

Chương 2: Các mô hình dữ liệu

GV: Nguyễn Hồng Hạnh

Email: hanhnh@nuce.edu.vn

Bộ môn Công nghệ Phần mềm

Khoa Công nghệ Thông Tin, Đại học Xây dựng

1

1

Nội dung

- Tổng quan về mô hình dữ liệu
- Mô hình phân cấp
- Mô hình mạng
- Mô hình quan hệ
- Mô hình thực thể liên kết
- Mô hình hướng đối tượng

2

2

Tổng quan về mô hình dữ liệu

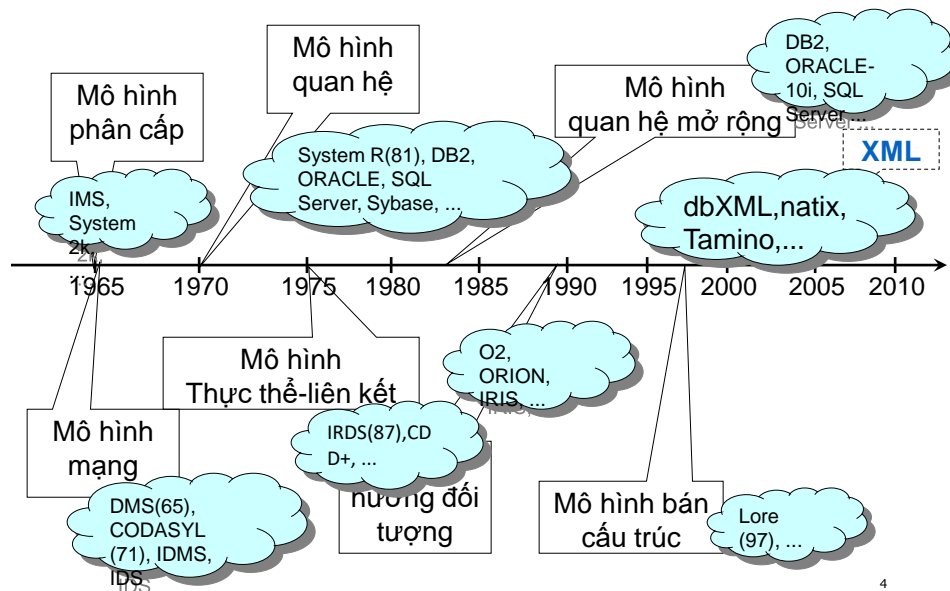
- Mô hình dữ liệu [Codd, 1980] gồm:
 - Một tập hợp các cấu trúc của dữ liệu
 - Một tập hợp các phép toán để thao tác với các dữ liệu
 - Một tập hợp các ràng buộc về dữ liệu
- Mô hình dữ liệu là một tập hợp các khái niệm dùng để mô tả:
 - Dữ liệu
 - Ngữ nghĩa của dữ liệu
 - Các mối quan hệ trong dữ liệu
 - Các ràng buộc dữ liệu

Nhiều mô hình còn bao gồm cả một tập các phép toán để thao tác các dữ liệu

3

3

Vài nét về lịch sử



4

4

Tổng quan...(tiếp)

- Mô hình thuộc dạng ngữ nghĩa: tập trung về ngữ nghĩa của dữ liệu như mô hình thực thể liên kết, sử dụng để hỗ trợ người dùng có cái nhìn khái quát về dữ liệu
- Mô hình thuộc dạng khái niệm: tập trung vào cách thức tổ chức dữ liệu tại mức khái niệm như mô hình mạng, mô hình liên kết, mô hình quan hệ, độc lập với DBMS và hệ thống phần cứng để cài đặt cơ sở dữ liệu

