

## *Bài 1*

# **Tổng quan về cơ sở dữ liệu**

## **Nội dung trình bày**

---

- Các định nghĩa cơ bản
- Đặc điểm của phương pháp csdl
- Người sử dụng csdl
- Mô hình dữ liệu
- Lược đồ và trạng thái csdl
- Kiến trúc 3 lược đồ
- Ngôn ngữ của DBMS
- Quá trình phát triển

## Các định nghĩa cơ bản (1)

---

- **Dữ liệu** là các dữ kiện có thể được ghi lại và có ý nghĩa ngầm định.
  - Ví dụ: họ tên, số điện thoại, địa chỉ của những người mà bạn biết.
- **Cơ sở dữ liệu** là một tập hợp dữ liệu có liên quan:
  - Biểu diễn một vài khía cạnh của thế giới thực.
  - Hợp lý về mặt logic.
  - Dành cho một mục đích cụ thể; một nhóm người dùng dự định; một số ứng dụng xác định.

---

3

## Các định nghĩa cơ bản (2)

---

- **Hệ quản trị csdl (DBMS)** là một hệ thống phần mềm ứng dụng cung cấp các chức năng giúp tạo và duy trì csdl trên máy tính:
  - *Định nghĩa*: chỉ định kiểu, cấu trúc và các ràng buộc của dữ liệu.
  - *Xây dựng*: lưu trữ dữ liệu vào các thiết bị lưu trữ.
  - *Thao tác*: truy xuất và sửa đổi dữ liệu.
  - *Chia sẻ*: cho phép nhiều người dùng và các chương trình truy cập đồng thời.
  - *Bảo vệ*: chống lại sự cố phần cứng, phần mềm, sự truy cập trái phép.
  - *Duy trì*: cho phép thay đổi khi có yêu cầu

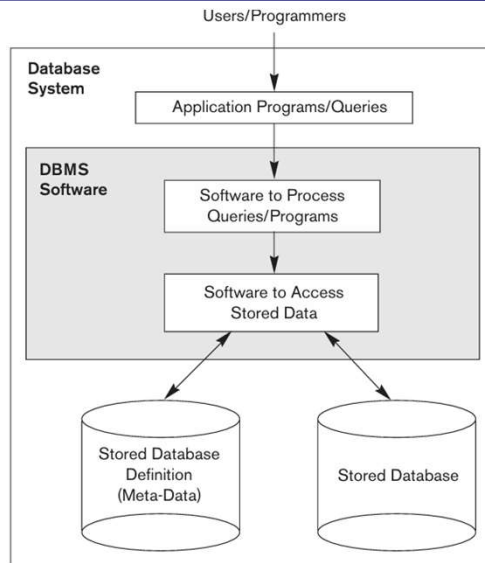
---

4

## Các định nghĩa cơ bản (3)

### ▪ Hệ thống cơ sở dữ liệu bao gồm:

- DBMS
- Các csdl
- Các ứng dụng csdl



5

## Ví dụ về một cơ sở dữ liệu (1)

Một phần csdl của môi trường UNIVERSITY trong thế giới thực gồm:

### ▪ Một vài thực thể:

- STUDENTs - sinh viên
- COURSEs - học phần
- SECTIONs - lớp học (của học phần)
- INSTRUCTORs - giảng viên (dạy các lớp)

### ▪ Một vài mối kết hợp giữa các thực thể:

- SECTIONs của COURSEs
- STUDENTs dự SECTIONs
- COURSEs có học phần trước COURSEs
- INSTRUCTORs dạy SECTIONs

6

## Ví dụ về một cơ sở dữ liệu (2)

### COURSE

| Course_name               | Course_number | Credit_hours | Department |
|---------------------------|---------------|--------------|------------|
| Intro to Computer Science | CS1310        | 4            | CS         |
| Data Structures           | CS3320        | 4            | CS         |
| Discrete Mathematics      | MATH2410      | 3            | MATH       |
| Database                  | CS3380        | 3            | CS         |

### SECTION

| Section_identifier | Course_number | Semester | Year | Instructor |
|--------------------|---------------|----------|------|------------|
| 85                 | MATH2410      | Fall     | 07   | King       |
| 92                 | CS1310        | Fall     | 07   | Anderson   |
| 102                | CS3320        | Spring   | 08   | Knuth      |
| 112                | MATH2410      | Fall     | 08   | Chang      |
| 119                | CS1310        | Fall     | 08   | Anderson   |
| 135                | CS3380        | Fall     | 08   | Stone      |

### STUDENT

| Name  | Student_number | Class | Major |
|-------|----------------|-------|-------|
| Smith | 17             | 1     | CS    |
| Brown | 8              | 2     | CS    |

**Figure 1.2**

A database that stores student and course information.

## Đặc điểm của phương pháp CSDL (1)

- Tính tự mô tả của hệ thống csdl
  - Lưu trữ csdl và cả thông tin mô tả về nó.
  - Thông tin mô tả được chứa trong danh mục của DBMS.
  - Cho phép DBMS hoạt động với các csdl khác nhau.

