



TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

CÁC MÔ HÌNH DỮ LIỆU

GV: Đỗ Bá Lâm

Email: lamdb@soict.hust.edu.vn

Viện Công nghệ thông tin và Truyền thông,
Trường Đại học Bách Khoa Hà Nội

Nội dung

Tổng quan các mô hình dữ liệu

Mô hình dữ liệu thực thể - liên kết

Mô hình dữ liệu quan hệ

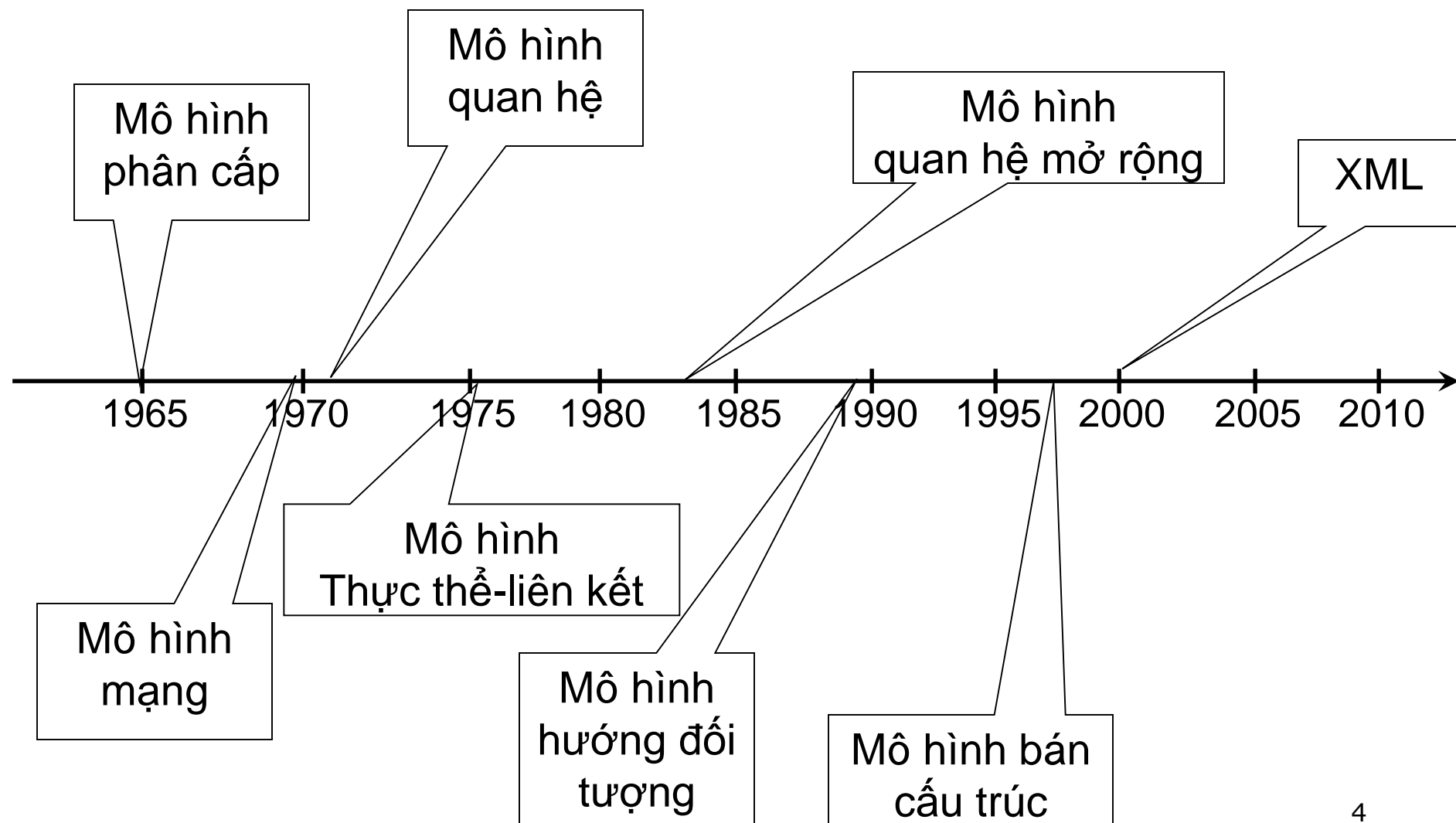
Biến đổi mô hình dữ liệu thực thể - liên kết sang mô hình dữ liệu quan hệ

Mô hình dữ liệu

*"A data model is a plan for building a database"**

- Mô hình dữ liệu là một tập hợp các khái niệm dùng để mô tả về dữ liệu. Bao gồm:
 - *Cấu trúc của dữ liệu*
 - *Các phép toán* để thao tác với dữ liệu
 - *Các ràng buộc* về dữ liệu

Lịch sử các mô hình dữ liệu

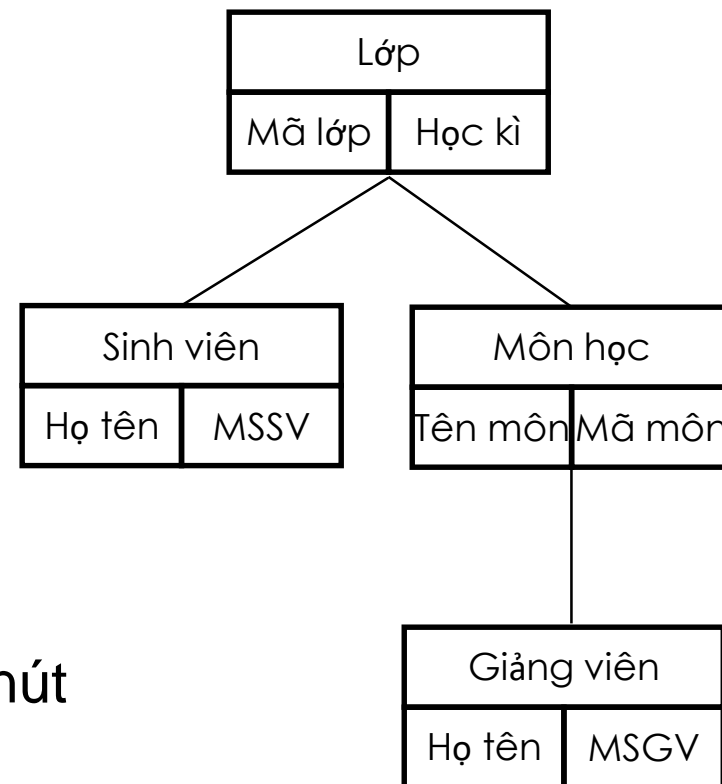


Một vài mô hình dữ liệu

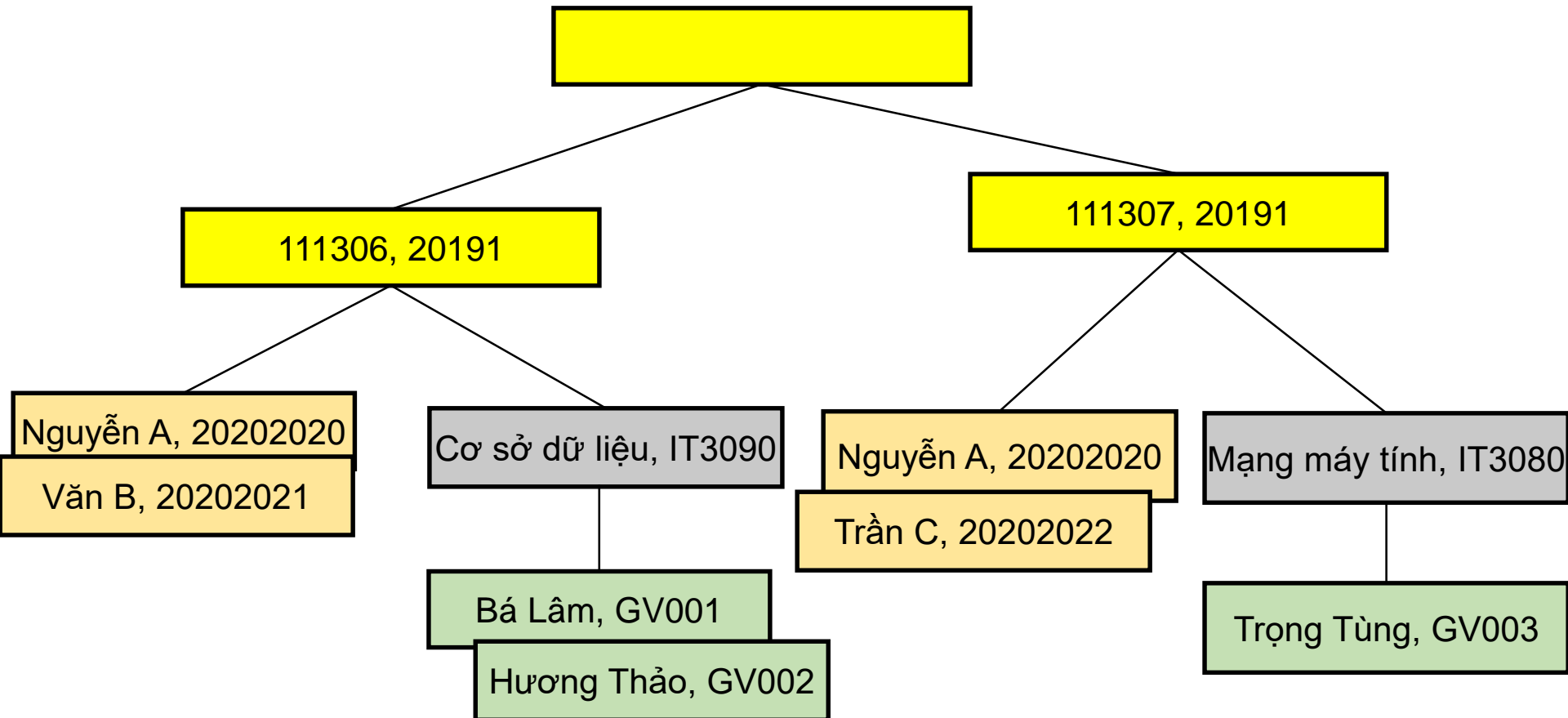
- Mô hình dữ liệu phân cấp
- Mô hình dữ liệu mạng
- Mô hình dữ liệu quan hệ
- Mô hình dữ liệu thực thể - liên kết
- Mô hình dữ liệu hướng đối tượng

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

- Ra đời khoảng năm 1960-1965
- Biểu diễn bằng cây,
 - 1 CSDL = tập các cây = rừng
- Các khái niệm cơ bản
 - Bản ghi (*record*)
 - Kiểu bản ghi (*record type*)
 - Các trường (*field*)
 - Móc nối/Liên kết (*link*):
 - giữa nút cha-nút con
 - mỗi nút con chỉ có **duy nhất một** nút cha
 - Các phép toán: GET, GET UNIQUE, GET NEXT, GET NEXT WITHIN PARENT, ...



