

PHÁT TRIỂN ỨNG DỤNG CSDL 1

Chương 1. PHÂN TÍCH DỮ LIỆU

GVLT: Hoàng Anh Tú

Nội dung trình bày

- Mô hình thực thể kết hợp nguyên thủy
- Mô hình thực thể kết hợp mở rộng
- Phương pháp phân tích dữ liệu
- Quy tắc mô hình hóa quan niệm dữ liệu
- Tiêu chuẩn chọn lựa khái niệm

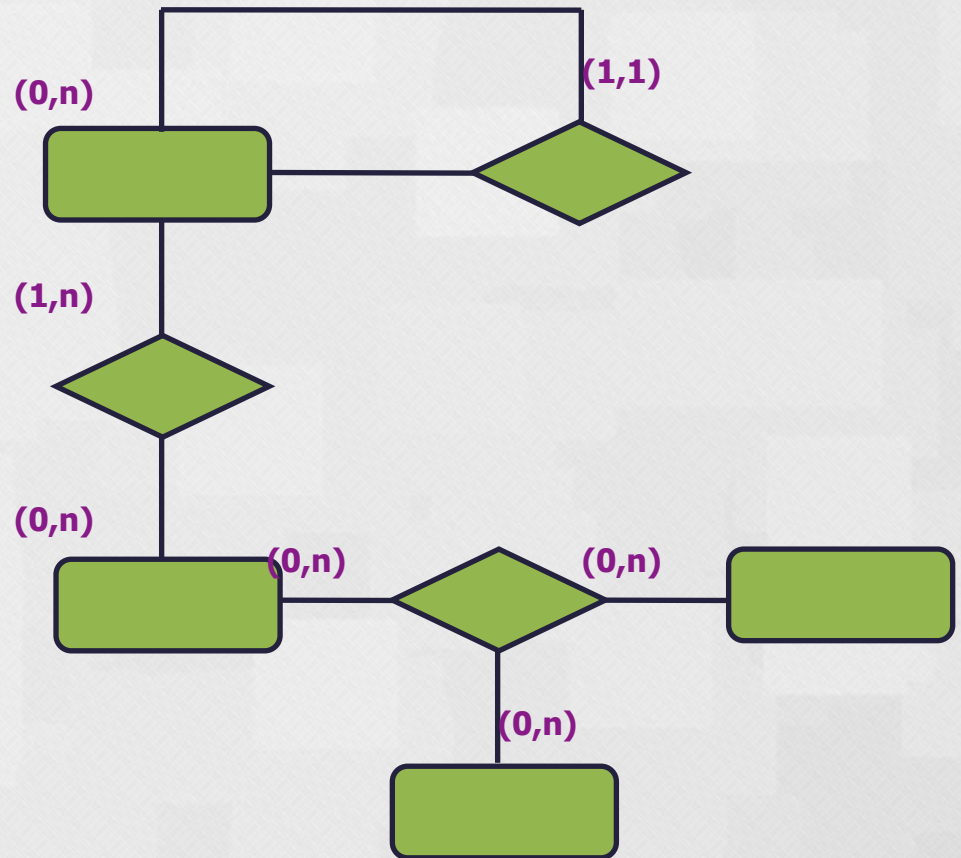
Mô hình thực thể kết hợp

- Entity Relationship Diagram (ERD)
- Được giới thiệu bởi Chen, 1976
- Phổ biến rộng rãi trong thiết kế quan niệm dữ liệu
- ANSI công nhận mô hình chuẩn, 1988
- Gồm
 - Mô hình nguyên thủy
 - Mô hình mở rộng

Mô hình thực thể kết hợp (tt)

○ Khái niệm

- Thực thể
- Thể hiện
- Thuộc tính
- Mối kết hợp
- Vai trò
- bảng số



Thực thể

- Biểu diễn 1 lớp khái niệm trong thế giới thực
 - Trực quan
 - Con người: nhân viên, sinh viên, khách hàng...
 - Nơi chốn: phòng học, chi nhánh, văn phòng...
 - Đối tượng: sách, máy móc, sản phẩm, xe...
 - Sự kiện: đăng ký, bán hàng, đặt trước, yêu cầu...
 - Không trực quan
 - Tài khoản, thời gian, khóa học, khả năng, nguồn vốn...

Thực thể (tt)

○ Ký hiệu

Tên thực thể

○ Ví dụ

Danh từ hoặc cụm danh từ

Nước giải khát

Khách hàng

Đơn đặt hàng

Mối kết hợp

- Biểu diễn sự kết hợp hệ ngữ nghĩa giữa 2 hay nhiều thực thể
 - Sự kiện nối kết
 - Mối quan hệ vật lý
- Ký hiệu



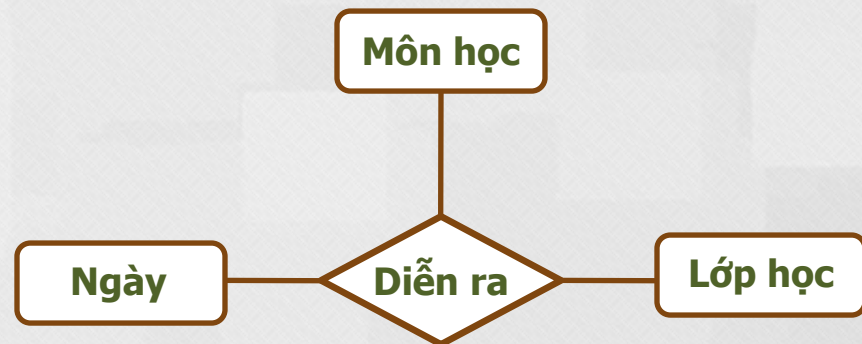
Động từ hoặc cụm từ mô tả mối quan hệ

Mối kết hợp (tt)