### RELATIONS

| Relation_name | No_of_columns |
|---------------|---------------|
| STUDENT       | 4             |
| COURSE        | 4             |
| SECTION       | 5             |
| GRADE_REPORT  | 3             |
| PREREQUISITE  | 2             |

### COLUMNS

| Column_name         | Data_type      | Belongs_to_relation |
|---------------------|----------------|---------------------|
| Name                | Character (30) | STUDENT             |
| Student_number      | Character (4)  | STUDENT             |
| Class               | Integer (1)    | STUDENT             |
| Major               | Major_type     | STUDENT             |
| Course_name         | Character (10) | COURSE              |
| Course_number       | XXXXNNNN       | COURSE              |
| ....                | ....           | ....                |
| ....                | ....           | ....                |
| ....                | ....           | ....                |
| Prerequisite_number | XXXXNNNN       | PREREQUISITE        |

**Figure 1.3**

An example of a database catalog for the database in Figure 1.2.

## Đặc điểm của phương pháp CSDL (2)

- Độc lập chương trình - dữ liệu
  - Cho phép thay đổi cấu trúc và tổ chức lưu trữ của dữ liệu mà không cần phải thay đổi các ứng dụng truy cập.
- Trừu tượng hóa dữ liệu
  - Cung cấp cho người dùng biểu diễn khái niệm về dữ liệu và ẩn đi các chi tiết về lưu trữ và triển khai.

| Data Item Name | Starting Position in Record | Length in Characters (bytes) |
|----------------|-----------------------------|------------------------------|
| Name           | 1                           | 30                           |
| Student_number | 31                          | 4                            |
| Class          | 35                          | 1                            |
| Major          | 36                          | 4                            |

**Figure 1.4**  
Internal storage format for a STUDENT record,  
based on the database catalog in Figure 1.3.

9

## Đặc điểm của phương pháp CSDL (3)

- Hỗ trợ nhiều chế độ xem dữ liệu
  - Mỗi người dùng có một chế độ xem csdl khác nhau.

### TRANSCRIPT

| Student_name | Student_transcript |       |          |      |            |
|--------------|--------------------|-------|----------|------|------------|
|              | Course_number      | Grade | Semester | Year | Section_id |
| Smith        | CS1310             | C     | Fall     | 08   | 119        |
|              | MATH2410           | B     | Fall     | 08   | 112        |
| Brown        | MATH2410           | A     | Fall     | 07   | 85         |
|              | CS1310             | A     | Fall     | 07   | 92         |
|              | CS3320             | B     | Spring   | 08   | 102        |
|              | CS3380             | A     | Fall     | 08   | 135        |

(a)

### COURSE\_PREREQUISITES

| Course_name     | Course_number | Prerequisites |
|-----------------|---------------|---------------|
| Database        | CS3380        | CS3320        |
|                 |               | MATH2410      |
| Data Structures | CS3320        | CS1310        |

(b)

**Figure 1.5**  
(a) The TRANSCRIPT view.  
(b) The COURSE\_PREREQUISITES view.

10

## Đặc điểm của phương pháp CSDL (4)

---

- Chia sẻ dữ liệu
  - Cho phép nhiều người dùng truy cập csdl cùng một lúc.
- Xử lý giao dịch đa người dùng
  - Tính năng kiểm soát đồng thời cho phép nhiều giao dịch thực hiện cùng lúc mà không xung đột với nhau, đảm bảo mỗi giao dịch được thực hiện chính xác hoặc hủy bỏ.
  - Tính năng xử lý giao dịch trực tuyến cho phép hàng trăm giao dịch đồng thời được thực hiện trong mỗi giây.

---

11

## Người sử dụng CSDL

---

- Nhóm người làm việc thường xuyên với csdl
  - Quản trị viên csdl
  - Thiết kế viên csdl
  - Người sử dụng cuối
  - Phân tích viên hệ thống
  - Lập trình viên ứng dụng csdl
  - Phân tích viên kinh doanh
- Nhóm người duy trì môi trường hệ csdl
  - Người thiết kế và triển khai DBMS
  - Người phát triển công cụ
  - Người vận hành và bảo trì hệ thống csdl

---

12

## Mô hình dữ liệu

---

### ▪ **Mô hình dữ liệu (data model)**

- Là tập hợp các khái niệm để mô tả cấu trúc của csdl (kiểu, các mối liên kết, các ràng buộc của dữ liệu).
- Hầu hết có các thao tác để truy cập csdl
  - Thao tác cơ bản
  - Thao tác do người dùng định nghĩa
- Phương tiện để trừu tượng hóa dữ liệu.

---

13

## Các dạng mô hình dữ liệu (1)

---

### ▪ **Mô hình dữ liệu khái niệm (cấp cao)**

- Cung cấp các khái niệm gần gũi với cách người dùng cảm nhận dữ liệu.
- Còn được gọi là mô hình dữ liệu dựa trên thực thể hoặc dựa trên đối tượng.
- Ví dụ: mô hình thực thể - mối kết hợp,

### ▪ **Mô hình dữ liệu vật lý (cấp thấp)**

- Cung cấp các khái niệm mô tả chi tiết về cách dữ liệu được lưu trữ trong máy tính.
- Thường được chỉ định theo cách đặc biệt thông qua các hướng dẫn quản trị và thiết kế DBMS.

