

CƠ SỞ DỮ LIỆU

CHƯƠNG 2
MÔ HÌNH DỮ LIỆU

Nội dung

1. Tổng quan về mô hình dữ liệu
2. Mô hình phân cấp
3. Mô hình mạng
4. Mô hình hướng đối tượng
5. Mô hình thực thể kết hợp
6. Mô hình quan hệ
7. So sánh và đánh giá

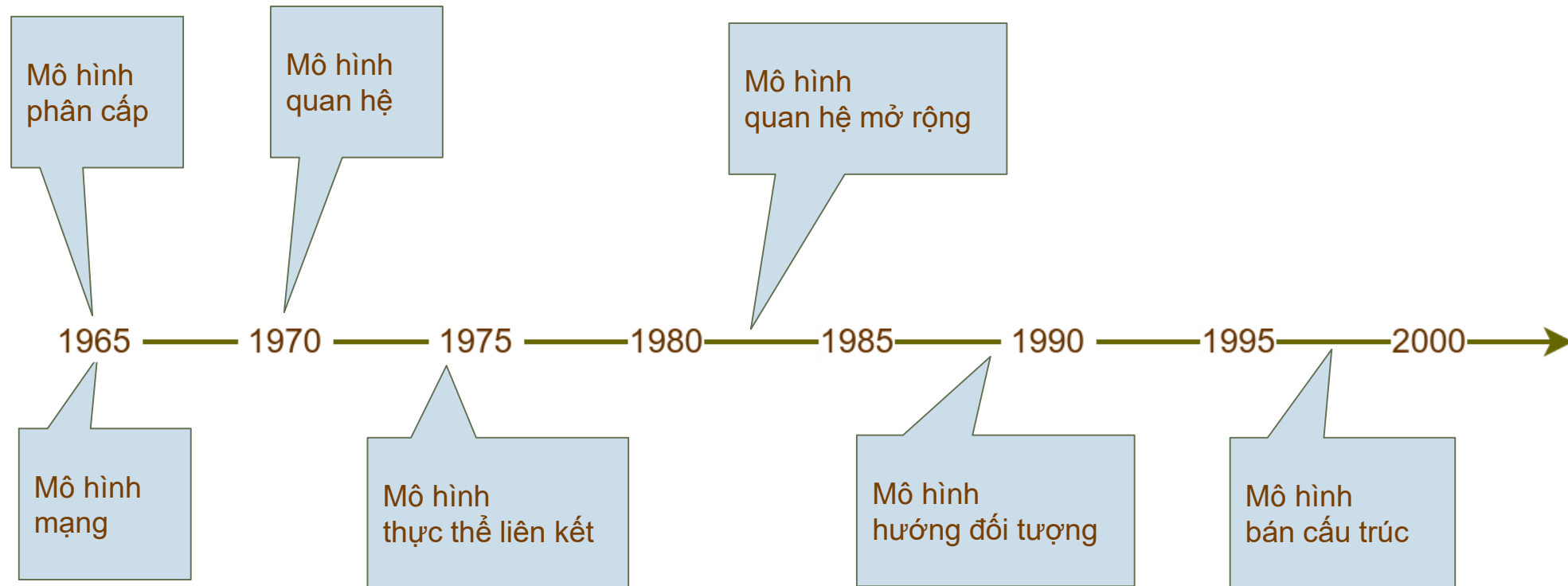
Tổng quan về mô hình dữ liệu

- 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

Tổng quan về mô hình dữ liệu

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

Tổng quan về mô hình dữ liệu



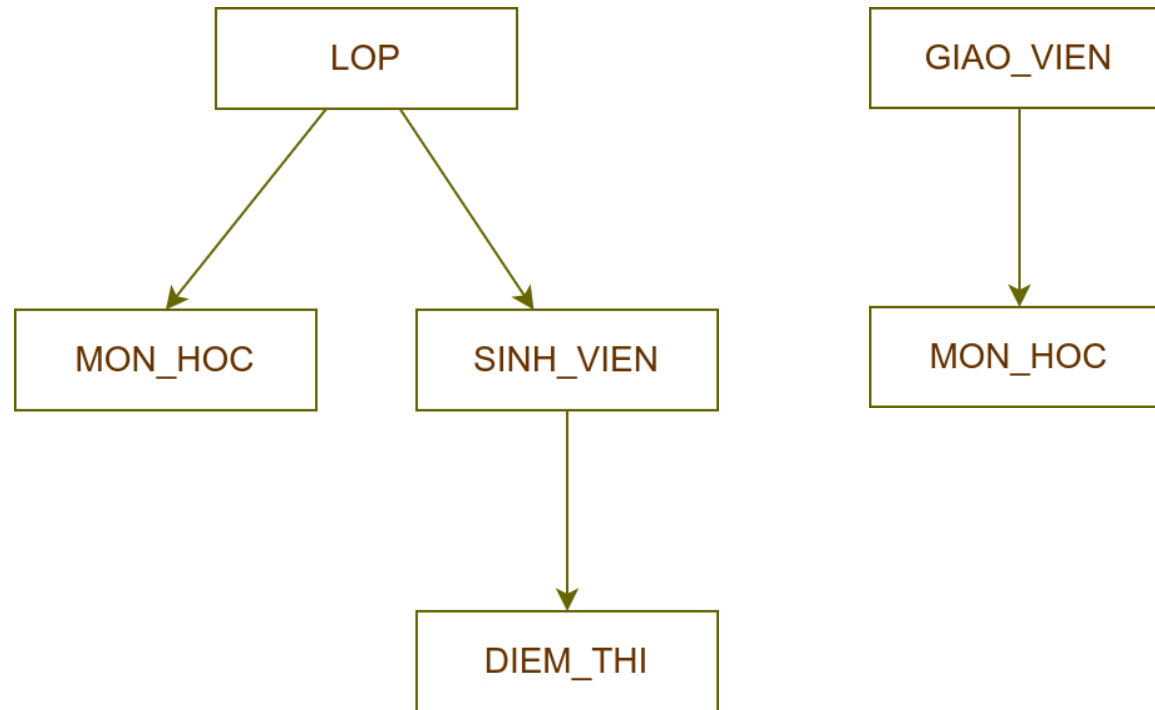
Đặt vấn đề

- 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

Mô hình cơ sở dữ liệu phân cấp

- Ra đời những năm 60-65
- Biểu diễn theo mô hình 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: chứa dữ liệu
 - Móc nối
 - Các phép toán: GET, GET UNIQUE, GET NEXT, GET NEXT WITHIN PARENT,...