○ Ví dụ



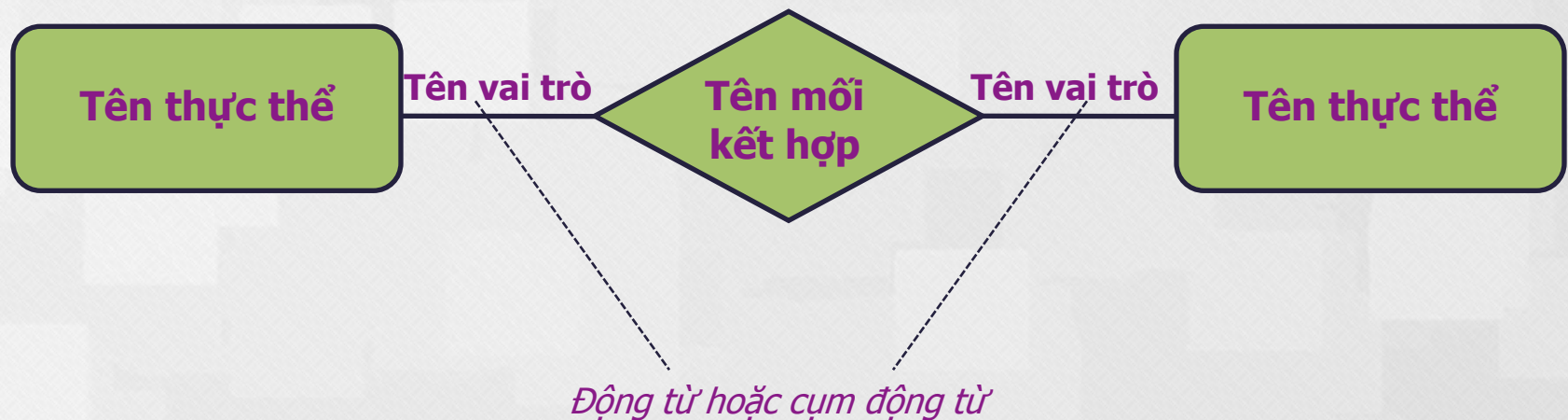
Mối quan hệ phản thân



Mối quan hệ đa phân

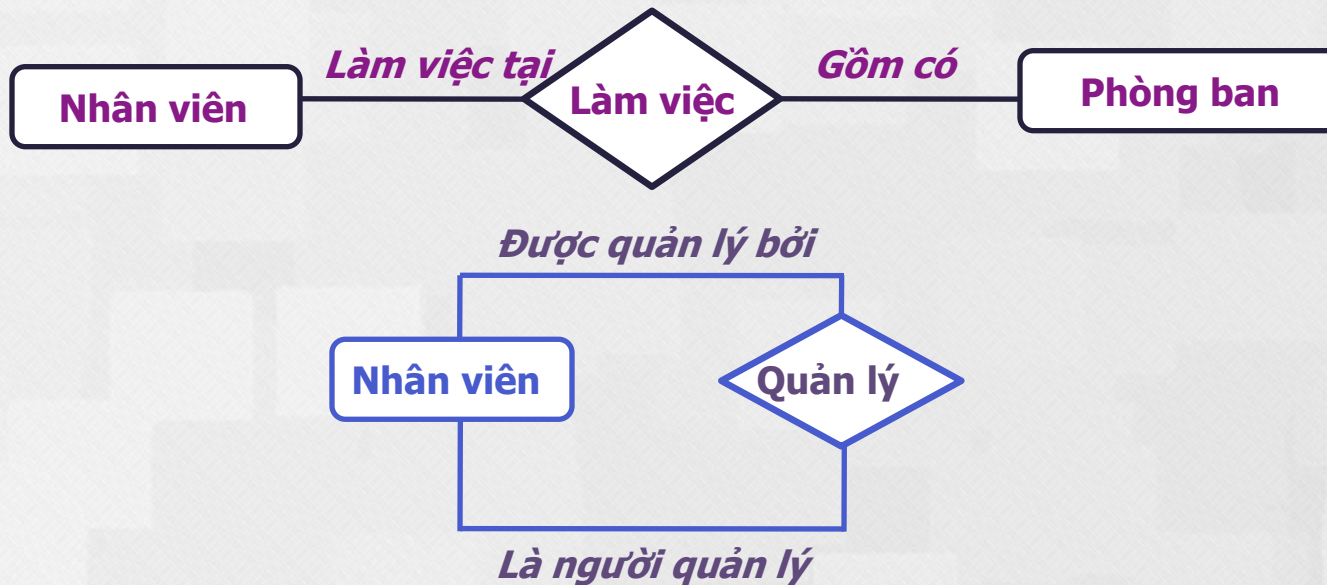
Vai trò

- Biểu diễn ngữ nghĩa của một thực thể tham gia vào mỗi kết hợp
 - Ký hiệu



Vai trò (tt)

○ Ví dụ



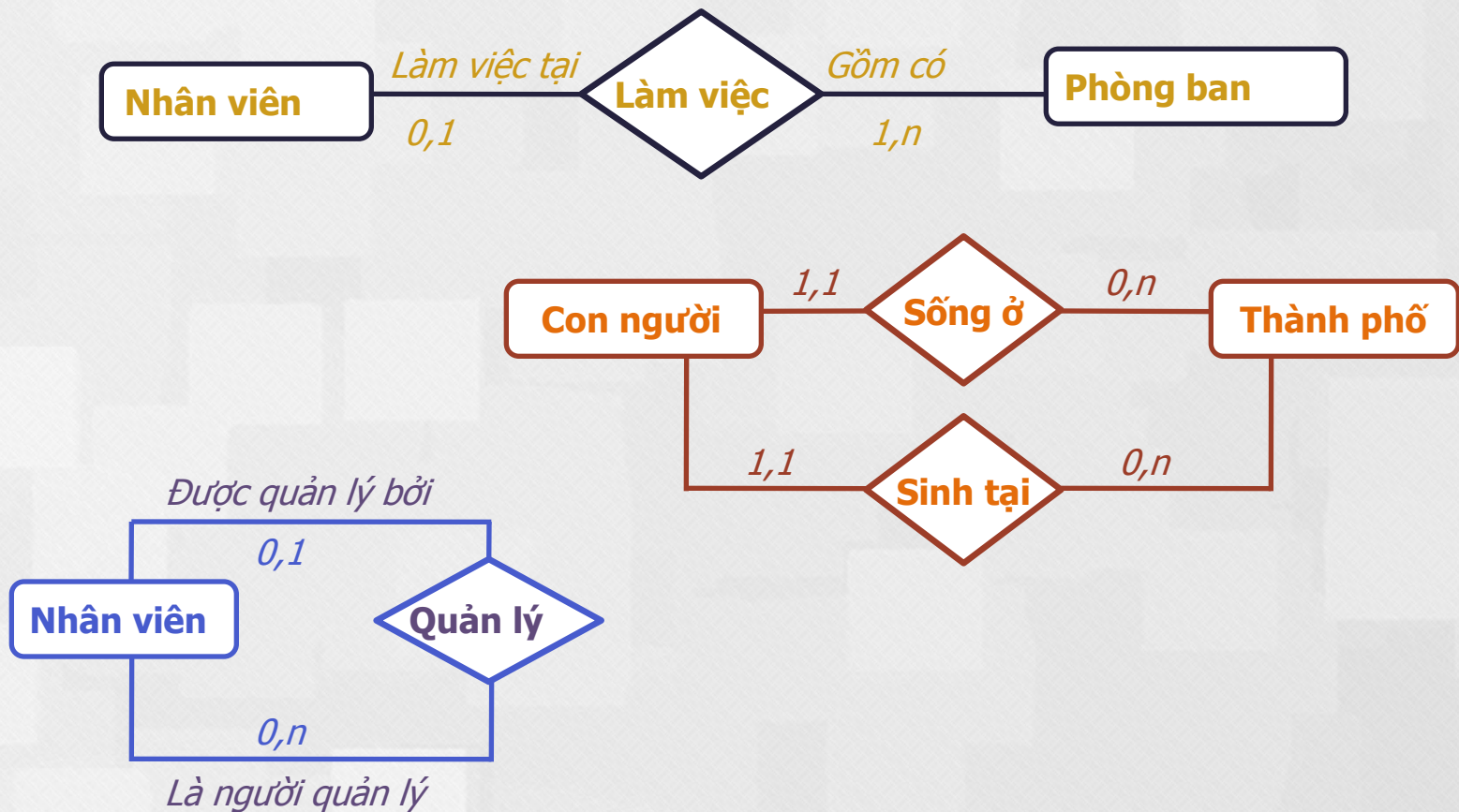
Thông thường tên-mối-kết-hợp sẽ lấy 1 trong các tên-vai-trò (ta bỏ qua tên-vai-trò)