5

5

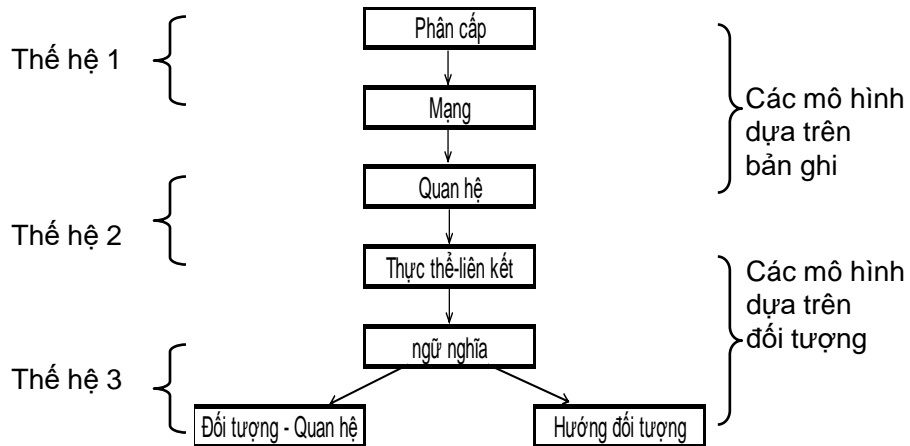
Một vài mô hình dữ liệu

- Mô hình phân cấp
- Mô hình mạng
- Mô hình quan hệ
- Mô hình thực thể liên kết
- Mô hình hướng đối tượng
- Mô hình bán cấu trúc
- Mô hình dữ liệu của XML

6

6

Phân loại các mô hình



7

7

Đặt vấn đề

- Đặc điểm của các mô hình dữ liệu?
- Sự khác nhau giữa các mô hình dữ liệu?
- Các mô hình dữ liệu phổ biến ngày nay

8

8

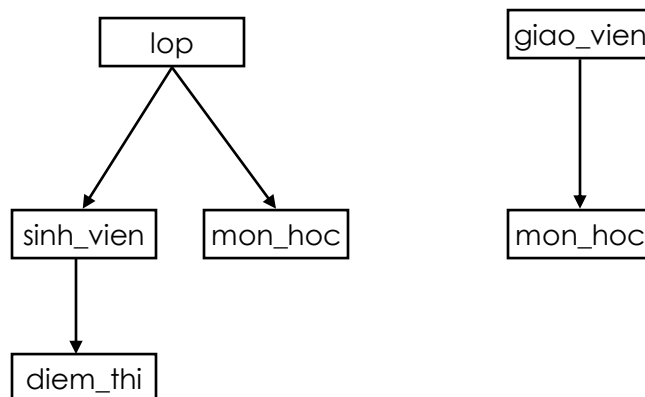
Mô hình dữ liệu phân cấp (*Hierarchical data model*)

- Ra đời những năm 60-65
- Biểu diễn bằng cây
 - Quan hệ cha-con
 - Mỗi nút có 1 cha duy nhất
 - 1 CSDL = 1 tập các cây = 1 rừng
- Các khái niệm cơ bản
 - Bản ghi
 - Móc nối
 - Các phép toán: GET, GET UNIQUE, GET NEXT, GET NEXT WITHIN PARENT,...

9

9

Mô hình dữ liệu phân cấp – Ví dụ



10

10

Mô hình dữ liệu phân cấp

- Ưu điểm
 - Dễ xây dựng và thao tác
 - Tương thích với các lĩnh vực tổ chức phân cấp
 - Ngôn ngữ thao tác đơn giản: duyệt cây.
- Nhược điểm:
 - Sự lặp lại của các kiểu bản ghi → dữ liệu dư thừa và không nhất quán.
 - Giải pháp: bản ghi ảo
 - Hạn chế trong biểu diễn ngữ nghĩa của các mối nối giữa các bản ghi (chỉ cho phép quan hệ 1-n)

11

11

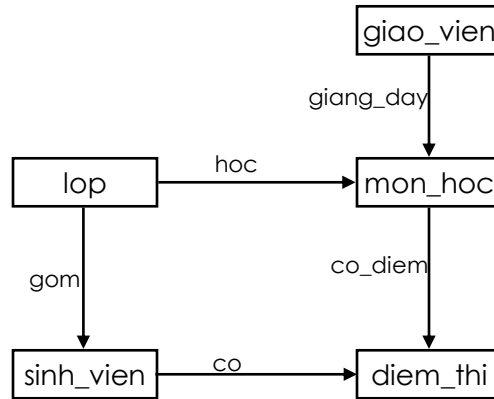
Mô hình dữ liệu mạng (*Network data model*)