Mô hình cơ sở dữ liệu phân cấp



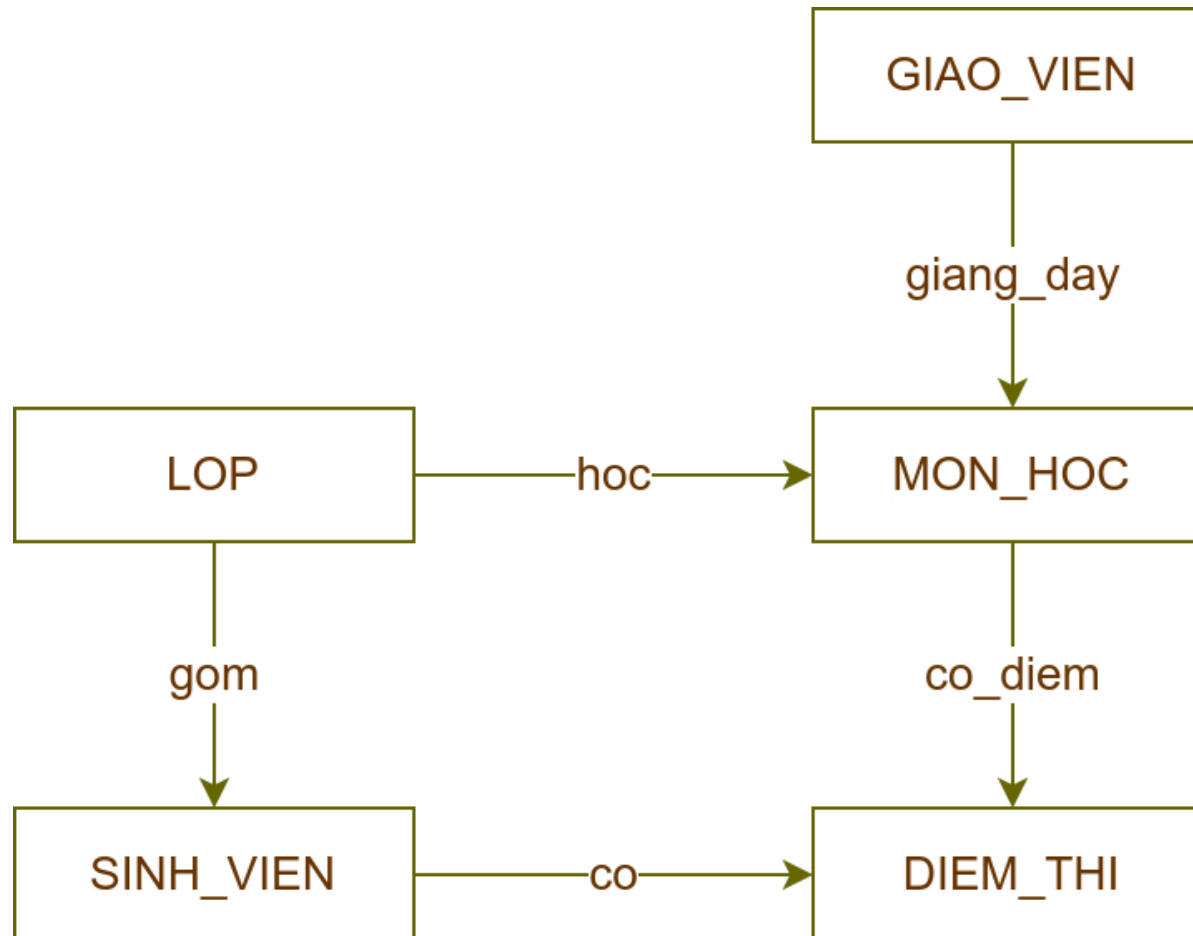
Mô hình dữ liệu phân cấp

- 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.
 - Hạn chế trong biểu diễn ngữ nghĩa của các mối nối giữa các bản ghi (chỉ cho phép quan hệ 1-n)

Mô hình dữ liệu mạng

- 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

Mô hình dữ liệu mạng



Mô hình dữ liệu mạng

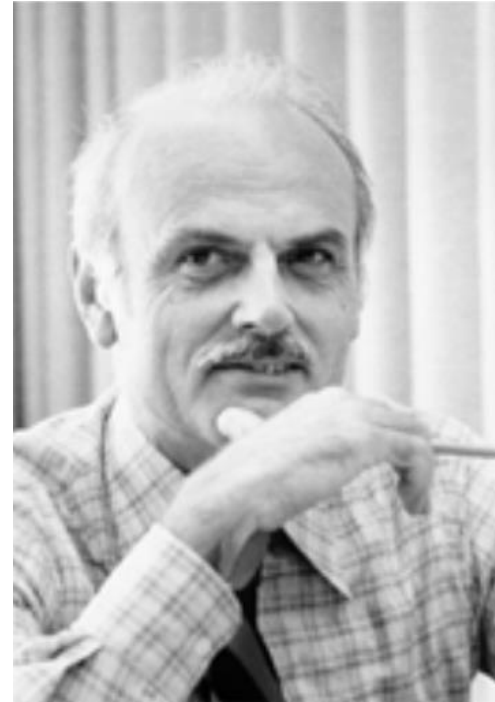
- Ưu điểm
 - Đơn giản
 - Có thể biểu diễn các ngữ nghĩa đa dạng với kiểu bản ghi và kiểu móc nối
 - Truy vấn thông qua phép duyệt đồ thị
- Nhược điểm
 - Số lượng các con trỏ lớn.
 - Hạn chế trong biểu diễn ngữ nghĩa của các móc nối giữa các bản ghi

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

Edgar Frank "Ted" Codd

- (19/08/1923 – 18/04/2003)
- Nhà khoa học máy tính người Anh, làm việc cho IBM.
- Cha đẻ của mô hình dữ liệu quan hệ
