mô tả cách lưu trữ dữ liệu của mô hình dữ liệu trong **hệ quản trị cơ sở dữ liệu**.
- Trong hệ quản trị cơ sở dữ liệu quan hệ, **lược đồ khái niệm** (*the conceptual schema*) mô tả tất cả các **quan hệ** được lưu trữ trong **cơ sở dữ liệu**

Lược đồ khái niệm

➤ Ví dụ:

STUDENT

Name	Student_number	Class	Major
------	----------------	-------	-------

COURSE

Course_name	Course_number	Credit_hours	Department
-------------	---------------	--------------	------------

PREREQUISITE

Course_number	Prerequisite_number
---------------	---------------------

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
--------------------	---------------	----------	------	------------

GRADE_REPORT

Student_number	Section_identifier	Grade
----------------	--------------------	-------

Lược đồ vật lý

- Specifies additional storage details. *The physical schema* summarizes how the relations described in the *conceptual schema* are actually stored on secondary storage devices such as disks and tapes

Lược đồ ngoài

- ***External schemas*** allow data access to be customized (and authorized) at the level of individual users or groups of users.
- Each external schema consists of a collection of one or more views and relations from the *conceptual schema*.

Ngôn ngữ CSDL

- **Ngôn ngữ định nghĩa dữ liệu (DDL):** được sử dụng bởi người quản trị CSDL và người thiết kế CSDL dùng để định nghĩa lược đồ khái niệm của CSDL.
- **Ngôn ngữ định nghĩa lưu trữ (SDL):** dùng để xác định lược đồ trong (mức vật lý).

Ngôn ngữ CSDL

➤ Ngôn ngữ thao tác dữ liệu (DML):

- Được sử dụng để truy vấn và cập nhật dữ liệu. Có hai loại ngôn ngữ thao tác dữ liệu

➤ Ngôn ngữ định nghĩa view người dùng (VDL):

- Xác định view của người dùng và ánh xạ view này với lược đồ khái niệm.