- Sự ra đời
 - Sử dụng phổ biến từ những năm 60, được định nghĩa lại vào năm 71
- Biểu diễn bằng đồ thị có hướng
- Các khái niệm cơ bản
 - Tập bản ghi (record)
 - Kiểu bản ghi (record type)
 - Các trường (field)
 - Móc nối
 - Tên của móc nối
 - Chủ (owner) – thành viên (member): theo hướng của móc nối
 - Kiểu móc nối: 1-1, 1-n, đệ quy
 - Các phép toán
 - Duyệt: FIND, FIND member, FIND owner, FIND NEXT
 - Thủ tục: GET

12

12

Mô hình dữ liệu mạng – Ví dụ



13

13

Mô hình dữ liệu mạng

- Ưu điểm
 - Đơn giản
 - Có thể biểu diễn các ngữ nghĩa đa dạng với kiểu bản ghi và kiểu móc nối
 - Truy vấn thông qua phép duyệt đồ thị (navigation)
- Nhược điểm:
 - Số lượng các con trỏ lớn
 - Hạn chế trong biểu diễn ngữ nghĩa của các móc nối giữa các bản ghi

14

14

Mô hình dữ liệu quan hệ

- Sự ra đời: vào năm 1970[Codd, 1970]
- Dữ liệu được biểu diễn dưới dạng bảng
- Là mô hình dữ liệu khái niệm phổ biến cho đến tận thời điểm hiện tại
- Dựa trên lý thuyết toán học, đồng thời cũng gần với cấu trúc tệp và cấu trúc dữ liệu nên có hai loại thuật ngữ liên quan:
 - Thuật ngữ toán học: quan hệ, bộ, thuộc tính
 - Thuật ngữ hướng dữ liệu: bảng, bản ghi, trường

15

15

Ví dụ
mô hình
dữ liệu
quan hệ

MON_HOC

maMH	tenmon	soHT
CNTT01	Nhập môn CSDL	4
CNTT02	Truyền DL và mạng	4
CNTT03	Phân tích và thiết kế hệ thống	4
HTTT01	Quản lý dự án	3

LOP

malop	lop	khoa	GVCN	loptruong
IT4	Tin 4	CNTT	Ng. V. Anh	Trần T. Bình
IT5	Tin 5	CNTT	Lê A. Văn	Ng. Đ. Trung
IT6	Tin 6	CNTT	Ng. T. Thảo	Trần M. Quế
IT7	Tin 7	CNTT	Ng. V. Quý	Ng. T. Phương

SINH VIEN

maSV	tenSV	ngaysinh	gt	diachi	malop
SV0011	Trần T. Bình	1/4/1981	0	21 T. Q. B	IT4
SV0025	Ng. Đ. Trung	3/2/1980	1	56 Đ. C. V	IT5
SV0067	Trần M. Quế	26/3/1982	0	45 H. B. T	IT6
SV0034	Ng. T. Phương	29/2/1980	0	86 L. T. N	IT7

16

Mô hình dữ liệu quan hệ

- **Thuộc tính** (~trường): là các đặc tính của một đối tượng
- Mỗi thuộc tính được xác định trên một miền giá trị nhất định gọi là **miền thuộc tính**
- Ví dụ:
 - Sinhviên (MãSV, TênSV, Nămsinh, GiớiTính, ĐịaChỉ)
 - $\text{dom}(\text{MãSV}) = \{\text{char}(5)\}$
 - $\text{dom}(\text{TênSV}) = \{\text{char}(30)\}$
 - $\text{dom}(\text{Nămsinh}) = \{\text{date}\}$
 - $\text{dom}(\text{GiớiTính}) = \{0, 1\}$
 - $\text{dom}(\text{ĐịaChỉ}) = \{\text{char}(50)\}$