- Nhận giải thưởng Turing, năm 1981



12 Quy tắc của Codd

- Quy tắc 0: Quy tắc nền tảng
- Quy tắc 1: Quy tắc thông tin
- Quy tắc 2: Quy tắc truy cập được đảm bảo
- Quy tắc 3: Xử lý có hệ thống các giá trị rỗng
- Quy tắc 4: Mục lục trực tuyến hoạt động / động dựa trên mô hình quan hệ
- Quy tắc 5: Quy tắc ngôn ngữ phụ dữ liệu toàn diện
- Quy tắc 6: Xem quy tắc cập nhật

12 Quy tắc của Codd

- Quy tắc 7: Hoạt động ở mức quan hệ
 - Chèn mức cao, cập nhật và xóa
- Quy tắc 8: Độc lập dữ liệu vật lý
- Quy tắc 9: Độc lập dữ liệu logic
- Quy tắc 10: Độc lập về tính chính trực
- Quy tắc 11: Độc lập về phân phối
- Quy tắc 12: Quy tắc không lật đổ

Mô hình dữ liệu quan hệ

LOP

MaLop	TenLop	Khoa	GVCN	LopTruong
VT1	Viễn Thông 1	VT	Nguyễn Văn A	Trần A
VT2	Viễn Thông 2	VT	Nguyễn Văn A	Trần B
VT3	Viễn Thông 3	VT	Nguyễn Văn A	Trần C
VT4	Viễn Thông 4	VT	Nguyễn Văn A	Nguyễn D