Bảng số

- Ràng buộc về số lượng các thực thể tham gia vào mỗi kết hợp
- Ký hiệu bởi 1 cặp (min, max)
 - Min : qui định giá trị tối thiểu các thực thể khi tham gia vào mỗi kết hợp
 - Giá trị đi từ 0, 1, 2, ... đến k (k là hằng số)
 - Max : qui định giá trị tối đa các thực thể khi tham gia vào mỗi kết hợp
 - Giá trị đi từ 1, 2, ... đến n

Bảng số (tt)

○ Ví dụ



Bảng số (tt)

- Phân loại mỗi kết hợp dựa vào bảng số



- Một – Một

$$\max(E1,R)=1 \ \& \ \max(E2,R)=1$$

- Một – Nhiều

$$\max(E1,R)=1 \ \& \ \max(E2,R)=n$$

- Nhiều – Một

$$\max(E1,R)=n \ \& \ \max(E2,R)=1$$

- Nhiều – Nhiều

$$\max(E1,R)=n \ \& \ \max(E2,R)=n$$

Thể hiện

○ Thể hiện của thực thể

- Sự xuất hiện cụ thể của các phần tử

Khách hàng

Khách hàng A

Khách hàng B

Khách hàng C

...

Giáo viên

Giáo viên 1

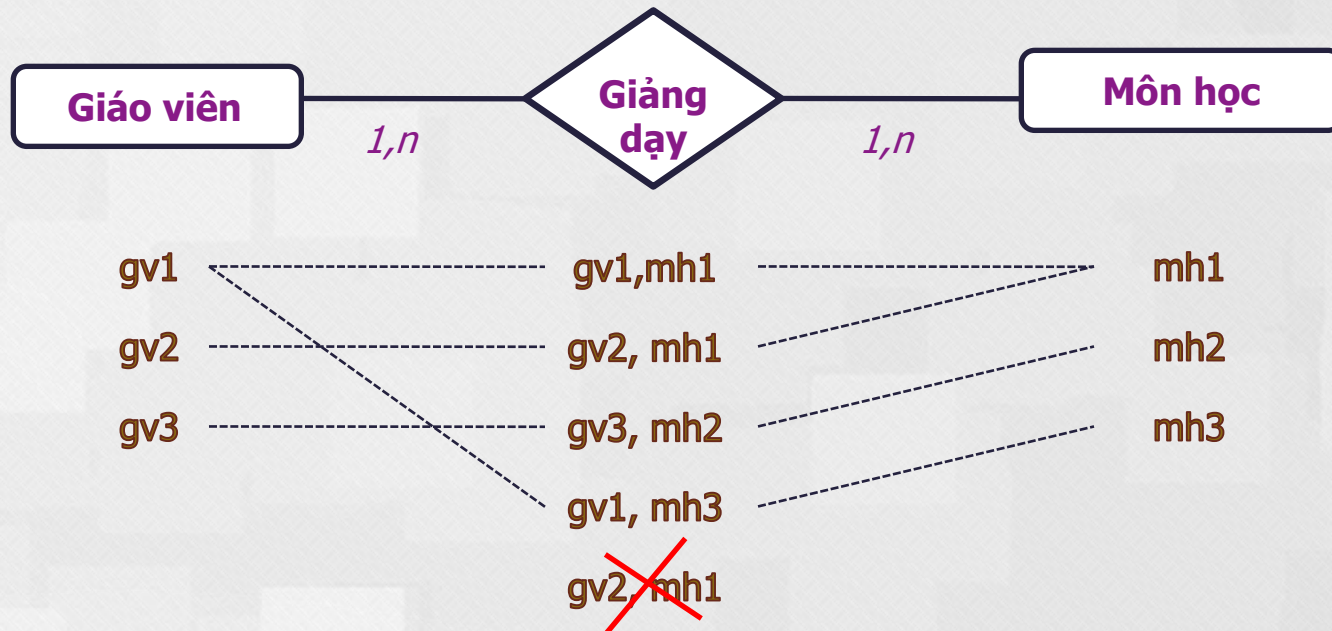
Giáo viên 2

Giáo viên 3

...

Thể hiện (tt)

- Thể hiện của mỗi kết hợp
 - Tổ hợp không trùng lặp các thực thể tham gia vào mỗi kết hợp



Thuộc tính

- Biểu diễn đặc trưng của

- Thực thể

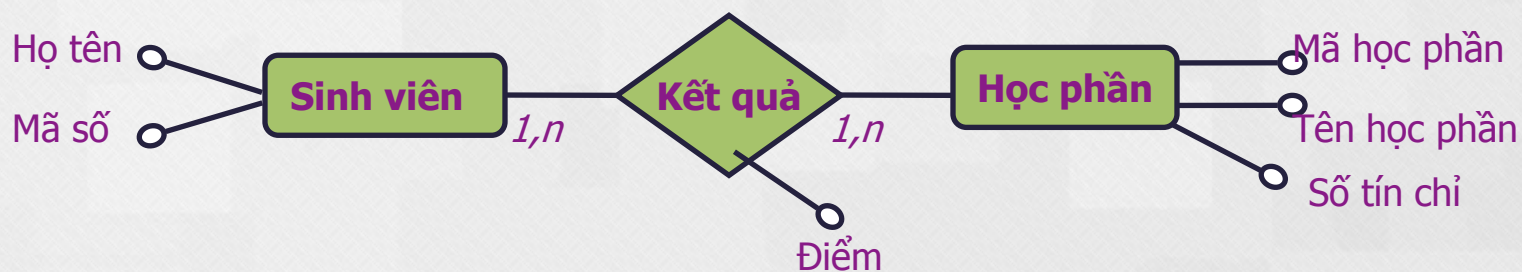
- Mỗi kết hợp

- Ký hiệu

—○ Tên thuộc tính

Thuộc tính (tt)