17

17

Mô hình dữ liệu quan hệ

- **Quan hệ** (~bảng): Cho n miền giá trị D_1, D_2, \dots, D_n không nhất thiết phân biệt, r là một quan hệ trên n miền giá trị đó nếu r là một tập các n -bộ (d_1, d_2, \dots, d_n) sao cho $d_i \in D_i$
- Một **quan hệ** có thể được biểu diễn dưới dạng 1 **bảng** trong đó 1 **dòng** trong bảng tương đương với 1 **bộ**, một **cột** trong bảng tương đương với 1 **thuộc tính** của quan hệ
- **Bậc** của 1 quan hệ là số các thuộc tính trong quan hệ
- **Lực lượng** của 1 quan hệ là số các bộ trong quan hệ

18

18

Mô hình dữ liệu quan hệ

- **Định nghĩa:** K là **khoá tối thiểu** của r nếu K là một khoá của r và bất kỳ tập con thực sự nào của K đều không phải là khoá của r
- **Định nghĩa:** Một tập con $K \subseteq U$ được gọi là **khoá ngoài** của quan hệ r(U) tham chiếu đến một quan hệ r' nếu K là khoá chính của r'

19

19

Mô hình dữ liệu quan hệ - Ví dụ:

- Quan hệ: SinhViên(MãSV, TênSV, NămSinh, GiớiTính, TênLớp)

SV001	Nguyễn Văn An	1982	1	Tin 7
SV002	Nguyễn Văn An	1985	1	HTTT
SV003	Lê Văn Cường	1981	1	HTTT
SV004	Nguyễn Thùy Linh	1981	0	BK65

- Siêu khoá: {MãSV, HọTên};
- Khoá tối thiểu: {MãSV}; {HọTên, NămSinh}
- Khoá ngoài: TênLớp nếu coi nó là khoá chính của quan hệ Lớp

20

20

Mô hình dữ liệu quan hệ -nhận xét

- Ưu điểm
 - Dựa trên lý thuyết tập hợp
 - Khả năng tối ưu hoá các xử lý phong phú
- Nhược điểm
 - Hạn chế trong biểu diễn ngữ nghĩa
 - Cấu trúc dữ liệu không linh hoạt

21

21

Mô hình thực thể liên kết (*Entity-Relationship data model*)

- Cho phép mô tả các dữ liệu có liên quan trong một xí nghiệp trong thế giới thực dưới dạng các đối tượng và các mối quan hệ của chúng.
- Được sử dụng cho bước đầu thiết kế CSDL, làm nền tảng để ánh xạ sang một mô hình khái niệm nào đó mà Hệ quản trị CSDL sẽ sử dụng
- Trong mô hình thực thể liên kết, CSDL được mô hình hóa như là:
 - Một tập hợp các thực thể
 - Liên hệ giữa các thực thể này

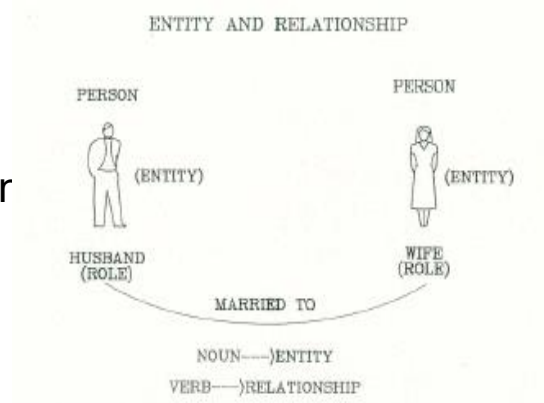
22

22

Mô hình thực thể liên kết

Các khái niệm cơ bản

- Thực thể, tập thực thể
- Thuộc tính
- Khoá
- Liên kết, tập liên kết



23

23

Mô hình thực thể liên kết...