---

14

## Các dạng mô hình dữ liệu (2)

### ▪ Mô hình dữ liệu triển khai (biểu diễn)

- Cung cấp các khái niệm nằm giữa hai khái niệm trên, được dùng bởi nhiều DBMS thương mại.
- Ví dụ: mô hình quan hệ.

### ▪ Mô hình dữ liệu tự mô tả

- Kết hợp mô tả dữ liệu với các giá trị dữ liệu.
- Ví dụ: XML, kho lưu trữ khóa-giá trị, và một số hệ thống NoSQL.

15

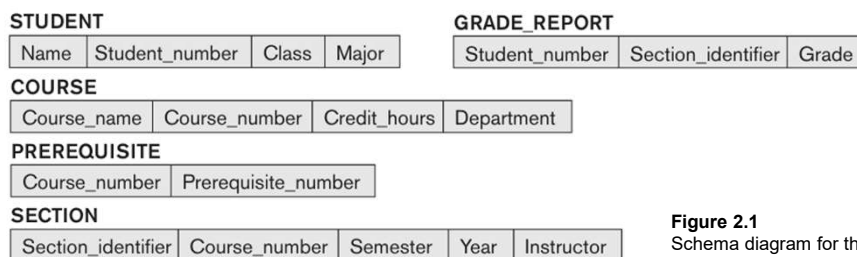
## Lược đồ csdl

### ▪ Lược đồ csdl (database schema)

- Mô tả của một csdl, bao gồm các mô tả về cấu trúc csdl, kiểu của dữ liệu và các ràng buộc trên csdl.

### ▪ Sơ đồ lược đồ (schema diagram)

- Biểu diễn bằng hình vẽ hầu hết các khía cạnh của một lược đồ csdl.



**Figure 2.1**  
Schema diagram for the  
database in Figure 1.2

16



## Trạng thái csdl

### ▪ Trạng thái csdl (database state)

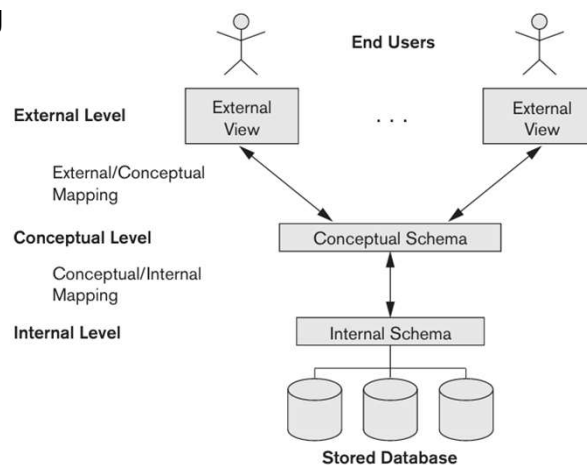
- Tất cả dữ liệu được lưu trữ trong csdl tại một thời điểm cụ thể. Còn được gọi là **thể hiện csdl**.
- **Trạng thái rỗng** là trạng thái tại thời điểm csdl được định nghĩa trong DBMS.
- **Trạng thái ban đầu** là trạng thái tại thời điểm tất cả dữ liệu lần đầu được đưa vào hệ thống.
- **Trạng thái hợp lệ** là trạng thái mà tất cả dữ liệu thỏa mãn các ràng buộc của csdl.
- Trạng thái csdl thay đổi mỗi khi csdl được cập nhật.

17

## Kiến trúc 3 lược đồ

### ▪ Kiến trúc 3 lược đồ để hỗ trợ các đặc điểm:

- Độc lập chương trình - dữ liệu.
- Nhiều chế độ xem dữ liệu.
- Độc lập logic.
- Độc lập vật lý.



18

## Ngôn ngữ của DBMS

---

- Ngôn ngữ định nghĩa dữ liệu (DDL)
  - Được sử dụng để định nghĩa lược đồ khái niệm của csdl.
  - Trong nhiều DBMS, DDL cũng được sử dụng để xác định các lược đồ trong và ngoài.
- Ngôn ngữ thao tác dữ liệu (DML)
  - Được sử dụng để truy cập csdl.
  - Các lệnh DML có thể được nhúng vào ngôn ngữ lập trình.
  - Các lệnh DML độc lập có thể được sử dụng trực tiếp được gọi là ngôn ngữ truy vấn.

---

19

## Quá trình phát triển

---

- Giữa 1960s: hệ csdl trên mô hình phân cấp và mô hình mạng.
- Đầu 1970s: hệ csdl trên mô hình quan hệ.
- Đầu 1980s: DBMS quan hệ.
- Cuối 1980s và đầu 1990s: các ứng dụng csdl hướng đối tượng với
  - DBMS hướng đối tượng.
  - DBMS quan hệ tích hợp các khái niệm hướng đối tượng.
- Giữa 1990s: các hệ csdl quan hệ thêm XML và các dữ liệu đa phương tiện.
- Đầu 2000s: hệ dữ liệu lớn với các csdl phi quan hệ.

---

20