Ví dụ một cây cho quan hệ Lớp-SV-Môn học-GV

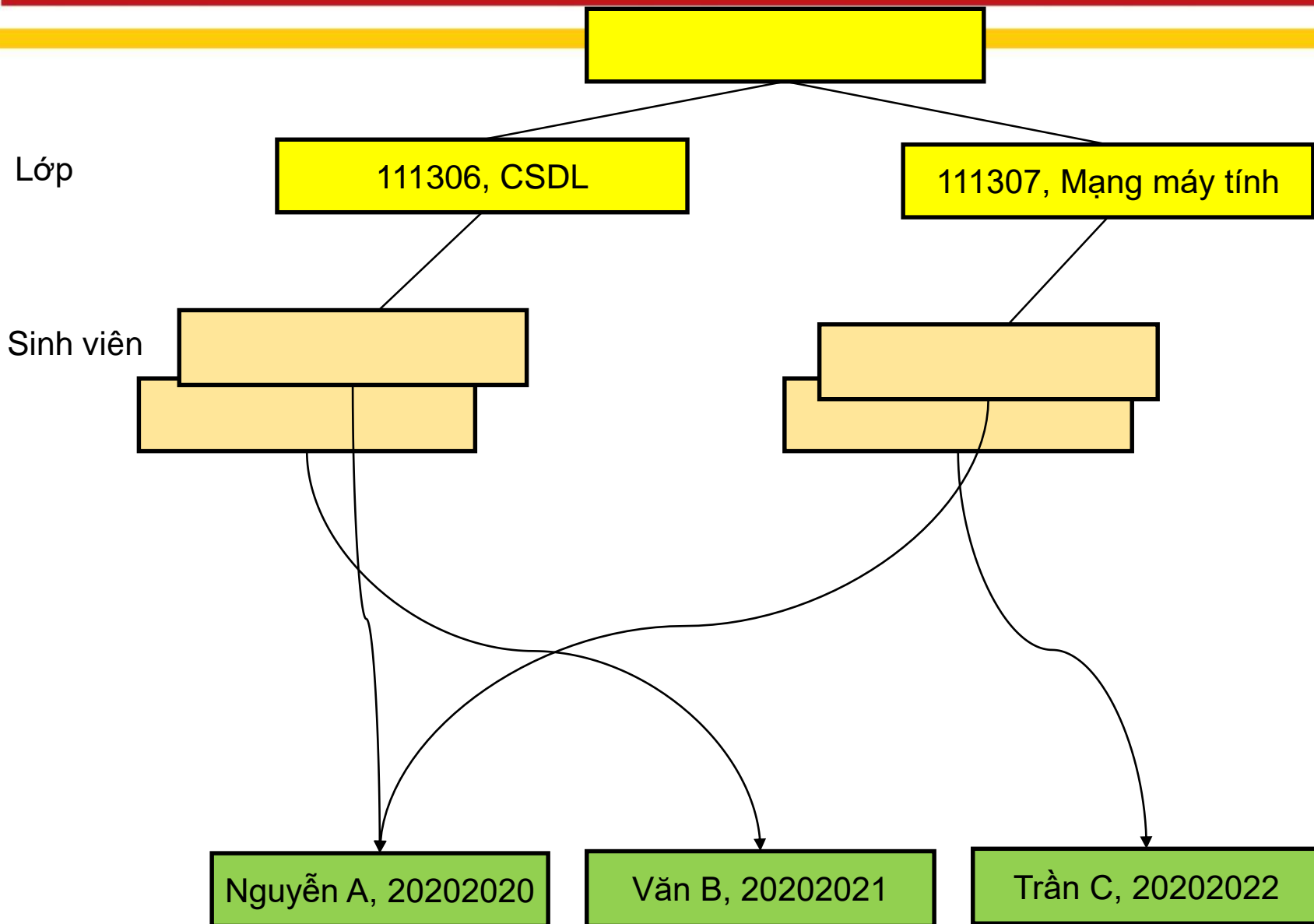


1 CSDL = tập các cây = rừng

Nhận xét

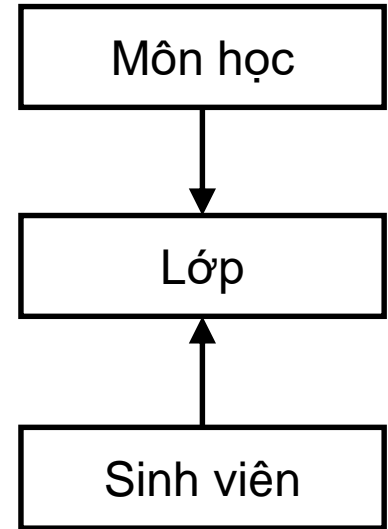
- Ưu điểm
 - Dễ xây dựng và thao tác
 - Tương thích với các lĩnh vực tổ chức phân cấp (vd: tổ chức nhân sự trong các đơn vị, ...)
 - Ngôn ngữ thao tác đơn giản (duyet cây)
- Nhược điểm
 - Lặp dữ liệu, do đó dữ liệu không nhất quán khi có sự cập nhật
 - Hạn chế trong biểu diễn ngữ nghĩa của các mối nối giữa các bản ghi (chỉ hỗ trợ quan hệ 1-n)
 - Giải pháp: sử dụng **bản ghi ảo**, con trỏ tới bản ghi thật
 - Giải quyết lặp dữ liệu
 - Hỗ trợ quan hệ nhiều – nhiều (m-n)

Bản ghi ảo

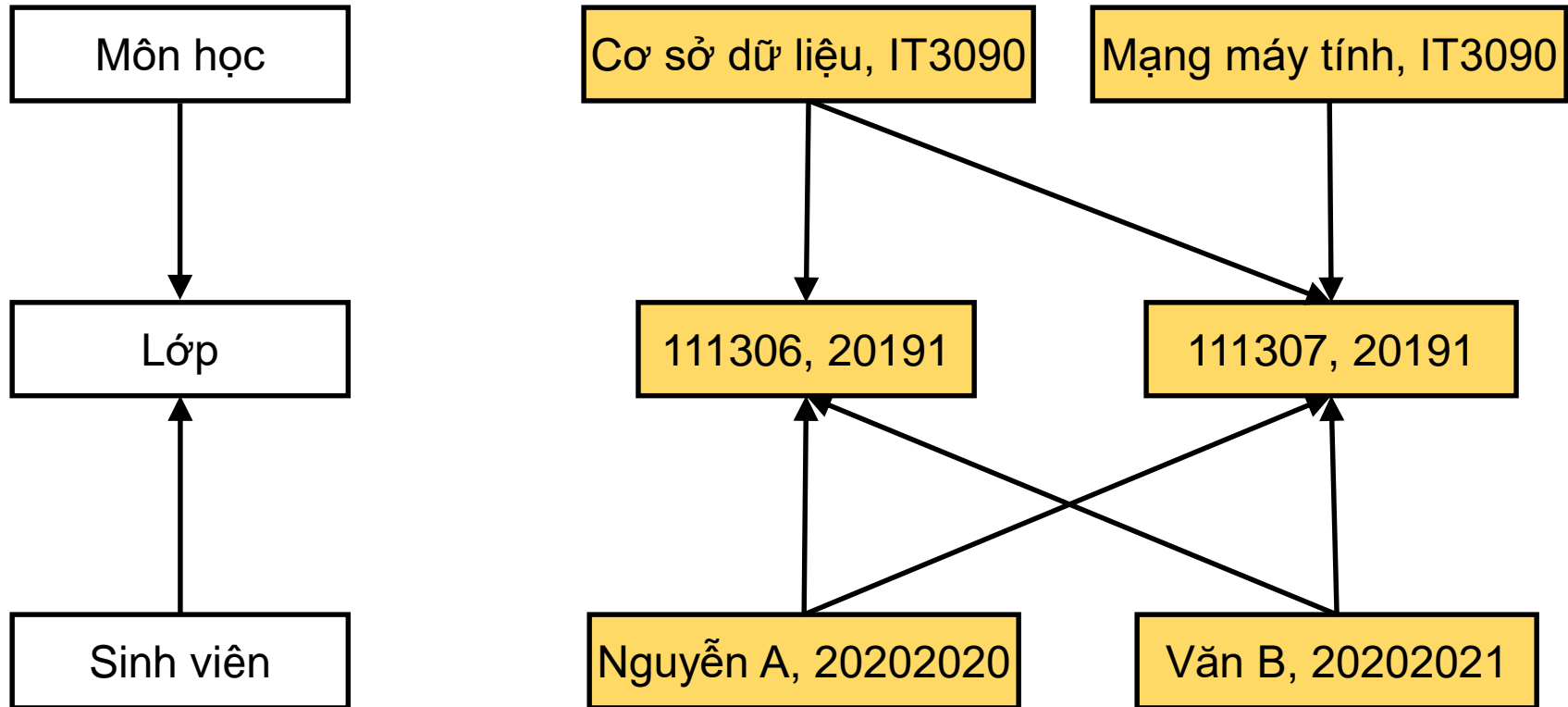


Mô hình dữ liệu mạng (*Network data model*)

- Ra đời vào năm 1969 bởi Charles Bachman
- Biểu diễn: bằng đồ thị có hướng
- Các khái niệm cơ bản
 - Bản ghi (*record*)
 - Kiểu bản ghi (*record type*)
 - Các trường (*field*)
 - Móc nối/Liên kết (*link*)
 - giữa Chủ (*owner*) – Thành viên (*member*)
 - một Thành viên (con) có thể **liên kết tới nhiều** Chủ (cha)
 - Các phép toán
 - Duyệt: FIND, FIND member, FIND owner, FIND NEXT
 - Thủ tục: GET



Ví dụ



Chủ (cha): Lớp

Thành viên (con): Môn học và Sinh viên

Nhận xét

- Ưu điểm
 - Đơn giản về mặt khái niệm, dễ dàng trong thiết kế
 - Truy cập dữ liệu nhanh chóng
 - Cho phép biểu diễn quan hệ một-nhiều (1-n), nhiều-nhiều (m-n)
- Nhược điểm
 - Cấu trúc của mô hình tương đối phức tạp vì cần nhiều con trỏ để tạo liên kết giữa các bản ghi
 - Sự thay đổi về cấu trúc CSDL khó khăn

Mô hình dữ liệu quan hệ (*Relational data model*)

- Sự ra đời: vào năm 1970
- Biểu diễn: dưới dạng bảng
- Các khái niệm cơ bản
 - Bộ giá trị (Tuple):
 - Bản ghi (record) về đối tượng
 - Biểu diễn dưới dạng các hàng/dòng trong bảng
 - Trường/thuộc tính (Attribute/field)
 - Tên, kiểu, miền giá trị
 - Biểu diễn dưới dạng các cột trong bảng
 - Lược đồ quan hệ (Relation Schema): tập hợp các thuộc tính
 - Quan hệ/thể hiện (Relation/Instance): tập hợp các bộ
 - Biểu diễn dưới dạng bảng
 - Khóa (Key): một nhóm các thuộc tính, cho phép **xác định duy nhất** mỗi bộ trong quan hệ
 - Các phép toán: hợp, giao, tích đề-các, lựa chọn, chiếu, kết nối, ...

<u>MaMH</u>	TênMH
IT3090	CSDL
IT3080	Mạng máy tính

Ví dụ

Row/Tuple/Record

Column/Attribute/Field

name	birth	gpa	grad
Anderson	1987-10-22	3.9	2009
Jones	1990-4-16	2.4	2012
Hernandez	1989-8-12	3.1	2011
Chen	1990-2-4	3.2	2011

Column Types

→ VARCHAR(30)

DATE

FLOAT

INT



TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

MON_HOC

maMH	tenmon	SoTC
IT3080	CSDL	3
IT3090	Mạng máy tính	3
IT2000	Nhập môn CNTT&TT	2

LOP

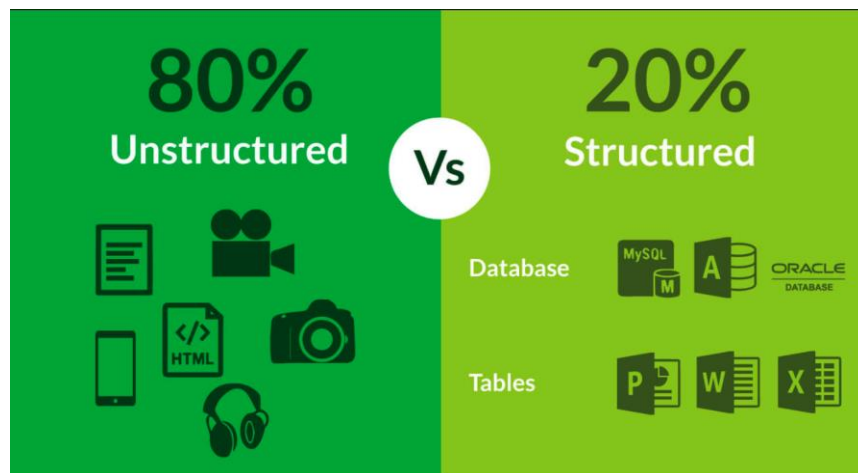
malop	lop	khoa	GVCN	loptruong
IT1	KHMT1	CNTTTT	Ng. V. Anh	SV0011
IT2	KHMT2	CNTTTT	Lê A. Văn	SV0025
IT3	KTMT1	CNTTTT	Ng. T. Thảo	SV0067
IT4	KTMT2	CNTTTT	Ng. V. Quý	SV0034

SINH_VIEN

maSV	tenSV	ngaysinh	nam	diachi	Malop
SV0011	Trần T. Bình	1/4/1981	0	21 T. Q. B	IT1
SV0025	Ng. Đ. Trung	3/2/1980	1	56 Đ. C. V	IT2
SV0067	Trần M. Quế	26/3/1982	0	45 H. B. T	IT3
SV0034	Ng. T. Phương	29/2/1980	0	86 L. T. N	IT4