- **Thực thể**: một đối tượng trong thế giới thực, tồn tại độc lập và phân biệt được với các đối tượng khác
- **Tập thực thể**: một tập hợp các thực thể có tính chất giống nhau
- Ví dụ:
 - Thực thể: một sinh viên, một lớp
 - Tập thực thể: toàn thể sinh viên của 1 lớp, toàn thể các lớp của 1 khoa

24

24

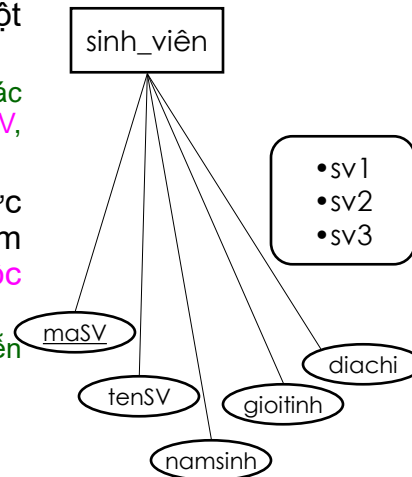
Mô hình thực thể liên kết...

- **Thuộc tính** là đặc tính của một tập thực thể

- Tập thực thể **SinhViên** có các thuộc tính như: **TênSV**, **NămSinh**,...

- Mỗi thực thể trong tập thực thể có một giá trị đặc tính nằm trong **miền giá trị của thuộc tính**

- Sinh viên 1 có: Họ tên là Nguyễn Hải Anh, Năm sinh 1980



25

25

Mô hình thực thể liên kết...

Khóa

- Một hay một tập thuộc tính mà giá trị của chúng có thể xác định duy nhất một thực thể trong tập thực thể
 - Tập thực thể SinhViên có thể dùng MãSV làm khóa
- Trong sơ đồ ER, thuộc tính nào được chọn làm khóa sẽ được **gạch chân**

26

26

Mô hình thực thể liên kết...

Liên kết - Tập liên kết

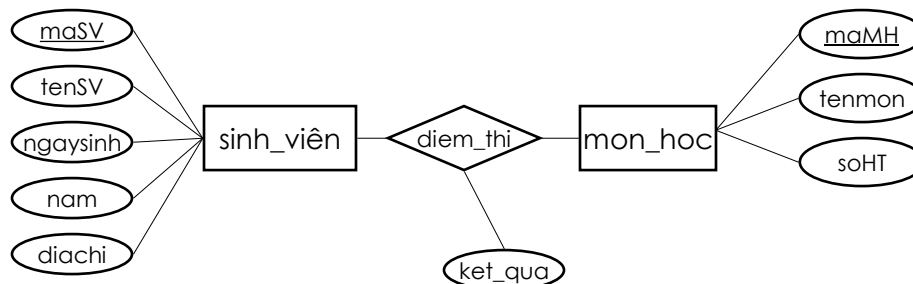
- Một **liên kết** là một **mối liên hệ có nghĩa** giữa nhiều thực thể
 - Cho một thực thể SinhViên1 và LớpA, liên kết ThànhViên chỉ ra rằng SinhViên1 là 1 thành viên của LớpA
- **Tập liên kết** là một tập hợp các liên kết cùng kiểu
 - Giữa tập thực thể SinhViên và Lớp có 1 tập liên kết ThànhViên, chỉ ra rằng mỗi sinh viên đều là thành viên của 1 lớp nào đó
- Một liên kết có thể có thuộc tính

27

27

Mô hình thực thể liên kết...

Liên kết - Tập liên kết - Ví dụ:



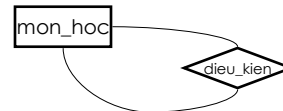
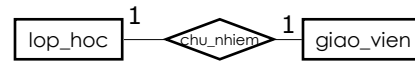
28

28

Mô hình thực thể liên kết...