○ Ví dụ



sv1(A, 1)

sv2(B, 2)

sv3(C, 3)

sv1, hp1, 6

sv2, hp1, 7

sv3, hp2, 8

sv1, hp3, 9

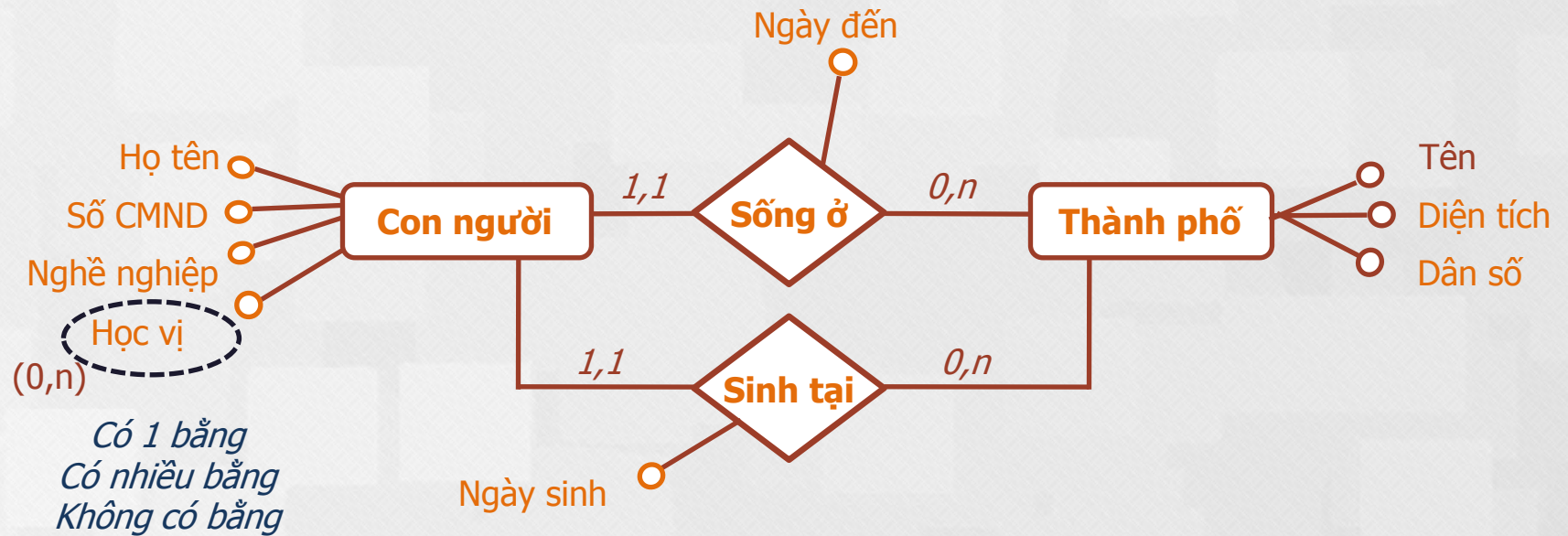
hp1(01, CSDL, 4)

hp2(02, CNPM, 4)

hp3(03, CTDL, 4)

Thuộc tính (tt)

○ Ví dụ



Sử dụng bảng số cho những thuộc tính đa trị

Nội dung trình bày

- Mô hình thực thể kết hợp nguyên thủy
- Mô hình thực thể kết hợp mở rộng
- Phương pháp phân tích dữ liệu
- Quy tắc mô hình hóa quan niệm dữ liệu
- Tiêu chuẩn chọn lựa khái niệm

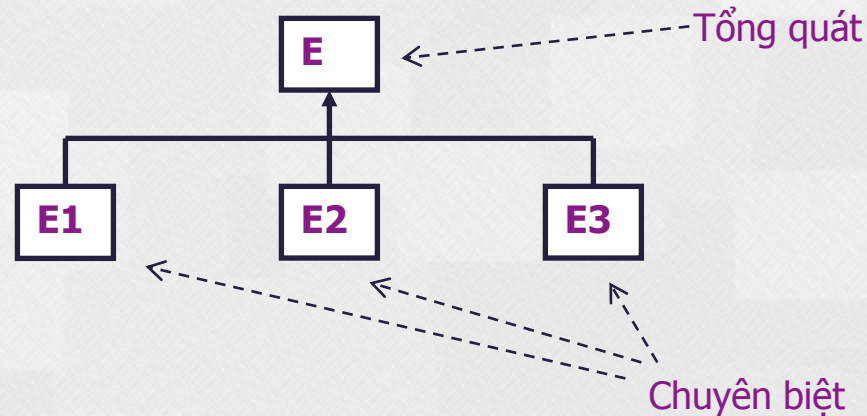
Mô hình thực thể kết hợp mở rộng

◉ Bổ sung vào mô hình thực thể kết hợp

- ◉ Thuộc tính kết hợp
- ◉ Định danh
- ◉ Tổng quát hóa
- ◉ Tập con
- ◉ Mối kết hợp mở rộng