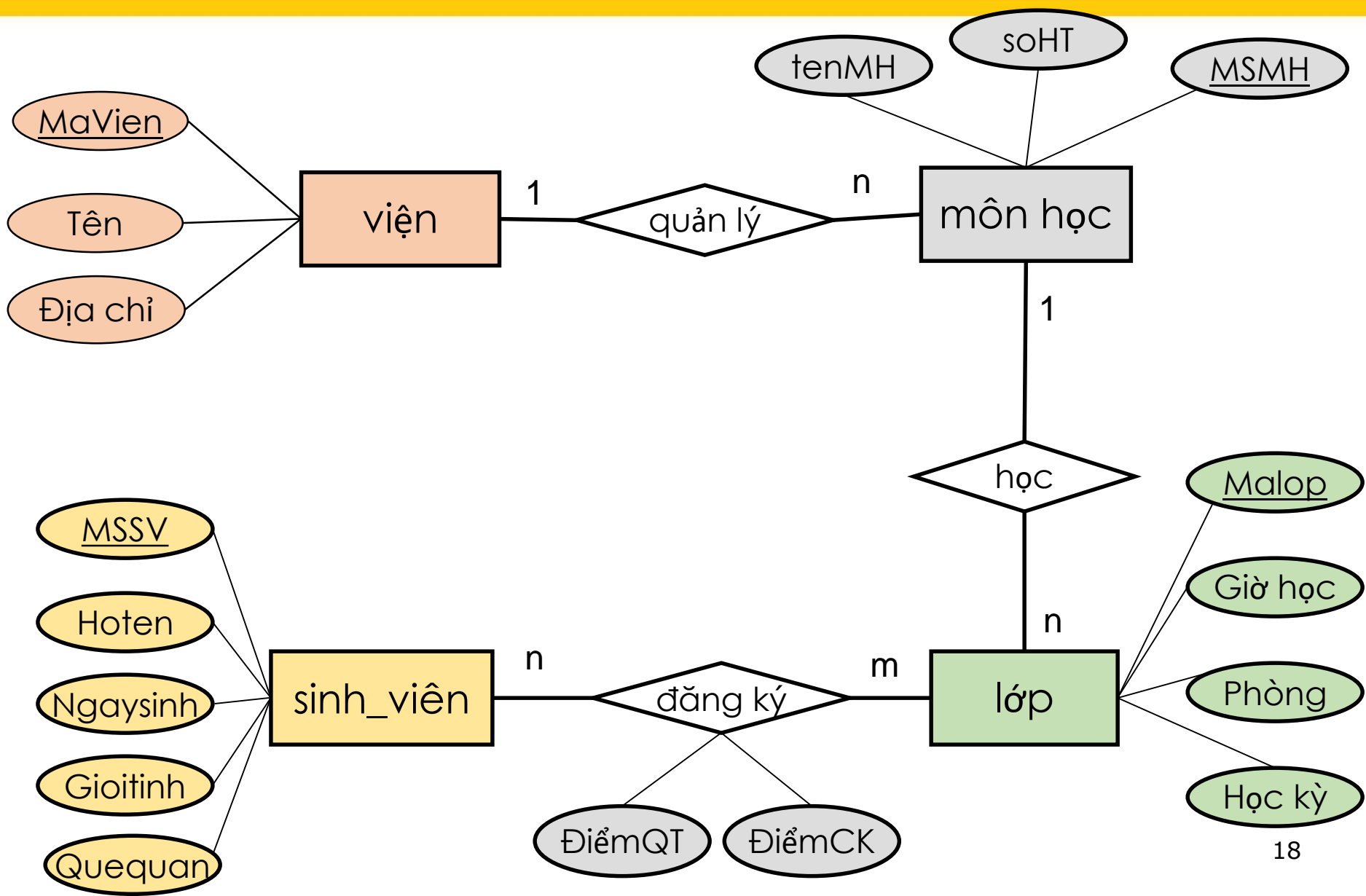
Nhận xét

- Ưu điểm
 - Dễ hiểu và đơn giản với người dùng thông qua cấu trúc dạng bảng
 - Hỗ trợ nhiều phép toán dựa trên lý thuyết tập hợp
- Nhược điểm
 - Hạn chế về ngữ nghĩa
 - Cấu trúc dữ liệu bảng hai chiều cố định, không phù hợp với nhiều dạng dữ liệu đa dạng (video, ảnh, email, file văn bản...)



Mô hình dữ liệu thực thể - liên kết (*Entity-Relational data model*)

- Sự ra đời
 - Xuất phát từ nhu cầu mô hình hoá ngữ nghĩa dữ liệu và phát triển phần mềm
 - Ra đời năm 1975
- Biểu diễn: bằng sơ đồ thực thể - liên kết
- Các khái niệm cơ bản
 - Thực thể: một đối tượng trong thế giới thực
 - Thuộc tính: một đặc tính của một tập thực thể
 - Khoá: một nhóm các thuộc tính cho phép xác định sự duy nhất của một thực thể
 - Liên kết: mối liên hệ có nghĩa giữa nhiều thực thể
 - Mỗi liên kết có thể có các thuộc tính
 - 1-1, 1-n, n-m, đệ quy



Nhận xét

- Ưu điểm
 - Dễ dàng biểu diễn cái mà con người nhận thức từ thế giới thực
 - Biểu diễn ngữ nghĩa phong phú của các thực thể và quan hệ giữa các thực thể
- Nhược điểm
 - Không dễ dàng ánh xạ vào những cấu trúc lưu trữ trên máy tính

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

- Sự ra đời
 - Vào những năm 1980s
- Biểu diễn: sơ đồ lớp
- Các khái niệm cơ bản
 - Đối tượng: một đối tượng trong thế giới thực
 - 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

Ví dụ

```
class sinh_vien {  
    string maSV;  
    string tenSV;  
    date ngaysinh;  
    boolean nam;  
    string diachi;  
    string lop;  
    string khoa;  
  
    string ten();  
    string ngay_sinh();  
    string dia_chi();  
    string lop();  
    void gan_DC(string DC_moi);  
    void gan_lop(string lop);  
}
```

Ví dụ

```
class lop {  
    string tenlop;  
    string khoa;  
}
```

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

Nhận xét

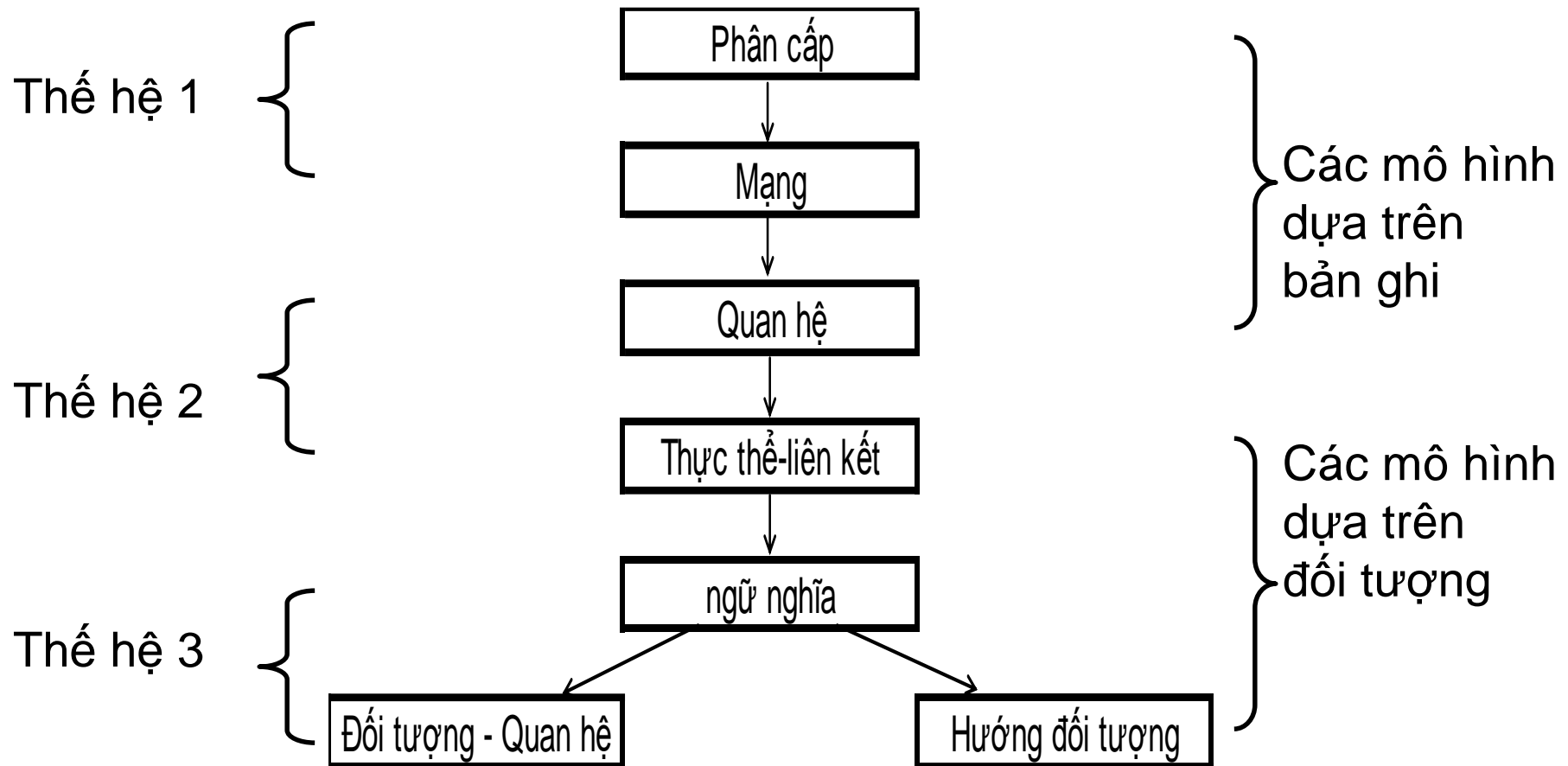
- Ưu điểm
 - Cho phép định nghĩa kiểu đối tượng phức tạp
 - Tính chất: bao đóng (*encapsulation*), kế thừa (*heritage*), đa hình (*polymorphism*)
- Nhược điểm
 - Cấu trúc lưu trữ phức tạp
 - Khó khăn trong tối ưu câu truy vấn

So sánh và đánh giá

Nhắc lại: Mô hình dữ liệu là một tập hợp các khái niệm dùng để mô tả về dữ liệu (cấu trúc dữ liệu, thao tác dữ liệu, và ràng buộc dữ liệu)

	Mô hình dữ liệu phân cấp	Mô hình dữ liệu mạng	Mô hình dữ liệu quan hệ	Mô hình dữ liệu TT-LK	Mô hình dữ liệu HĐT
Biểu diễn ngữ nghĩa DL	hạn chế	hạn chế	<i>tương đối đa dạng</i>	đa dạng	đa dạng
Lưu trữ DL	dữ liệu lặp lại	s/d nhiều con trỏ	dễ dàng và hiệu quả	khó lưu trữ	<i>cấu trúc phức tạp</i>
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	<i>không hiệu quả</i>

Phân loại các mô hình

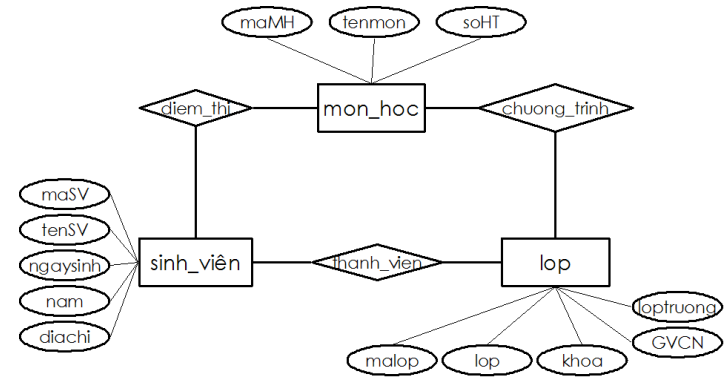


Các bước xây dựng một CSDL



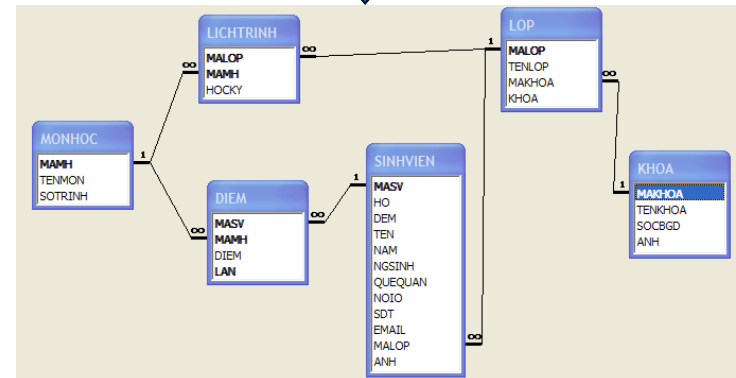
Mô tả ứng dụng

1: PHÂN TÍCH



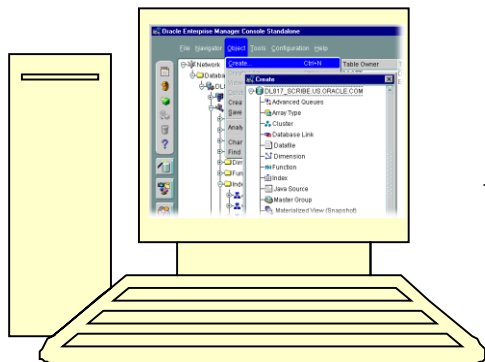
Mô hình hoá dữ liệu (vd: Thực thể-liên kết)