Ràng buộc của kết nối

- **1-1**: Liên kết 1 thực thể của một tập thực thể với nhiều nhất 1 thực thể của tập thực thể khác
- **1-n**: Liên kết 1 thực thể của một tập thực thể với nhiều thực thể của tập thực thể khác
- **n-n**: Liên kết 1 thực thể của một tập thực thể với nhiều thực thể của tập thực thể khác và ngược lại
- **đệ quy**: Liên kết giữa các thực thể cùng kiểu



29

29

Mô hình dữ liệu hướng đối tượng (Object-oriented data model)

- Sự ra đời
 - Khoảng đầu những năm 90
- Biểu diễn: sơ đồ lớp
- Các khái niệm cơ bản
 - **Đối tượng**: một đối tượng trong thế giới thực, được xác định bởi một định danh duy nhất
 - **Thuộc tính**: biểu diễn một đặc tính của đối tượng,
 - **Phương thức**: thao tác được thực hiện trên đối tượng.
 - Tất cả các truy nhập vào thuộc tính của đối tượng đều phải được thực hiện thông qua các phương thức này.
 - **Lớp**: một cách thức để khai báo một tập các đối tượng có chung một tập thuộc tính và phương thức

30

30

Ví dụ:

```
class sinh_vien {
    string maSV;
    string tenSV;
    date ngaysinh;
    boolean nam;
    string diachi;
    string lop;

    string ten();
    string ngay_sinh();
    string dia_chi();
    string lop();
    void gan_DC(string DC_moi);
    void gan_lop(string lop);
}
```

31

31

Nhận xét:

- Ưu điểm
 - Cho phép định nghĩa kiểu đối tượng phức tạp
 - Tính chất: bao đóng (*encapsulation*), kế thừa (*heritage*), đa hình (*polymorphism*)
- Nhược điểm
 - Cấu trúc lưu trữ phức tạp và có thể sử dụng nhiều con trỏ
 - Khả năng tối ưu hoá các xử lý bị hạn chế trong nhiều trường hợp

32

32

So sánh và đánh giá

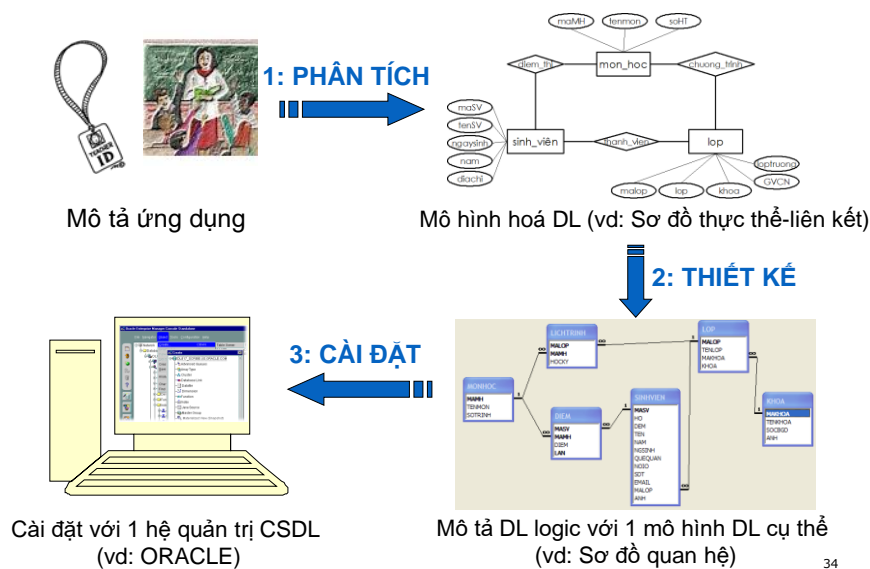
Nhắc lại: Mô hình dữ liệu là một tập hợp các khái niệm dùng để mô tả cấu trúc của một CSDL

	Mô hình mạng	Mô hình phân cấp	Mô hình quan hệ	Mô hình TT-LK	Mô hình HDT
biểu diễn ngữ nghĩa DL	hạn chế	hạn chế	tương đối đa dạng	đa dạng	đa dạng
lưu trữ DL	s/d nhiều con trỏ	dữ liệu lặp lại	dễ dàng và hiệu quả	khó lưu trữ	cấu trúc phức tạp
khả năng truy vấn	đơn giản	đơn giản	đa dạng		đa dạng
hiệu quả của truy vấn	ít khả năng tối ưu	ít khả năng tối ưu	tối ưu hoá tốt	không được xem xét (không hiệu quả)	không h/q khi s/d nhiều con trỏ

33

33

Các bước xây dựng một hệ CSDL



34

34