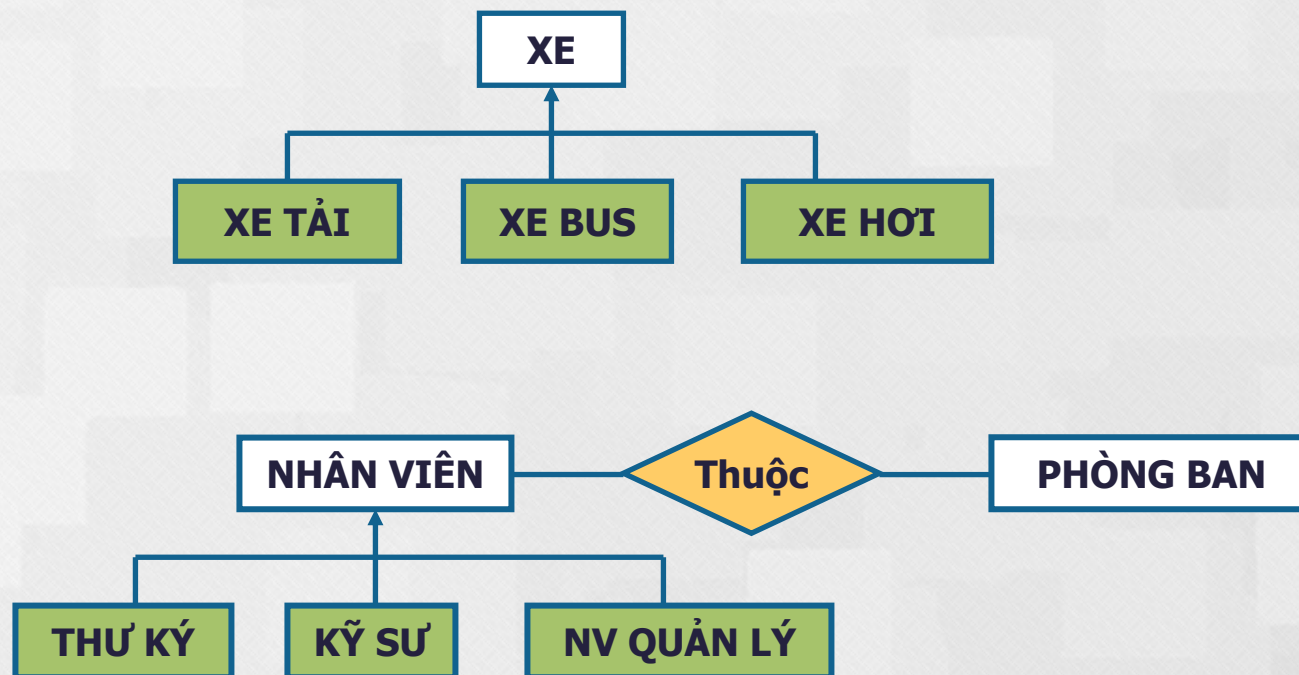
Cấu trúc phân cấp

- Thiết lập cấu trúc cây phân cấp giữa các thực thể

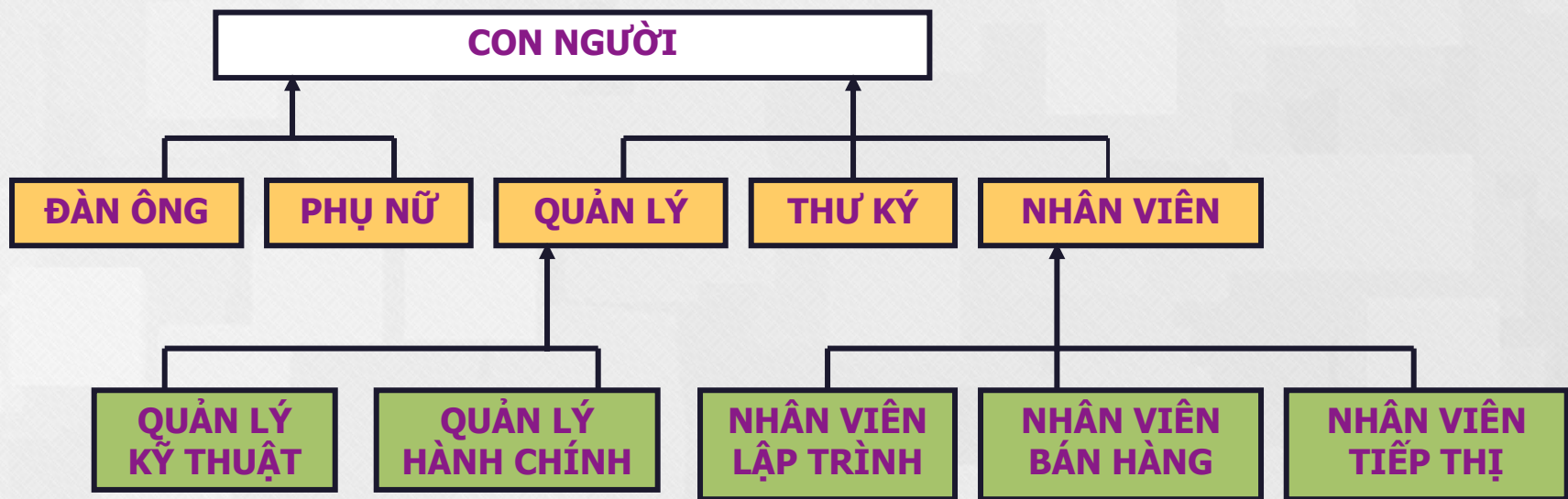


- E là một tổng quát hóa của 1 nhóm thực thể E1, E2, E3 khi
 - Mỗi đối tượng của lớp E1, E2, E3 cũng là đối tượng của lớp E