2: THIẾT KẾ



Mô tả dữ liệu logic với một mô hình dữ liệu cụ thể (vd: Mô hình dữ liệu quan hệ)

3: CÀI ĐẶT



Cài đặt với một hệ quản trị CSDL (vd: MySQL, SQL Server)

Nội dung

Tổng quan các mô hình dữ liệu

Mô hình dữ liệu thực thể - liên kết

Mô hình dữ liệu quan hệ

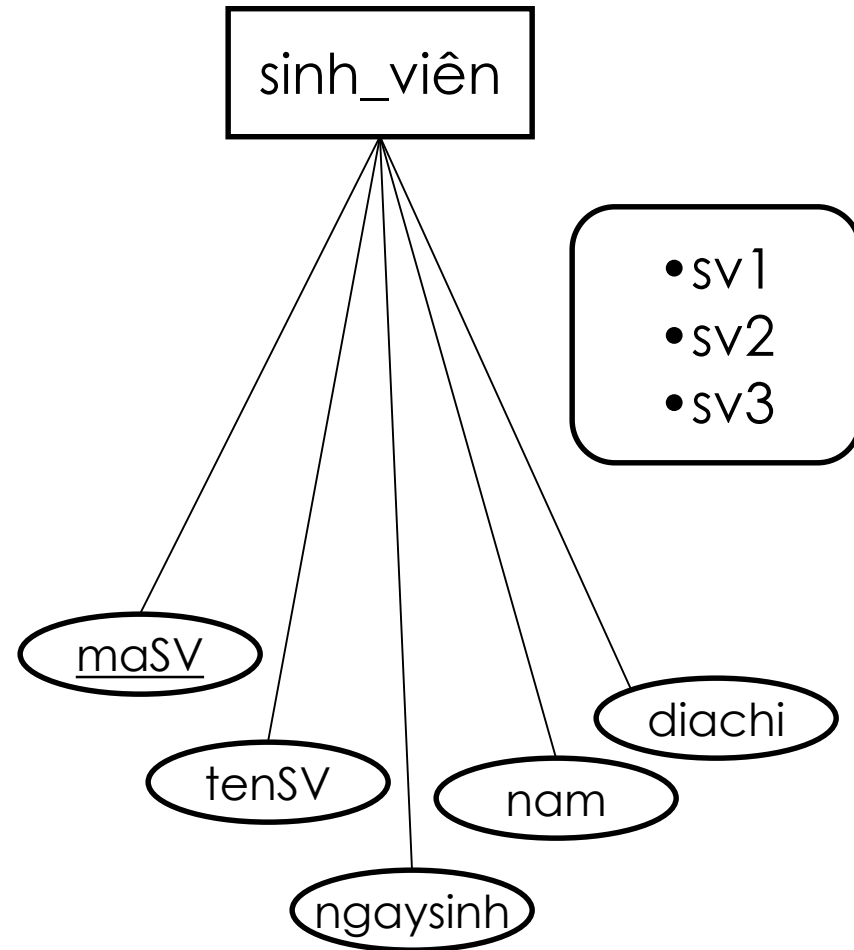
Biến đổi mô hình dữ liệu thực thể - liên kết sang mô hình dữ liệu quan hệ

Đặc điểm mô hình dữ liệu Thực thể - Liên kết

- Thích hợp để mô hình hoá dữ liệu cho CSDL
 - Gần gũi với nhận thức của con người → dễ sử dụng
 - Dễ chuyển đổi sang mô hình dữ liệu quan hệ
 - Dựa trên các khái niệm chính
 - **Thực thể**: một đối tượng trong thế giới thực
 - **Tập thực thể**: các thực thể có cùng các tính chất
 - **Thuộc tính**: đặc tính của một tập thực thể
 - **Khoá**: xác định sự duy nhất của một thực thể
 - **Liên kết**: mối liên hệ có nghĩa giữa nhiều thực thể
 - **Tập liên kết**: tập hợp các liên kết cùng kiểu
- Được biểu diễn bởi **sơ đồ thực thể - liên kết**

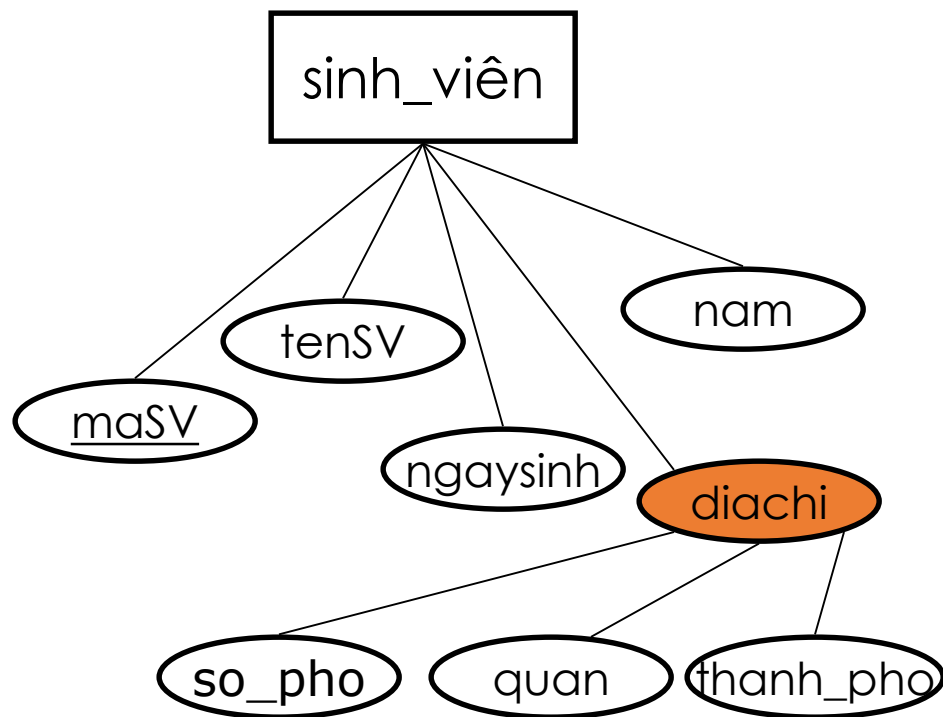
Thực thể và Thuộc tính

- **Thực thể**: một đối tượng trong thế giới thực
- **Tập thực thể**: gồm các thực thể có cùng tính chất
- **Thuộc tính**: đặc tính của một tập thực thể
 - **Miền giá trị**: tập các giá trị có thể
- **Khoá**: xác định sự duy nhất của một thực thể



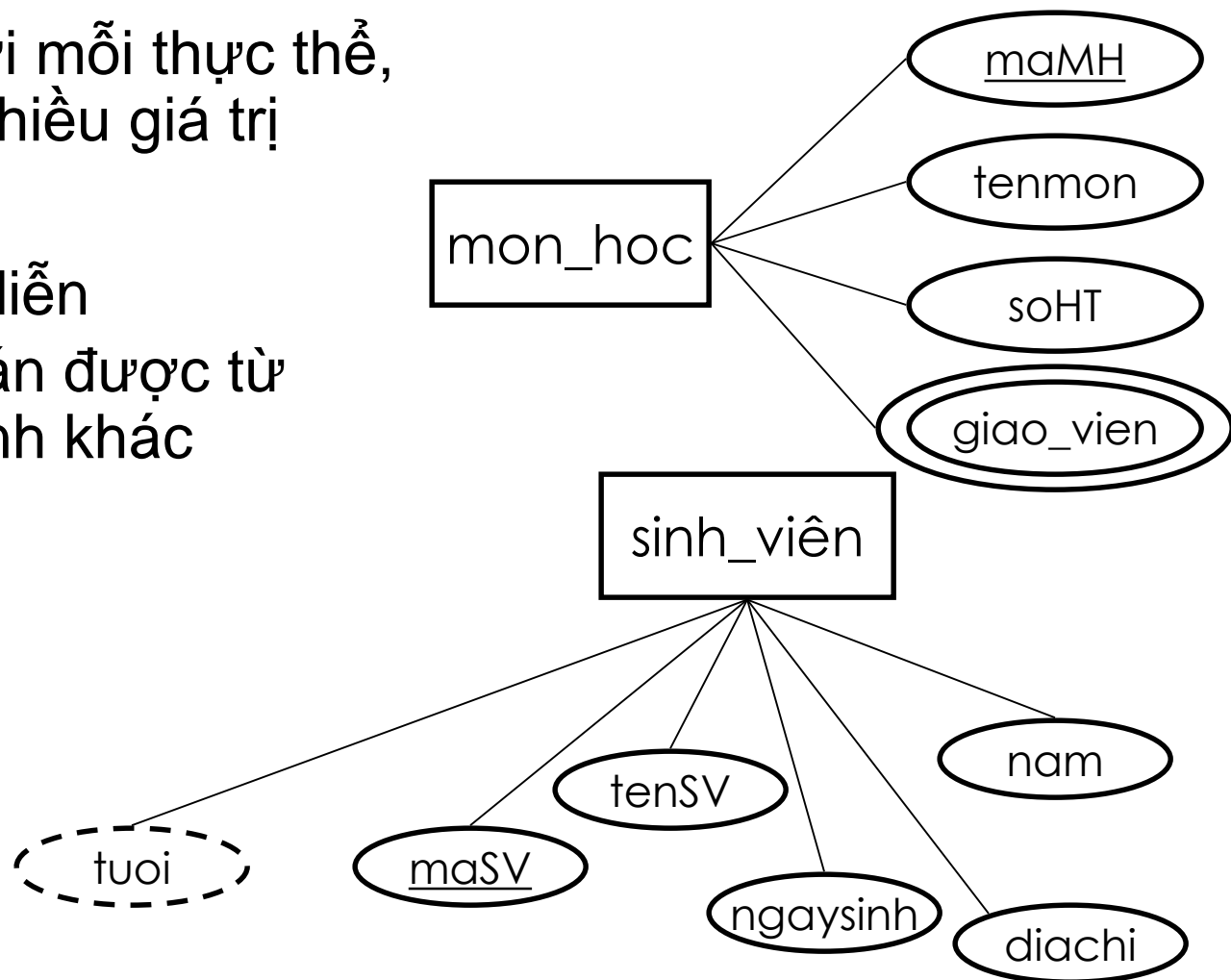
Kiểu thuộc tính

- 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



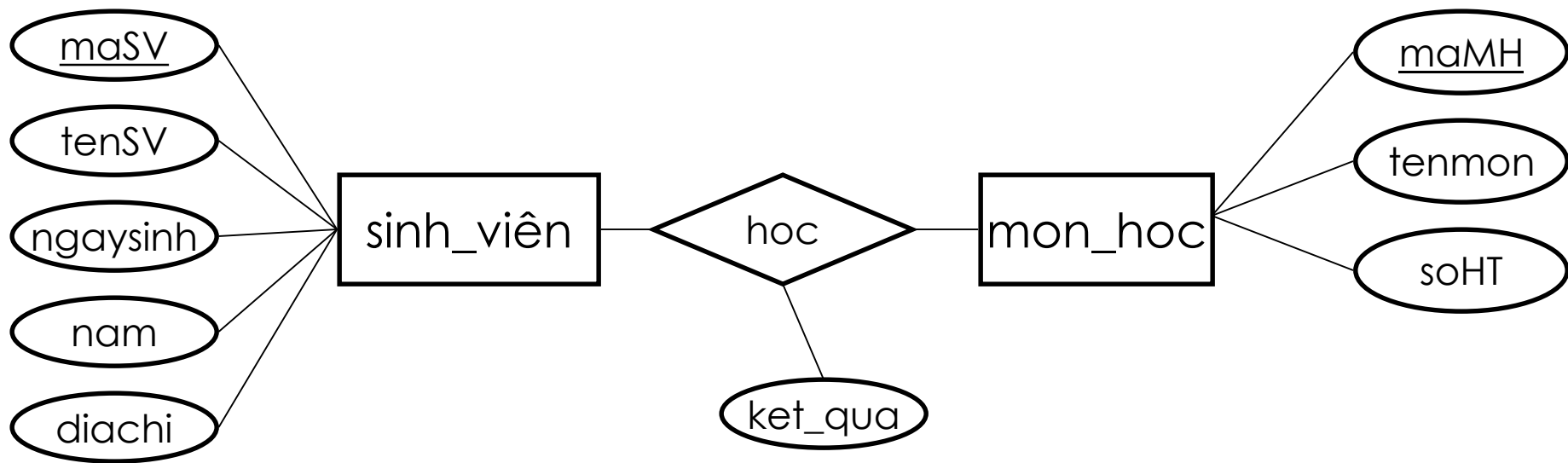
Kiểu thuộc tính (2)