MON_HOC

MaMH	TenMH	SoTC
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

SINH_VIEN

MaSV	TenSV	NgaySinh	GioiTinh	DiaChi	MaLop
SV0001	Trần T. Bình	1/4/2002	Nam	Hà Nội	VT1
SV0002	Ng. Đ. Trung	23/2/2002	Nam	Hà Nội	VT1
SV0003	Trần M. Quế	12/9/2002	Nữ	Thái Nguyên	VT3
SV0004	Ng. T. Phương	7/5/2002	Nữ	Bắc Ninh	VT4

Mô hình dữ liệu quan hệ

- Các khái niệm cơ bản
 - Thuộc tính, miền thuộc tính
 - Quan hệ
 - Khóa

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)
 - $\text{dom}(\text{MãSV}) = \{\text{char}(5)\}$
 - $\text{dom}(\text{TênSV}) = \{\text{char}(30)\}$
 - $\text{dom}(\text{Nămsinh}) = \{\text{date}\}$
 - $\text{dom}(\text{GiớiTính}) = \{0, 1\}$
 - $\text{dom}(\text{ĐịaChỉ}) = \{\text{char}(50)\}$

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

Mô hình dữ liệu quan hệ

- Định nghĩa (tiếp): Cho $U = \{A_1, A_2, \dots, A_n\}$ là một tập hữu hạn các thuộc tính trong đó $\text{dom}(A_i) = D_i$, r là quan hệ trên tập thuộc tính U ký hiệu là $r(U)$ nếu:
 - $r \subseteq D_1 \times D_2 \times \dots \times D_n$
- U được gọi là sơ đồ quan hệ (lược đồ quan hệ)

Mô hình dữ liệu quan hệ

- Định nghĩa Khóa của quan hệ r trên tập thuộc tính $U = \{A_1, A_2, \dots, A_n\}$ là một tập $K \subseteq U$ sao cho với bất kỳ 2 bộ t_1, t_2 thuộc r đều tồn tại một thuộc tính A thuộc K mà $t_1[A] \neq t_2[A]$
- Một quan hệ có thể có nhiều khóa
- Nếu K là khóa của r thì mọi K' sao cho $K \subseteq K'$ đều là khoá của r .
 K' được gọi là siêu khoá của r

Mô hình dữ liệu quan hệ

- Định nghĩa: K là khóa 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 khóa chính của r'

Mô hình dữ liệu quan hệ

- Quan hệ: SinhViên, Lớp
- Siêu khóa: {MãSV, HọTên};
- Khóa tối thiểu: {MãSV}; {HọTên, NămSinh}
- Khóa ngoài: MãLớp nếu coi nó là khóa chính của quan hệ Lớp

SINH_VIEN

MaSV	TenSV	NgaySinh	GioiTinh	DiaChi	MaLop
SV0001	Trần T. Bình	1/4/2002	Nam	Hà Nội	VT1
SV0002	Ng. Đ. Trung	23/2/2002	Nam	Hà Nội	VT1
SV0003	Trần M. Quế	12/9/2002	Nữ	Thái Nguyên	VT3
SV0004	Ng. T. Phương	7/5/2002	Nữ	Bắc Ninh	VT4

LOP

MaLop	TenLop	Khoa	GVCN	LopTruong
VT1	Viễn Thông 1	VT	Nguyễn Văn A	Trần A
VT2	Viễn Thông 2	VT	Nguyễn Văn A	Trần B
VT3	Viễn Thông 3	VT	Nguyễn Văn A	Trần C
VT4	Viễn Thông 4	VT	Nguyễn Văn A	Nguyễn D

Mô hình dữ liệu quan hệ

- 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

Mô hình thực thể liên kết

- 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

Mô hình thực thể liên kết (ERD)