Ví dụ



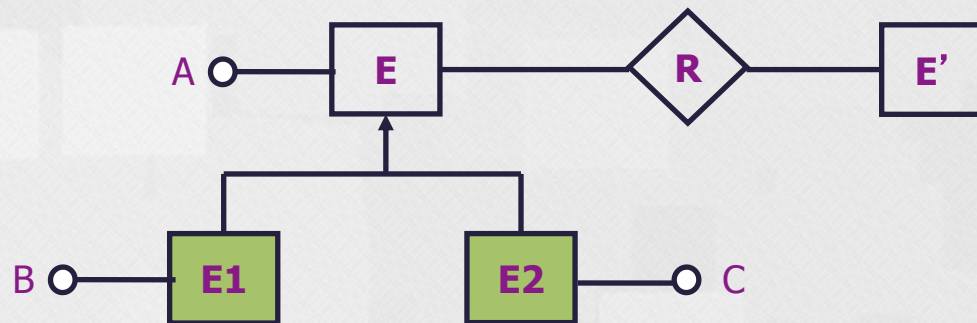
Ví dụ



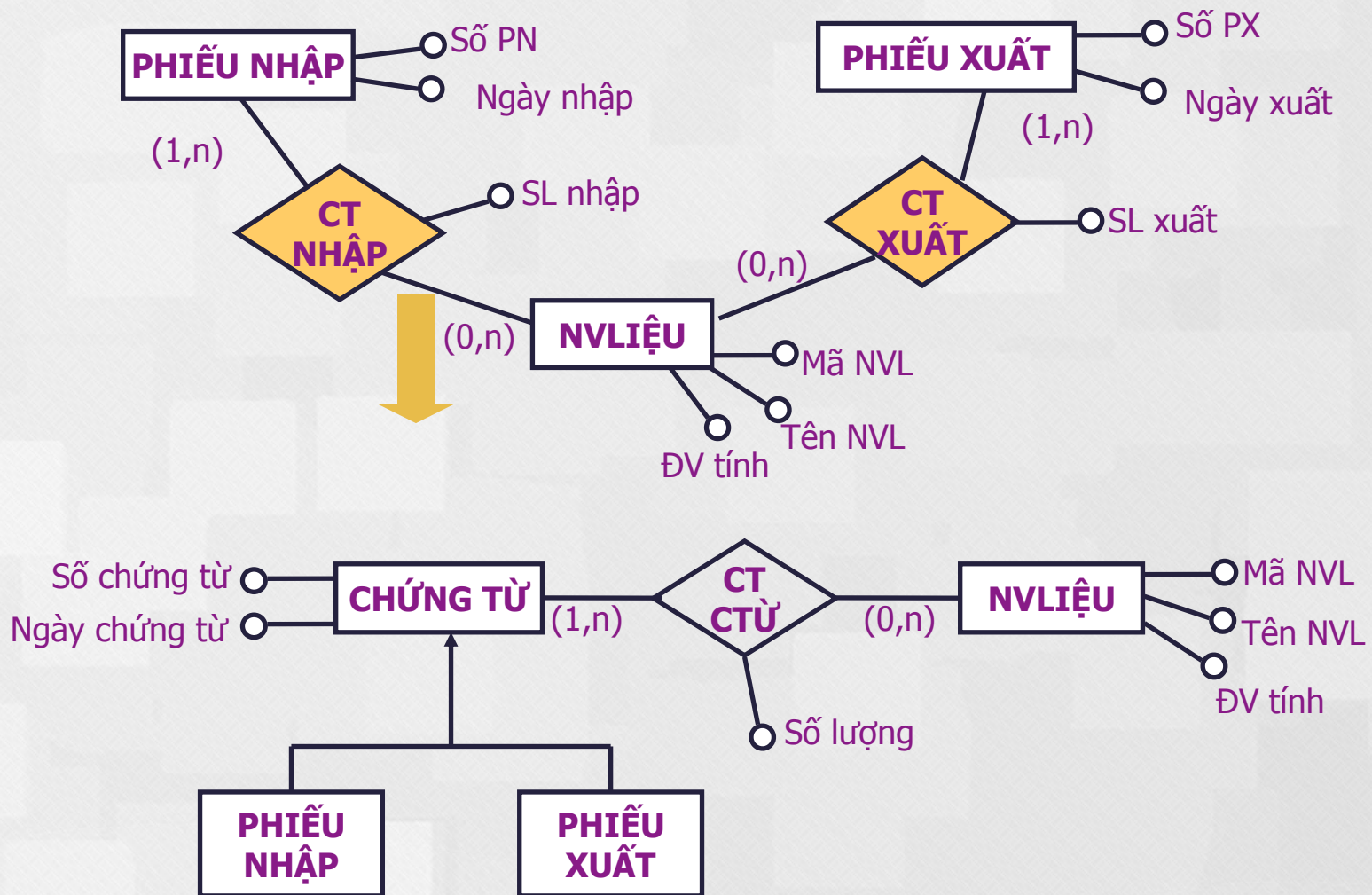
Cấu trúc phân cấp (tt)

○ Tính kế thừa

- Thực thể chuyên biệt kế thừa thuộc tính và mối kết hợp của thực thể tổng quát



Cấu trúc phân cấp (tt)



Cấu trúc phân cấp (tt)

- Nguyên tắc

- Xây dựng thực-thể-chuyên-biệt khi xác định được đặc trưng riêng của nó
- Xây dựng thực thể tổng quát khi xác định được các đặc trưng chung của các thực thể

- Tái sử dụng

Cấu trúc phân cấp (tt)

- Tính bao phủ

- Sự tương quan giữa thực-thể-tổng-quát và thực-thể-chuyên-biệt
- Gồm
 - Toàn phần (t-total)
 - Bán phần (p-partial)
 - Riêng biệt (e-exclusive)
 - Chồng chéo (o-overlapping)

Cấu trúc phân cấp (tt)

○ Toàn phần

- Tất cả các phần tử của các thực thể chuyên biệt phủ toàn bộ tập phần tử của thực thể tổng quát

○ Bán phần

- Các phần tử của các thực thể chuyên biệt không phủ toàn bộ tập phần tử của thực thể tổng quát

Cấu trúc phân cấp (tt)

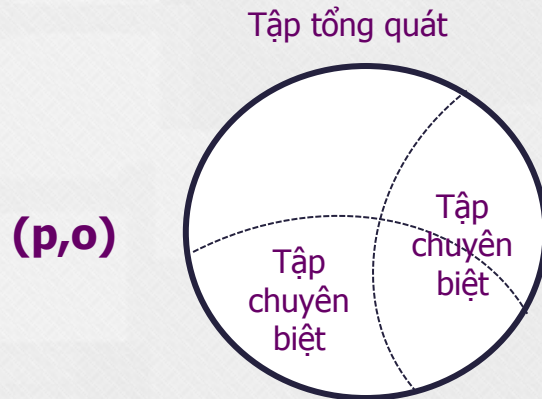
- Riêng biệt

- Phần tử của thực thể chuyên biệt này không là phần tử của thực thể chuyên biệt khác

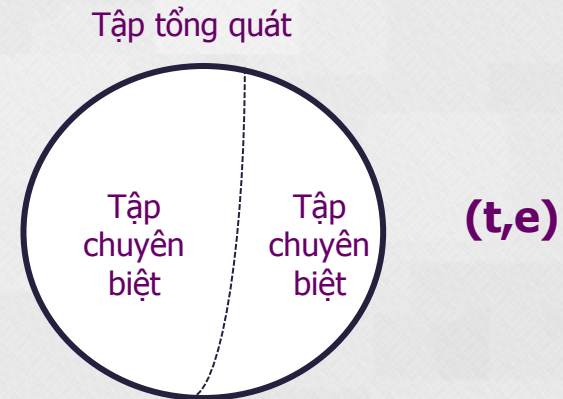
- Chồng chéo

- Phần tử của thực thể chuyên biệt này có thể là phần tử của thực thể chuyên biệt khác

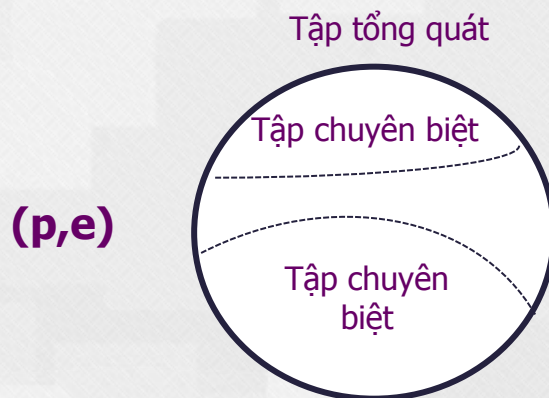
Cấu trúc phân cấp (tt)