- Thuộc tính đa 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



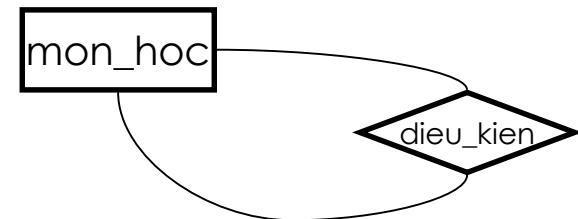
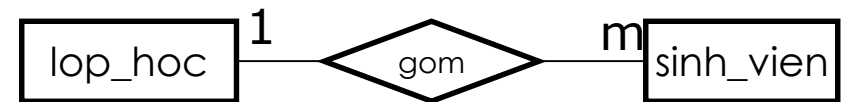
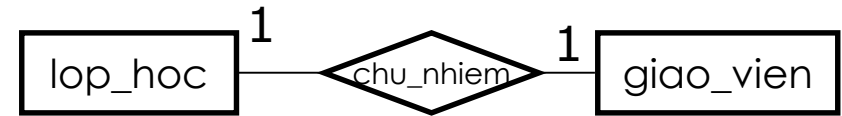
Liên kết

- Là mối liên hệ/sự kết hợp giữa nhiều thực thể
- Có thể có thuộc tính



Ràng buộc của kết nối

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



Tạo sơ đồ thực thể - liên kết

- B1: Xác định các thực thể
- B2: Xác định các liên kết giữa các thực thể
 - Bậc của liên kết
 - Bậc 1: đệ quy
 - Bậc 2: 2 thực thể
 - Bậc 3: 3 thực thể
 - ...
 - Ràng buộc (1-1, 1-n, n-m, đệ quy)

Bài tập 1

- Phân tích và thiết kế 1 CSDL gồm các thông tin trong quản lý học tập như sau:
 - Mỗi sinh viên khi vào trường sẽ được cấp một **mã số sinh viên** duy nhất sau khi cung cấp các thông tin cá nhân gồm **họ tên, ngày sinh, giới tính, quê quán**.
 - Ở đầu mỗi học kỳ, sinh viên sẽ đăng ký các lớp học trong kỳ đó. Thông tin mỗi lớp gồm **mã lớp, giờ học, phòng học, kỳ học**. Kết quả học tập của SV gồm **điểm quá trình và điểm cuối kỳ** sẽ được ghi lại.
 - Mỗi lớp sẽ học một môn học. Thông tin môn học được quản lý gồm: **mã số môn học, tên môn, số học trình**.
 - Mỗi môn học sẽ được quản lý bởi một viện cụ thể. Thông tin khoa viện: **mã viện, tên, và địa chỉ**



Bài tập 2

Phân tích và thiết kế 1 CSDL gồm các thông tin trong công ty

- Công ty được tổ chức bởi các phòng ban. Mỗi phòng ban có **tên, mã số và một 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 địa điểm** làm việc khác nhau
- Mỗi phòng điều phối một số dự án. Mỗi dự án có **tên và mã số, địa điểm thực hiện**
- Thông tin về nhân viên 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** 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 trong dự án cũng được lưu trữ
- Thông tin về con cái của từng nhân viên: **tên, giới tính, ngày sinh**



Nội dung

Tổng quan các mô hình dữ liệu

Mô hình thực thể - liên kết

Mô hình dữ liệu quan hệ

Biến đổi mô hình dữ liệu thực thể - liên kết sang mô hình dữ liệu quan hệ

Đặc điểm

- Dựa trên lý thuyết tập hợp
- Các khái niệm cơ bản
 - Thuật ngữ toán học: quan hệ, bộ và thuộc tính
 - Thuật ngữ hướng dữ liệu: bảng, bản ghi và trường
- Được biểu diễn bởi sơ đồ (lược đồ) quan hệ

Thuộc tính (Attribute) - Trường (Field)

- Là tính chất riêng biệt của một đối tượng cần được lưu trữ trong CSDL để phục vụ cho việc khai thác dữ liệu về đối tượng.
- Ký hiệu: A
- Kiểu dữ liệu: text, number, boolean, date/time
 - maSV: char(10)
 - gioitinh: boolean
- Miền giá trị $Dom(A)$: tập tất cả các giá trị A có thể nhận
 - $Dom(maSV) = \{char(10)\}$
 - $Dom(gioitinh) = \{0, 1\}$

Bộ (Tuple) - Bản ghi (Record)

- Các thông tin của một đối tượng
- Ký hiệu

$t(a_1, a_2, \dots, a_n)$

$t(a_1, a_2, \dots, a_n) \in \text{Dom}(A_1) \times \dots \times \text{Dom}(A_n)$

name	birth	gpa	grad
Anderson	1987-10-22	3.9	2009
Jones	1990-4-16	2.4	2012
Hernandez	1989-8-12	3.1	2011
Chen	1990-2-4	3.2	2011

Row/Tuple/Record

Column/Attribute/Field

Column Types → VARCHAR(30) DATE FLOAT INT

Quan hệ (Relation) - Bảng (Table)

- Lược đồ/sơ đồ quan hệ (Schema): được xác định trên một tập các thuộc tính A_i . Là sự trừu tượng hóa mức logic
 - Ký hiệu: $S(A_1, A_2, \dots, A_n)$
 - Sinhvien(maSV, tenSV, ngaysinh, gioitinh)
Monhoc(maMH, tenmon, soTC)
- Quan hệ/thể hiện của quan hệ (Relation/Instance): tập hợp các bộ giá trị
 - Ký hiệu $t(a_1, a_2, \dots, a_n)$ với
 - $a_i \in \text{DOM}(A_i)$
- Một quan hệ được biểu diễn dưới dạng một bảng
 - Bậc của quan hệ: số lượng thuộc tính trong quan hệ
 - Lực lượng của quan hệ: số các bộ giá trị trong quan hệ
- Lược đồ CSDL (Database schema): tập hợp các lược đồ quan hệ, dùng để biểu diễn toàn bộ CSDL

Khoá - Key

- Định nghĩa

- Cho $S(A_1, A_2, \dots, A_n)$, $K \subseteq \{A_i\}$,
- K là khoá nếu với $\forall t_1, t_2 \in R$, $t_1.K \neq t_2.K$

SINH_VIEN (maSV, tenSV, ngaysinh, diachi, lop)

- Tính chất

- $K \subseteq K'$, $K' \subseteq \{A_i\}$, K là khoá $\Rightarrow K'$ cũng là khoá

SINH_VIEN (maSV, tenSV, ngaysinh, diachi, lop)

SINH_VIEN (maSV, tenSV, ngaysinh, diachi, lop)

Phân loại khoá

- Khoá tối thiểu

- Cho $S(A_1, A_2, \dots, A_n)$, $K \subseteq \{A_i\}$,
- K là khoá tối thiểu nếu K là khoá và $\nexists K' \subset K$ mà K' là khoá

- Khoá chính (Primary key): 1 khoá tối thiểu tốt được chọn là **khóa chính**

- Khoá ngoài (Foreign Key)

- $S(A_1, A_2, \dots, A_n)$, $K \subseteq \{A_i\}$,
 $S'(A'_1, A'_2, \dots, A'_m)$
- K là khoá ngoài của S tham chiếu đến quan hệ S' nếu K là khoá chính của S'
- **malop** là Khóa ngoài của SINH_VIEN tham chiếu tới LOP, vì malop là khóa chính của LOP

SINH_VIEN(maSV, tenSV, ngay sinh, diachi, **malop**)

⋮

LOP(malop, lop, GVCN, loptruong)

Nội dung

Tổng quan các mô hình dữ liệu

Mô hình thực thể - liên kết

Mô hình dữ liệu quan hệ

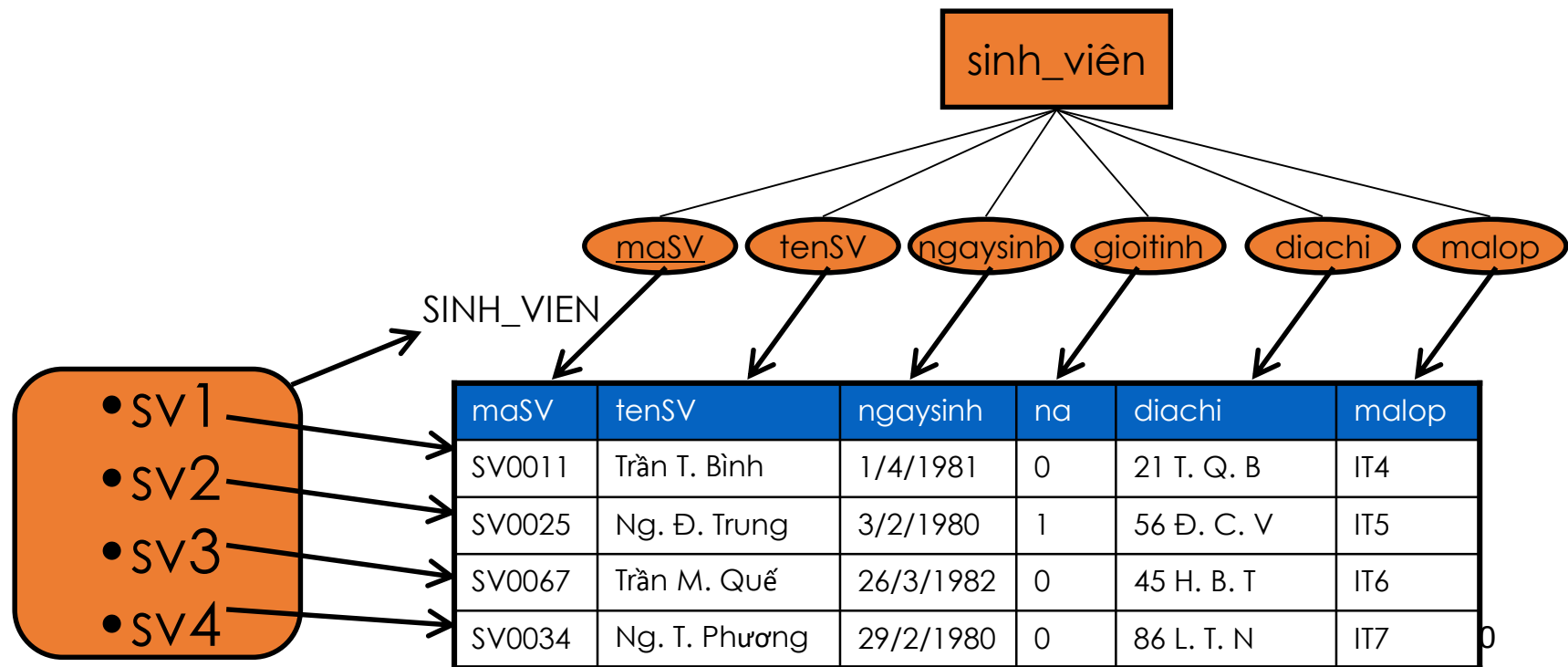
Biến đổi mô hình dữ liệu thực thể - liên kết sang mô hình dữ liệu quan hệ

Các bước biến đổi

- B1. Biến đổi các tập thực thể
- B2. Biến đổi các liên kết
- Lưu ý: các quan hệ có cùng khoá thì gộp lại