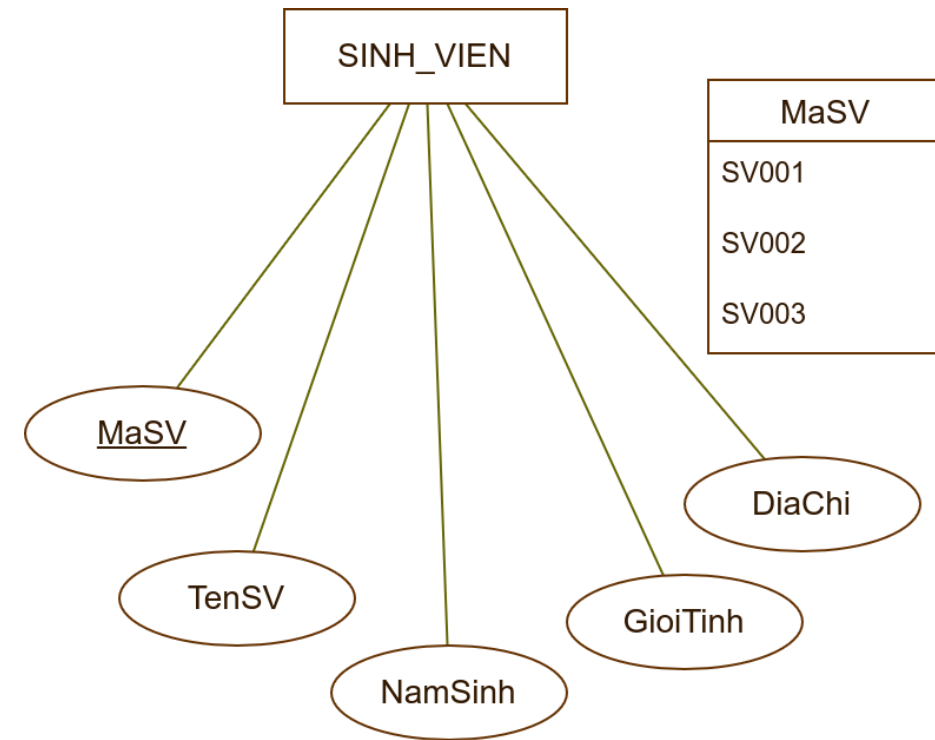
- Thực thể, tập thực thể
- Thuộc tính
- Khoá
- Liên kết, tập liên kết

Mô hình thực thể liên kết (ERD)

- 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

Mô hình thực thể liên kết (ERD)

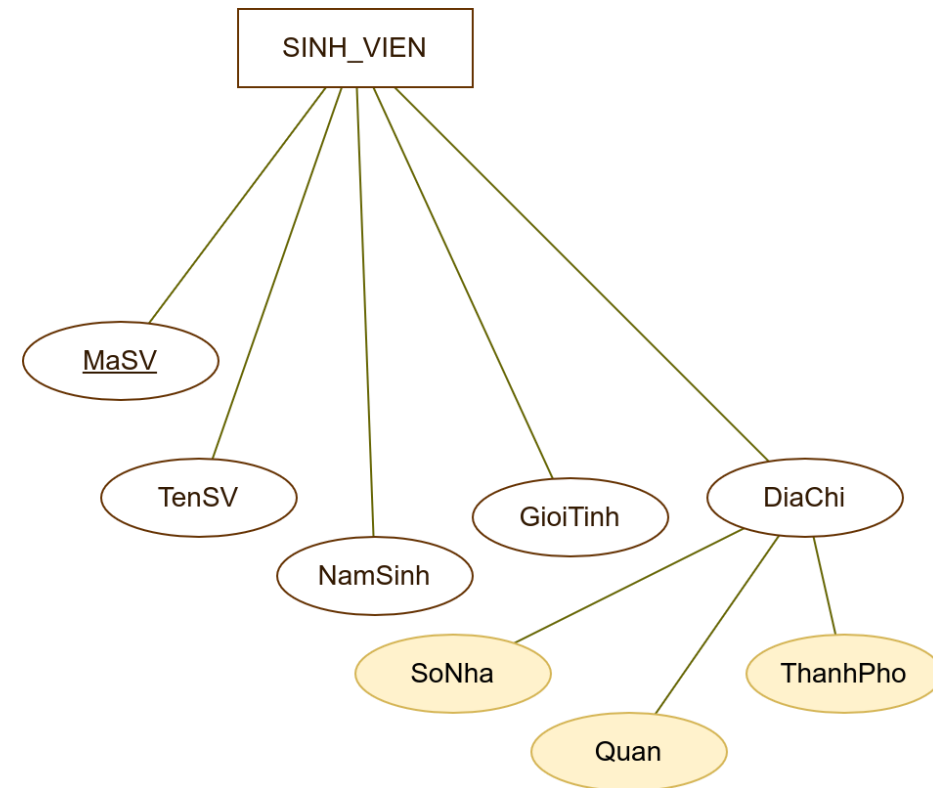
- Thuộc tính là đặc tính của một tập thực thể
 - Tập thực thể SinhViên có các thuộc tính như: TênSV, NămSinh,...
- Mỗi thực thể trong tập thực thể có một giá trị đặc tính nằm trong miền giá trị của thuộc tính
 - Sinh viên 1 có: Họ tên là Nguyễn Hải Anh, Năm sinh 1980