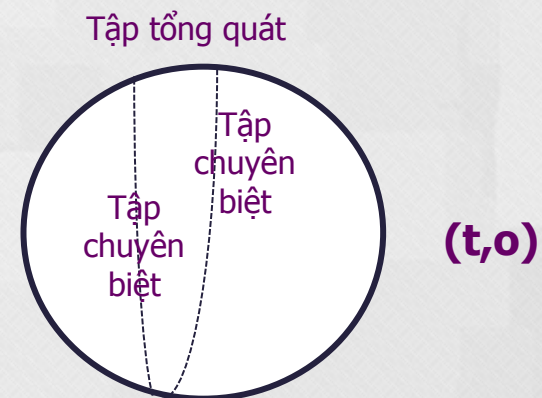
Chuyên biệt bán phần, chồng lấp



Chuyên biệt toàn phần, riêng biệt

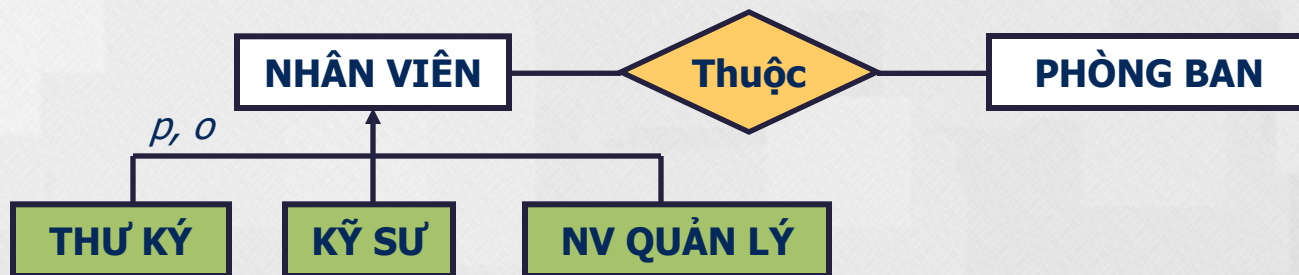
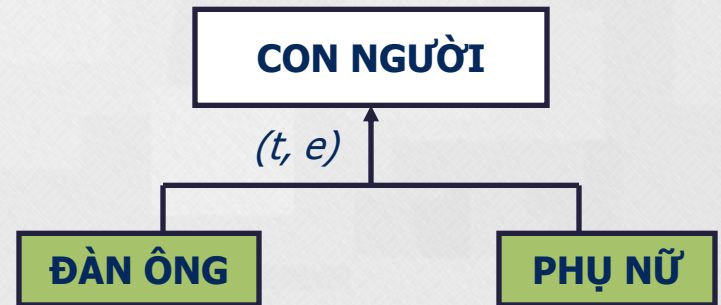
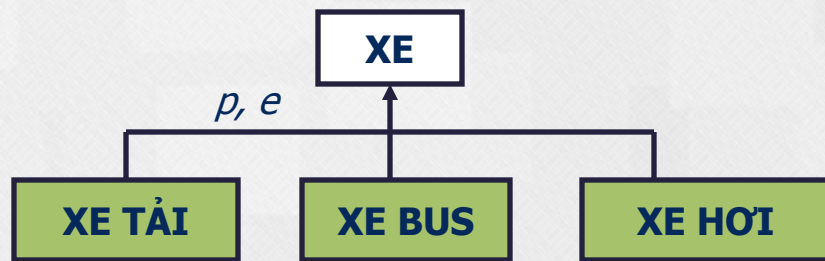


Chuyên biệt bán phần, riêng biệt

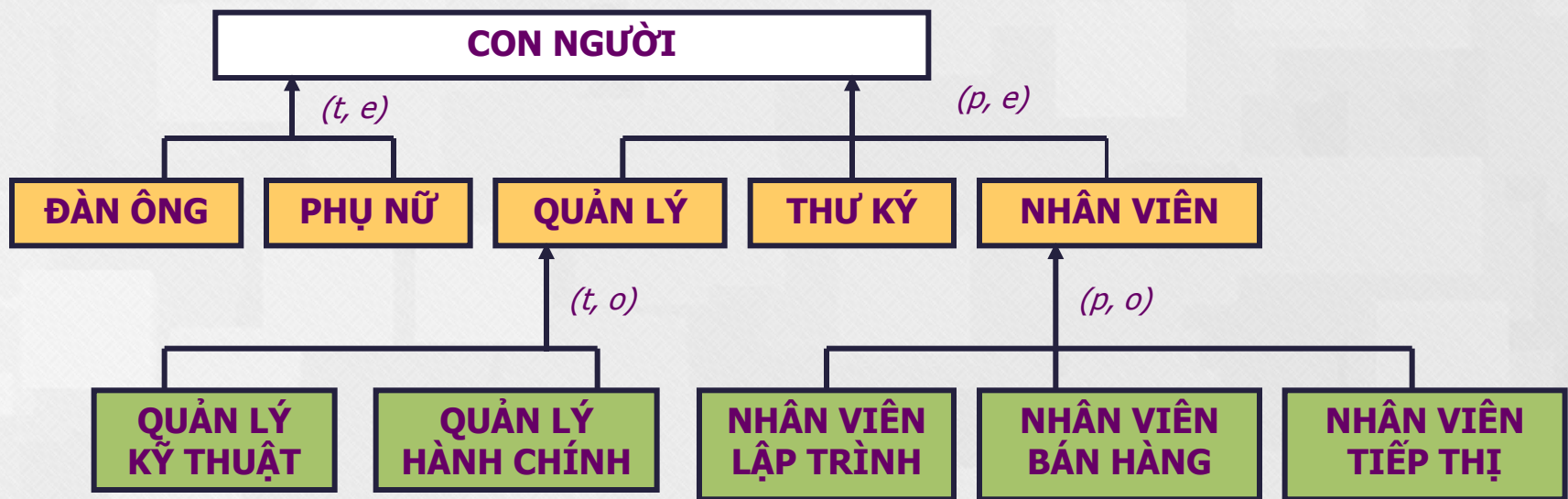


Chuyên biệt toàn phần, chồng lấp

Ví dụ

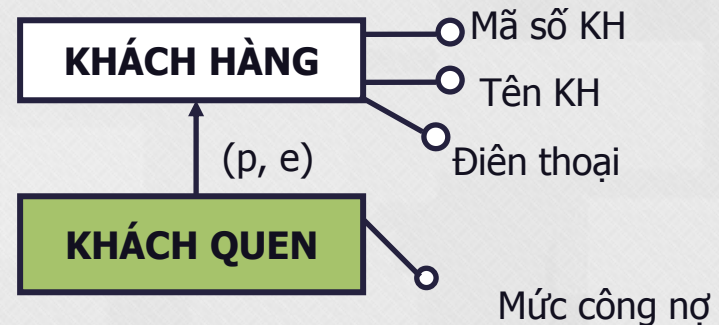
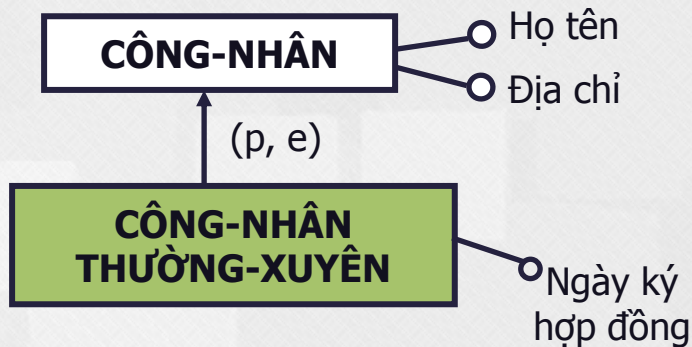