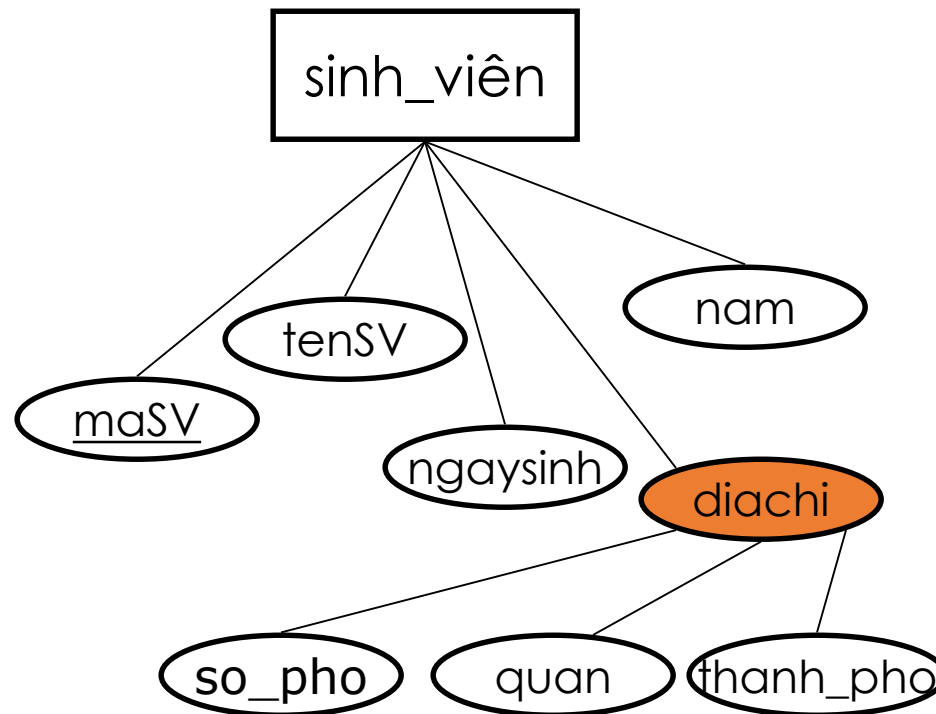
B1. Biến đổi các tập thực thể

- Biến đổi tập thực thể sang quan hệ
 - 1 thực thể → 1 bộ
 - thuộc tính → thuộc tính (trường),
 - khoá của tập thực thể → khoá của quan hệ



Thuộc tính phức

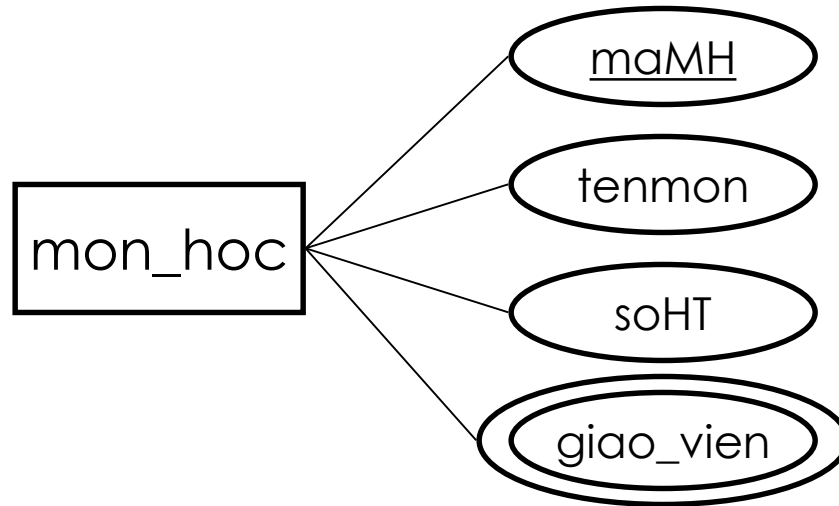
- Thay *thuộc tính phức* bởi *các thuộc tính thành phần*
sinh_vien(maSV, tenSV, ngaysinh, **so_pho**, **quan**, **thanh_pho**, nam)



Thuộc tính đa trị

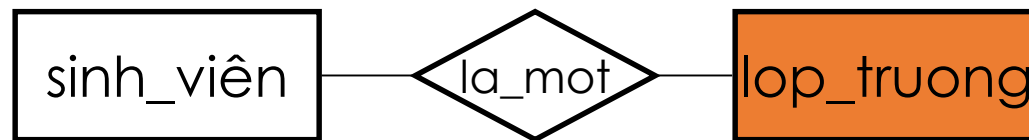
- Thêm 1 quan hệ mới xác định bởi thuộc tính đa trị và khoá của tập thực thể tương ứng

MH_GV (maMH, giao_vien)



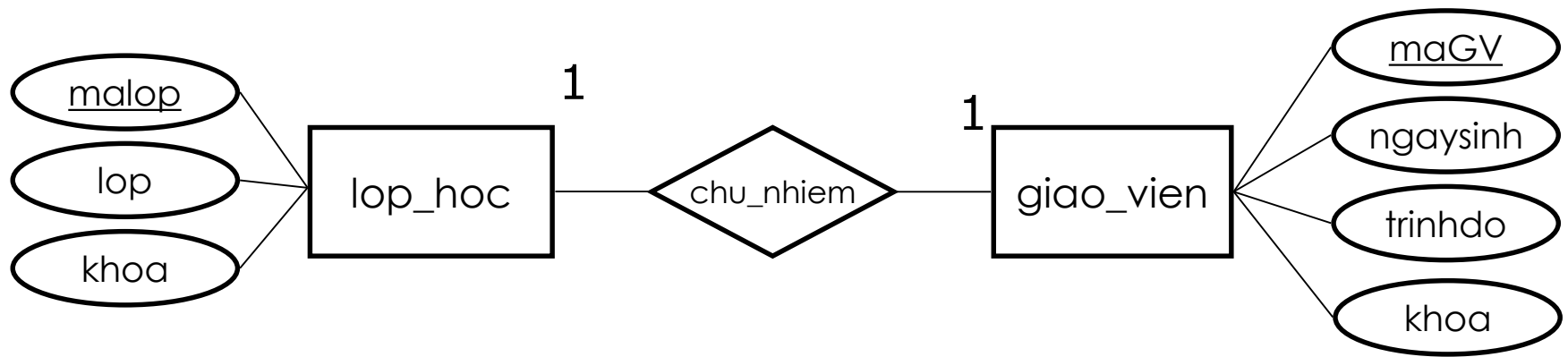
Một tập thực thể xác định từ tập thực thể khác (E) qua một liên kết

- Tạo một quan hệ chứa khoá của E
LOPTRUONG (maSV)



B2. Biến đổi các liên kết

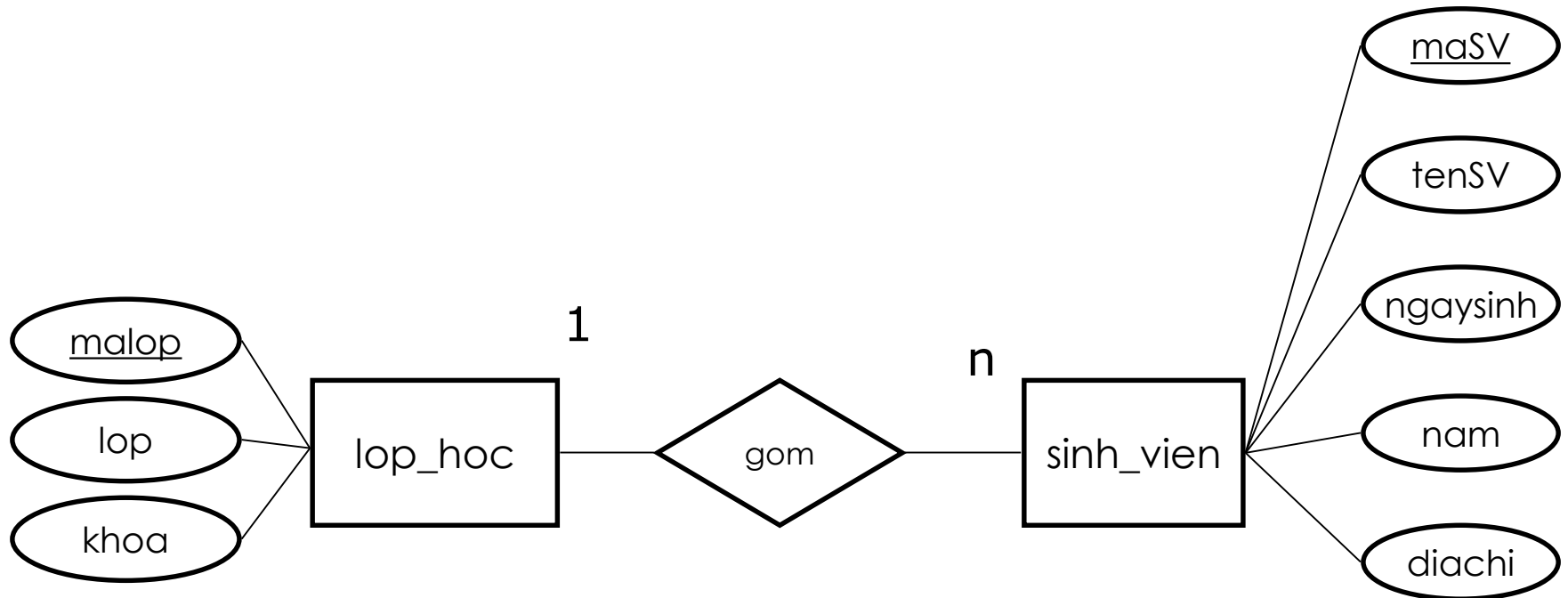
- Liên kết 1-1: dùng khoá ngoài
LOP_HOC(malop,lop,khoa,*maGV*)



B2. Biến đổi các liên kết

- **Liên kết 1-n:** dùng khoá ngoài
 - Thêm khoá chính của quan hệ bên 1 vào quan hệ bên n làm khoá ngoài

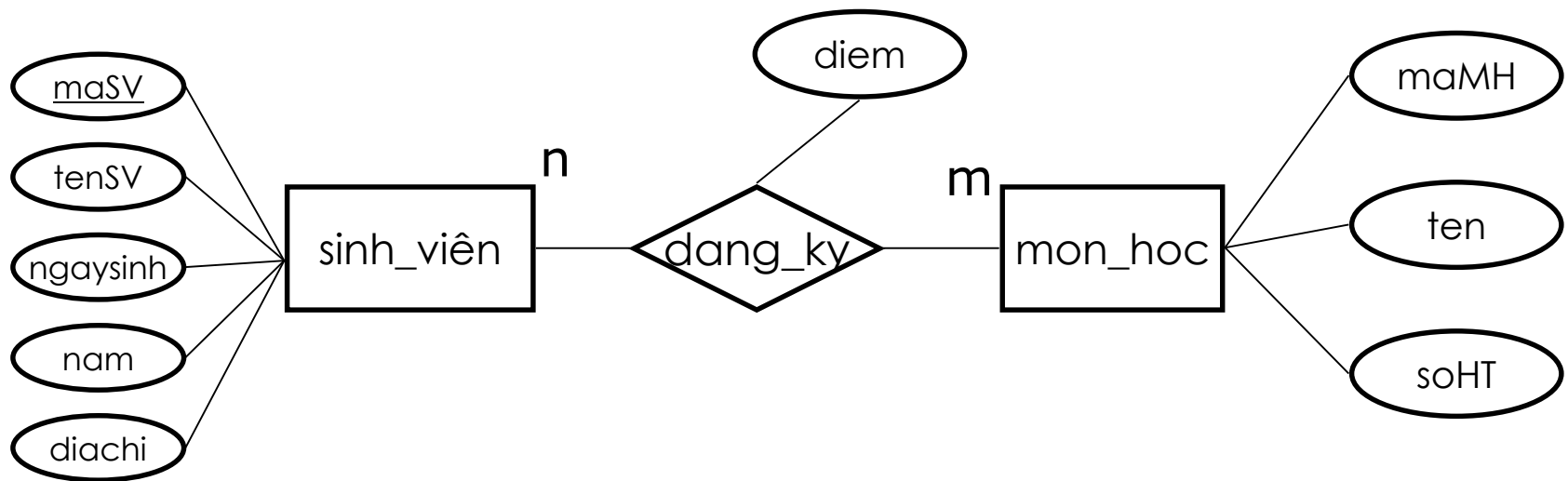
SINH_VIEN (maSV, tenSV, ngaysinh, nam, diachi, malop)



B2. Biến đổi các liên kết

- **Liên kết n-m:** 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)