Mô hình thực thể liên kết

Kiểu thuộc tính

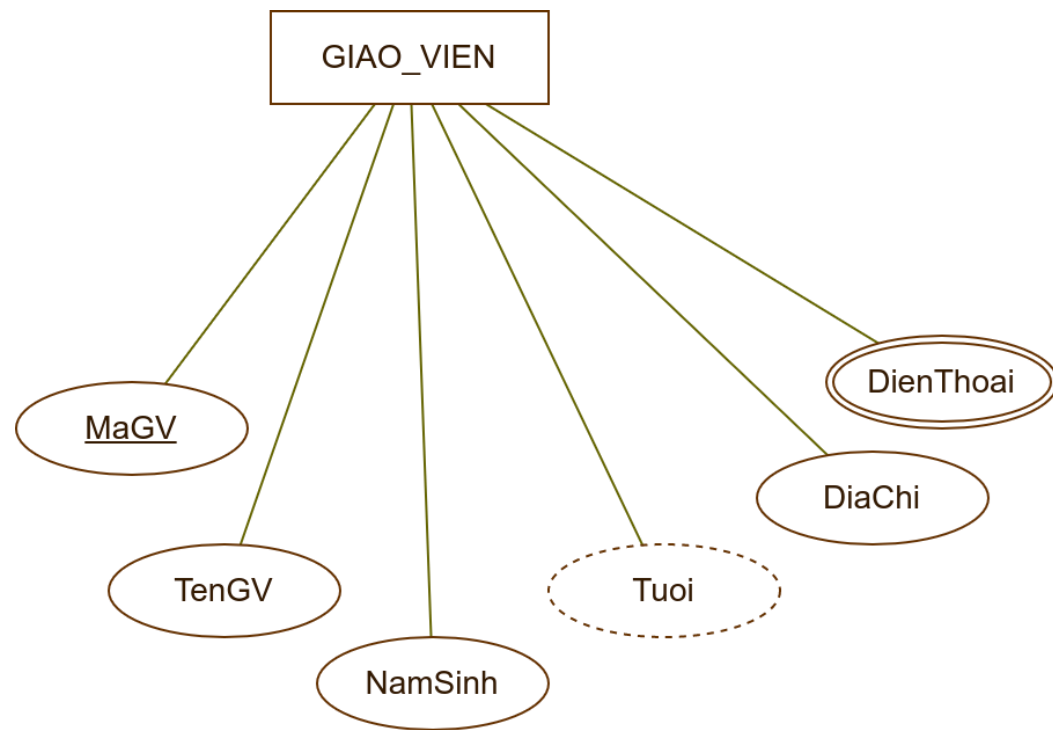
- Thuộc tính đơn giản (thuộc tính nguyên tố)
 - có kiểu dữ liệu nguyên tố
- Thuộc tính phức
 - có kiểu phức, định nghĩa bởi các thuộc tính khác



Mô hình thực thể liên kết

Kiểu thuộc tính

- Thuộc tính đa giá trị
 - tương ứng với mỗi thực thể, có thể nhận nhiều giá trị
- Thuộc tính suy diễn
 - có thể tính toán được từ (các) thuộc tính khác



Mô hình thực thể liên kết

Khoá

- 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
- Khóa gồm nhiều thuộc tính thì gọi là khóa phức
- Một tập thực thể có thể có nhiều khóa nhưng chỉ một trong số các khóa được chọn làm khóa chính
- Trong sơ đồ ER, thuộc tính nào được chọn làm khóa chính sẽ được gạch chân