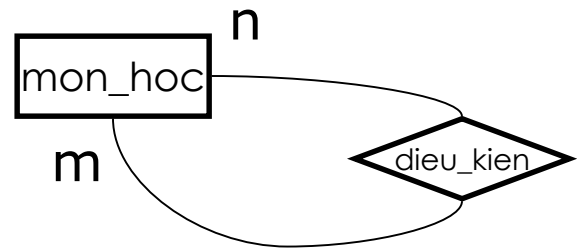


B2. Biến đổi các liên kết

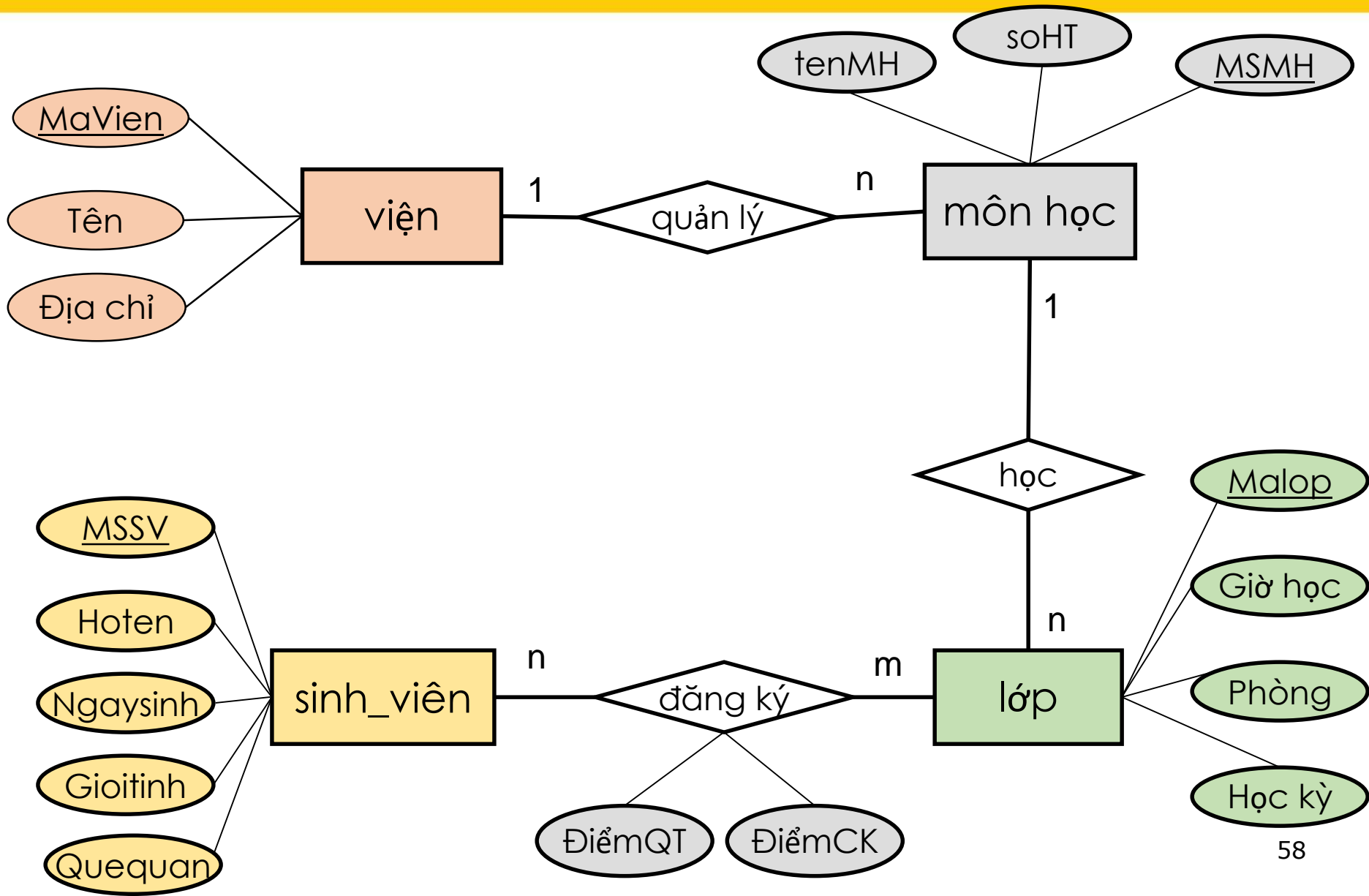
- **Độ quy:** quy về biến đổi của 1-1, 1-n, n-m

MonHoc (MaMH, TenMH, SoTC) ;

DieuKien (MaMH, MaMH-DieuKien)

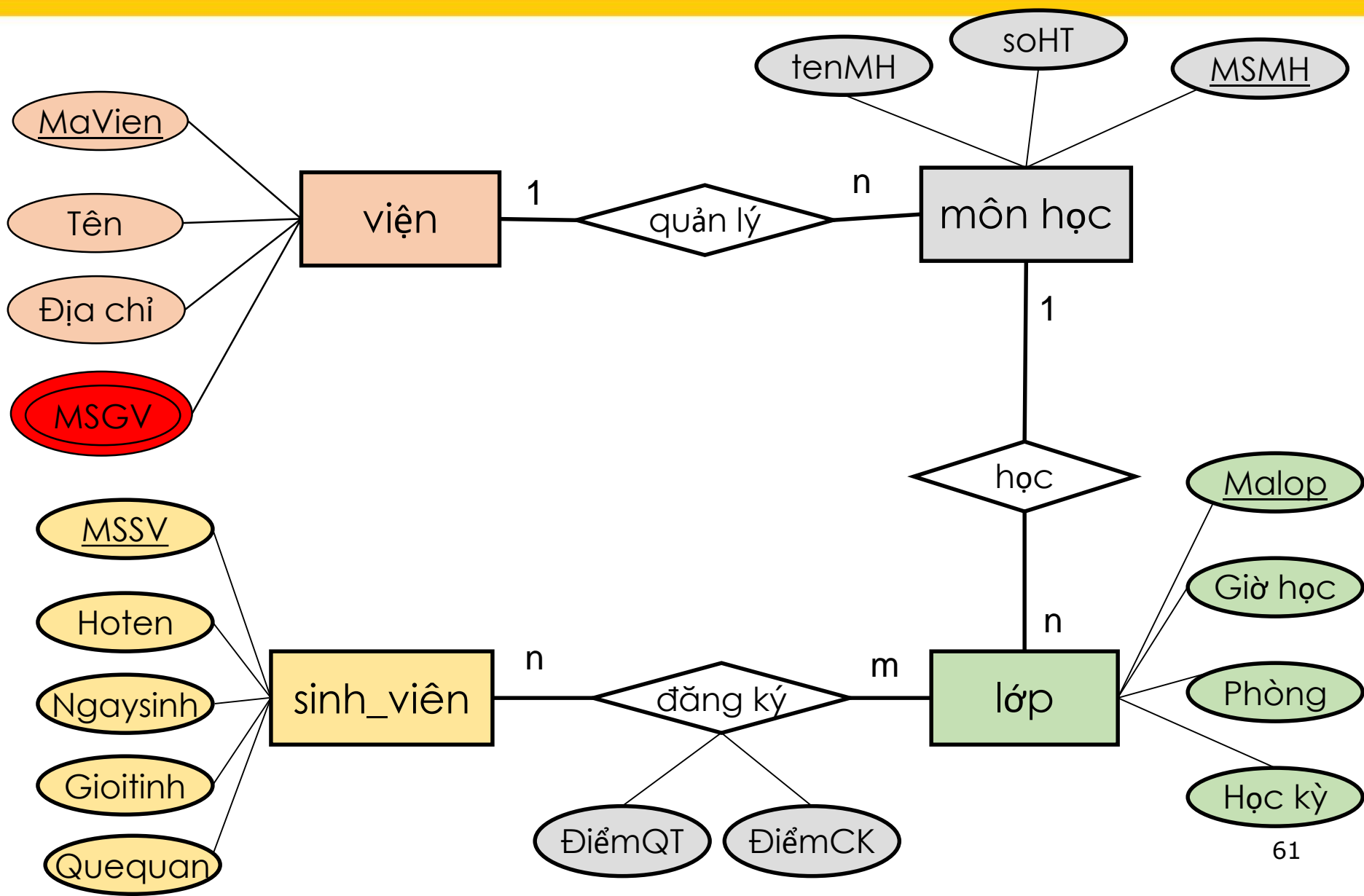


Quản lý đào tạo



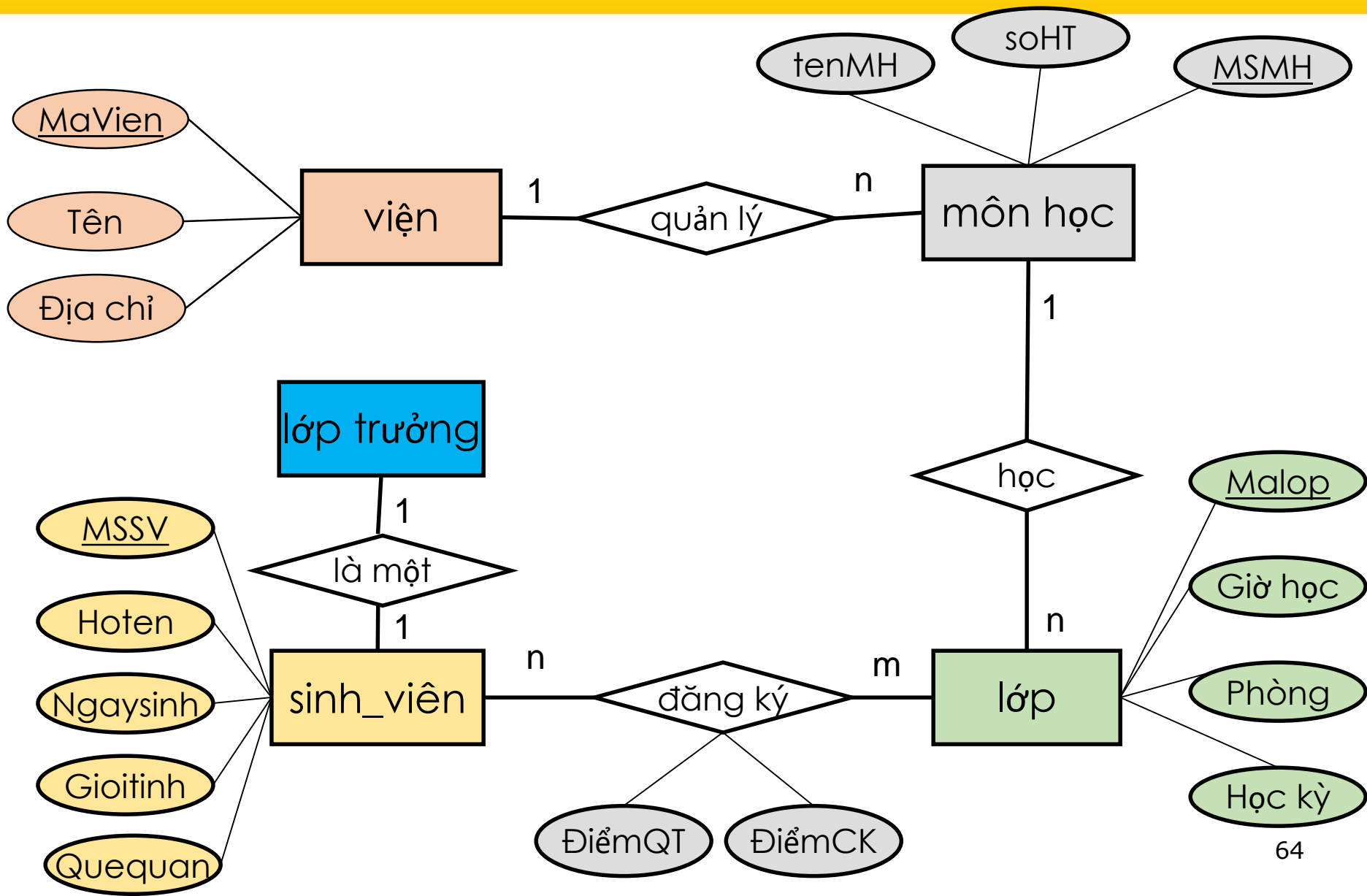


Có thuộc tính đa trị



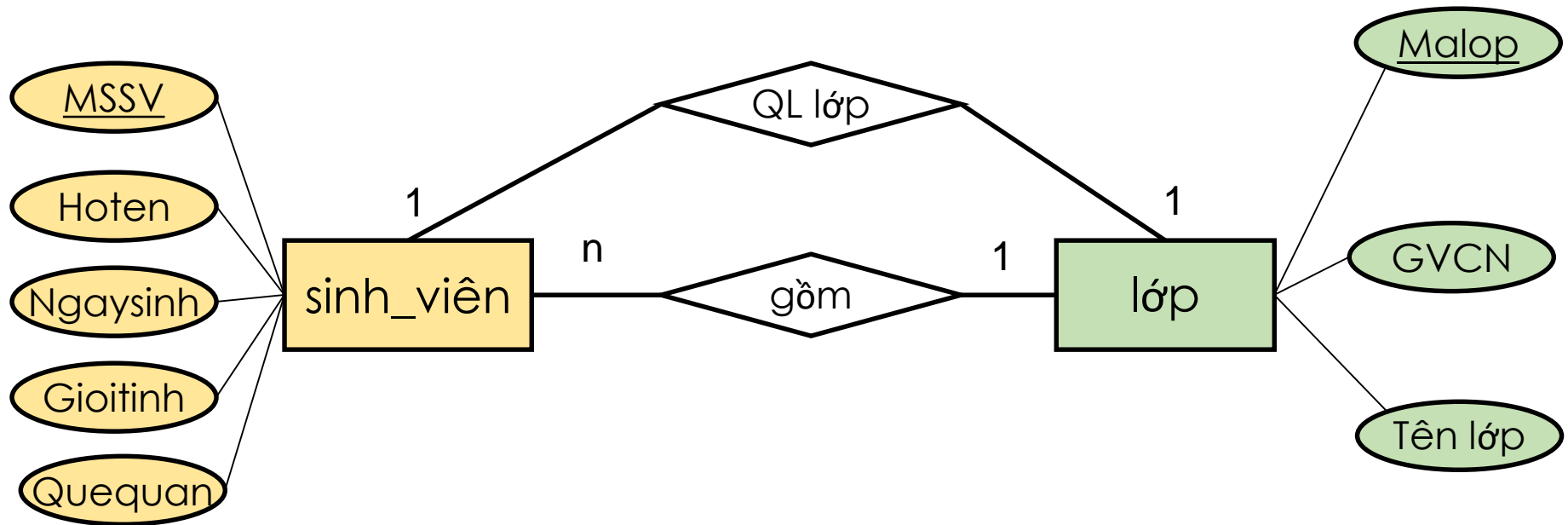


Có liên kết Là – một (is – a)

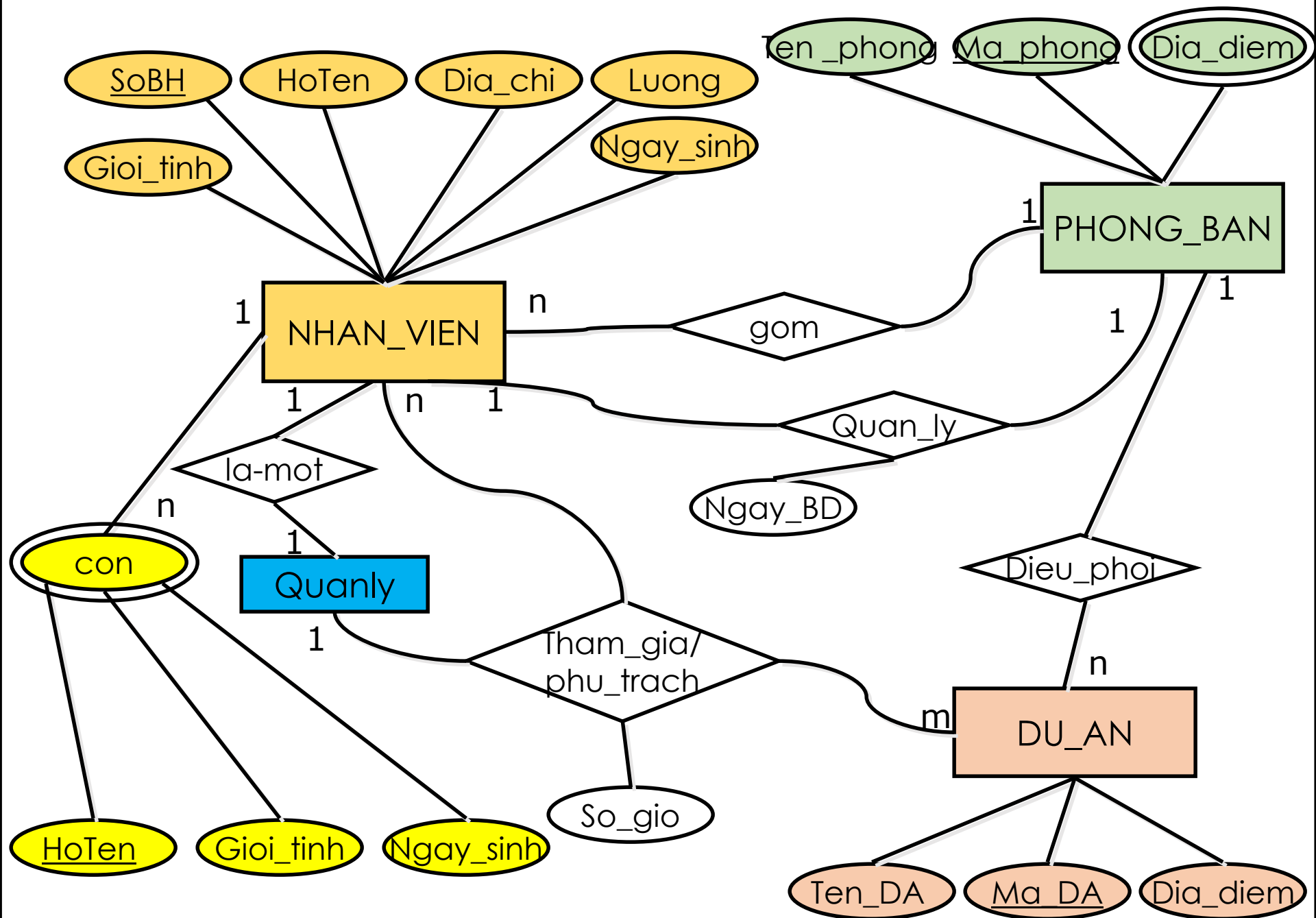




Nhiều liên kết giữa hai thực thể



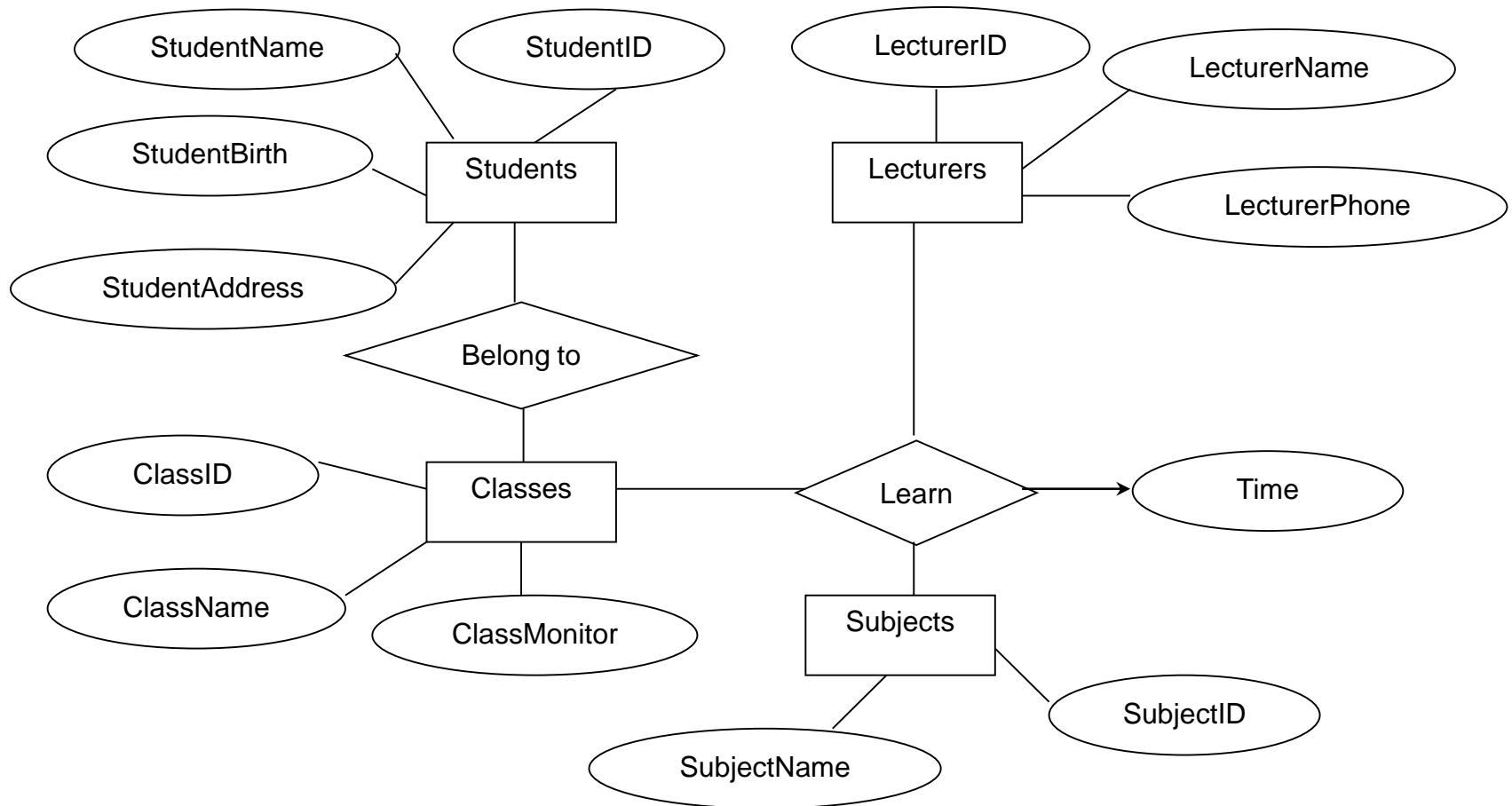






Bài tập

- Biến đổi sơ đồ thực thể-liên kết → sơ đồ quan hệ



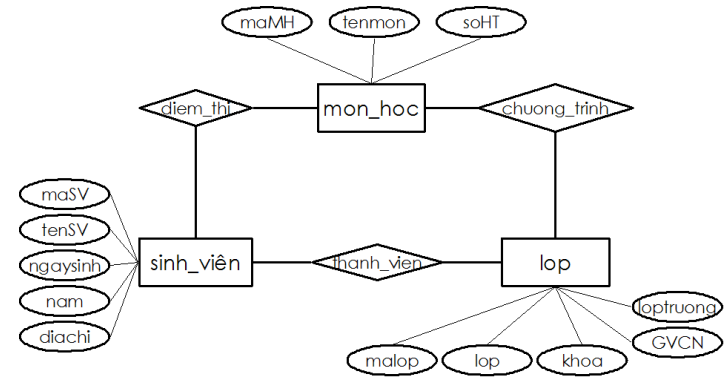


Các bước xây dựng một CSDL



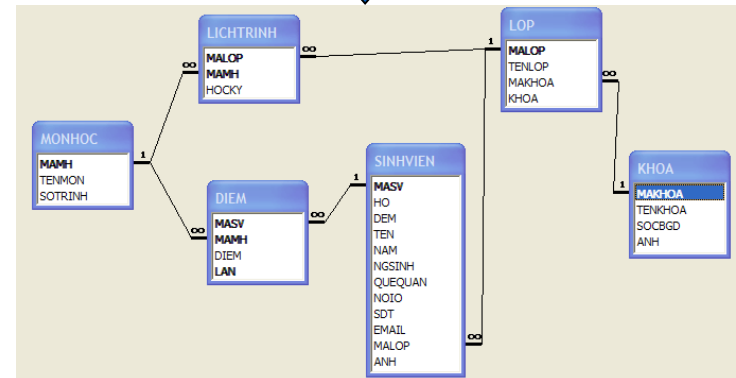
Mô tả ứng dụng

1: PHÂN TÍCH



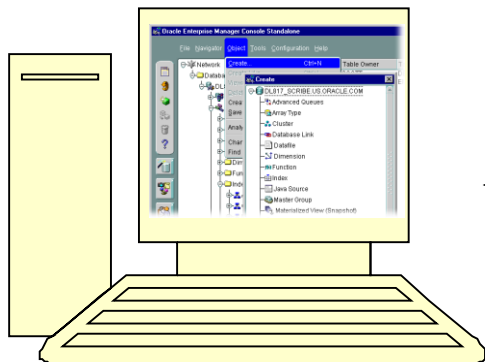
Mô hình hoá dữ liệu (vd: Thực thể-liên kết)

2: THIẾT KẾ



Mô tả dữ liệu logic với một mô hình dữ liệu cụ thể (vd: Mô hình dữ liệu quan hệ)

3: CÀI ĐẶT



Cài đặt với một hệ quản trị CSDL (vd: MySQL, SQL Server)