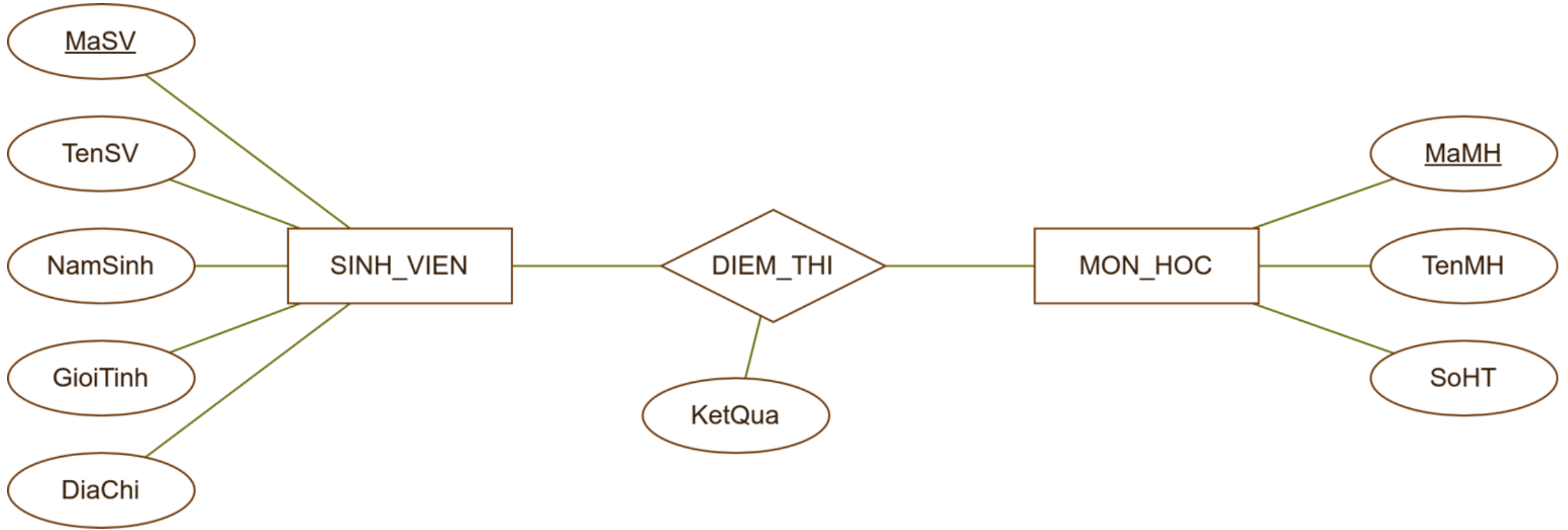
Mô hình thực thể liên kết

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

Mô hình thực thể liên kết

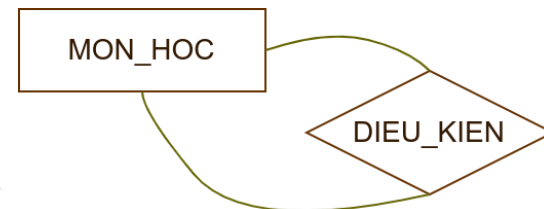
Liên kết - Tập liên kết



Mô hình thực thể liên kết

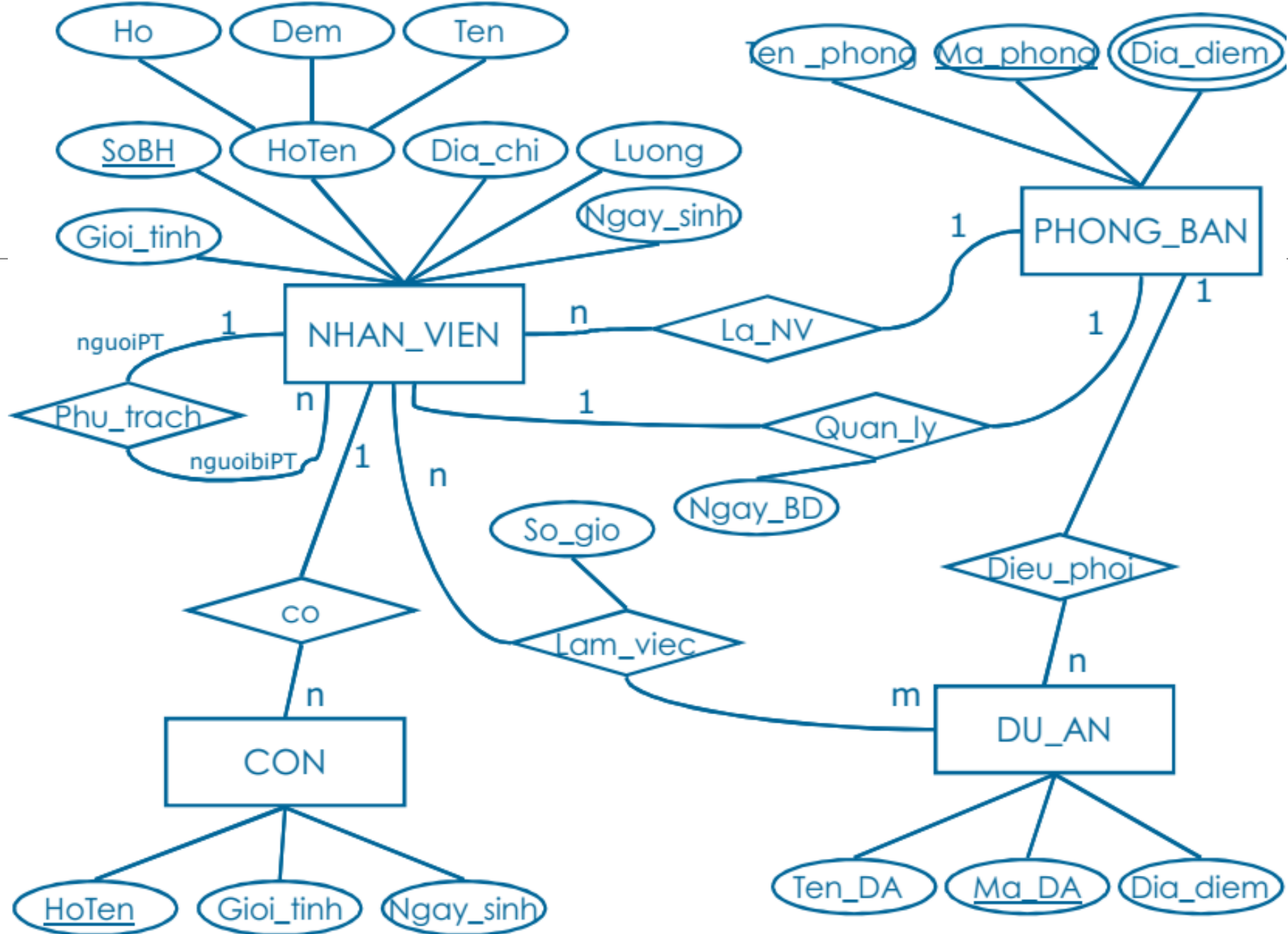
Ràng buộc của kết nối

- 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



Bài tập: Vẽ sơ đồ ER

- Bài toán: phân tích và thiết kế 1 CSDL gồm các thông tin trong 1 công ty (nhân viên, phòng ban, dự án)
 - Công ty được tổ chức bởi các phòng ban. Mỗi phòng ban có 1 tên duy nhất, 1 số duy nhất và 1 người quản lý (thời điểm bắt đầu công tác quản lý của người này cũng được lưu lại trong CSDL). Mỗi phòng ban có thể có nhiều trụ sở làm việc khác nhau
 - Mỗi phòng điều phối một số dự án. Mỗi dự án có 1 tên và 1 mã số duy nhất, thực hiện tại một địa điểm duy nhất
 - Các thông tin về nhân viên cần được quan tâm gồm: tên, số bảo hiểm, địa chỉ, lương, giới tính, ngày sinh. Mỗi nhân viên làm việc tại một phòng ban nhưng có thể tham gia nhiều dự án khác nhau. Những dự án này có thể được điều phối bởi các phòng ban khác nhau. Thông tin về số giờ làm việc trong từng dự án (theo tuần) cũng như người quản lý trực tiếp của các nhân viên cũng được lưu trữ
 - Thông tin về con cái của từng nhân viên: tên, giới tính, ngày sinh



Các công cụ online giúp vẽ ERD

- Lucidchart
 - <https://www.lucidchart.com>
- Drawio
 - <https://www.draw.io/>
- Một số mô hình ERD tham khảo
 - http://www.databaseanswers.org/data_models/index.htm

Biến đổi sơ đồ thực thể liên kết sang sơ đồ quan hệ

- Biến đổi tập các thực thể
- Biến đổi các liên kết
- Các khoá của các sơ đồ quan hệ
- Các sơ đồ quan hệ với khoá chung

Biến đổi các tập thực thể

- Bước 1: 1 tập thực thể 1 quan hệ
 - thuộc tính thuộc tính (trường)
 - 1 thực thể 1 bộ
 - khoá của tập thực thể khóa của quan hệ

Biến đổi các tập thực thể

- Bước 2: 1 tập thực thể xác định từ tập thực thể khác (E) qua 1 liên kết 1 quan hệ chứa khoá của E:
 - LOPTRUONG(maSV)



Biến đổi các tập thực thể

- Bước 3: Liên kết 1-1
 - Thêm 1 quan hệ mới xác định bởi các thuộc tính nằm trong khóa của các thực thể có liên quan
 - CHU_NHIEM_LOP(malop,maGV)
 - Dùng khóa ngoài
 - LOP_HOC(malop,lop,khoa,maGV)

Biến đổi các tập thực thể

- Bước 5: Liên kết n-n
 - Thêm 1 quan hệ mới xác định bởi các thuộc tính nằm trong khóa của các thực thể có liên quan và các thuộc tính của liên kết
 - DANG_KY(maSV, maMH, diem)

Thuộc tính đa trị

- Bước 6: Với mỗi thuộc tính đa trị
 - Thêm 1 quan hệ mới xác định bởi thuộc tính đa trị và khoá của tập thực thể tương ứng
 - MH_GV(maMH,giao_vien)

Mô hình dữ liệu hướng đối tượng

- 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

Mô hình dữ liệu hướng đối tượng

```
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);  
}
```

Mô hình dữ liệu hướng đối tượng

- 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

So sánh và đánh giá

	Mô hình mạng	Mô hình phân cấp	Mô hình quan hệ	Mô hình TT-LK	Mô hình HĐT
Biểu diễn ngữ nghĩa DL	hạn chế	hạn chế	tương đối đa dạng	đa dạng	đa dạng
Lưu trữ DL	sử dụng 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		sử dụng nhiều con trỏ