Ví dụ



Tập con

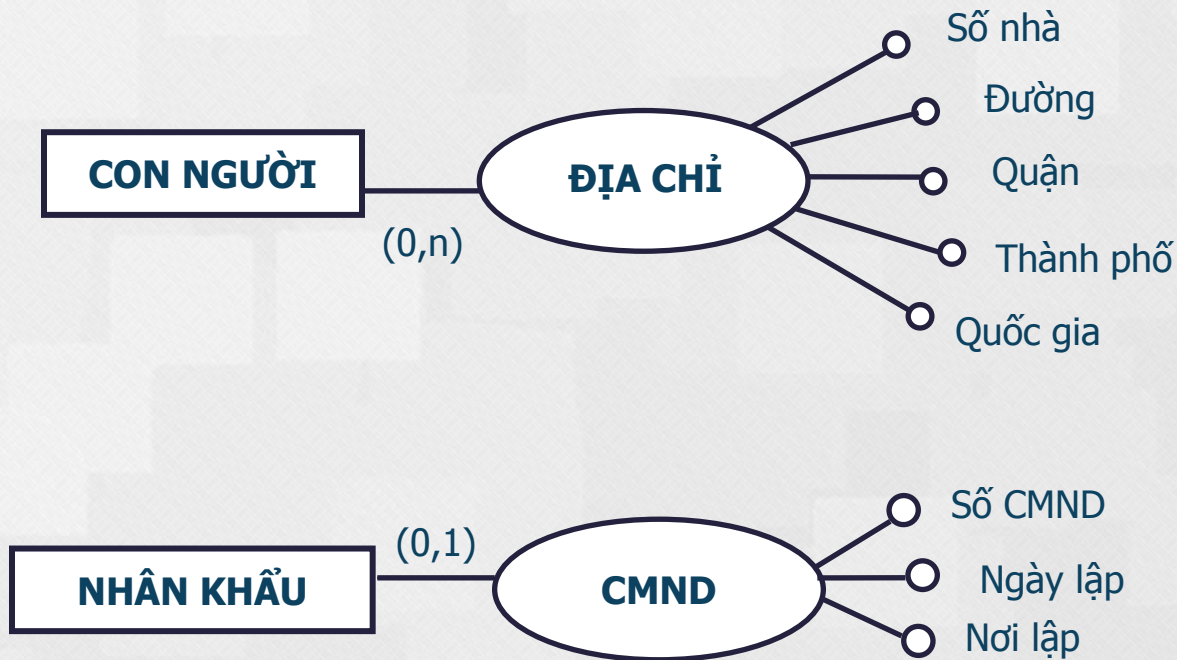
- Là trường hợp đặc biệt của tổng quát hóa
 - Chỉ có 1 thực thể chuyên biệt



Sự tương quan luôn là bán phần & riêng biệt

Thuộc tính kết hợp

- Là 1 nhóm các thuộc tính có liên hệ



Định danh

- Tập hợp các thuộc tính có tính chất đặc trưng duy nhất nhằm phân biệt tất cả các thể hiện của thực thể
- Ký hiệu



Định danh 1 thuộc tính



Định danh 2 thuộc tính trở lên

Định danh (tt)

- Gọi I là tập hợp

- $I = \{A_1, A_2, \dots, A_n, E_1, E_2, \dots, E_m\}$

- $n \geq 0, m \geq 0, n+m \geq 1$

- I là định danh

- (1) Không có 2 thể hiện của E có chung 1 giá trị định danh

- (2) Nếu bỏ bớt đi 1 A_i hay E_i thì (1) không còn đúng

Định danh (tt)

○ Phân loại

- Đơn giản ($m+n=1$)
- Bên trong ($m=0$)
- Bên ngoài ($n=0$)
- Phối hợp ($m>0$ & $n>0$)

Ví dụ

Định danh bên trong



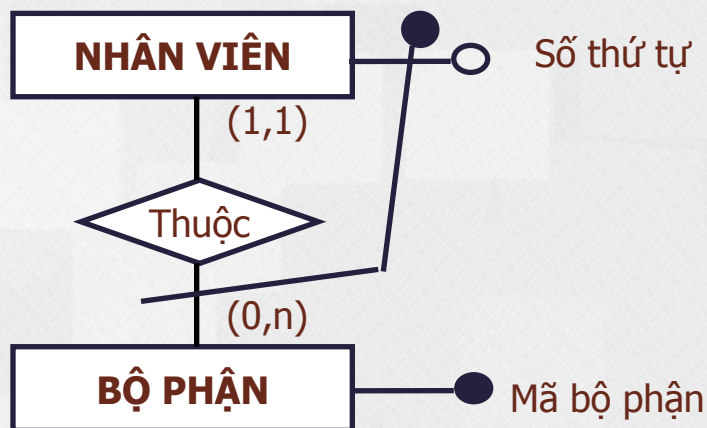
$I = \{\text{Số CMND}\}$

Định danh bên trong



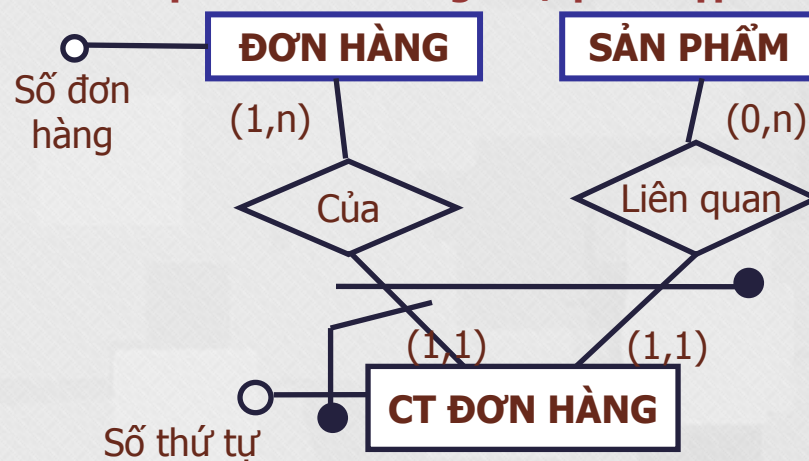
$I = \{\text{Họ tên, Ngày sinh}\}$ Định danh con người

Định danh phối hợp



$INV = \{\text{Số thứ tự, BỘ PHẬN}\}$

Định danh bên ngoài / phối hợp

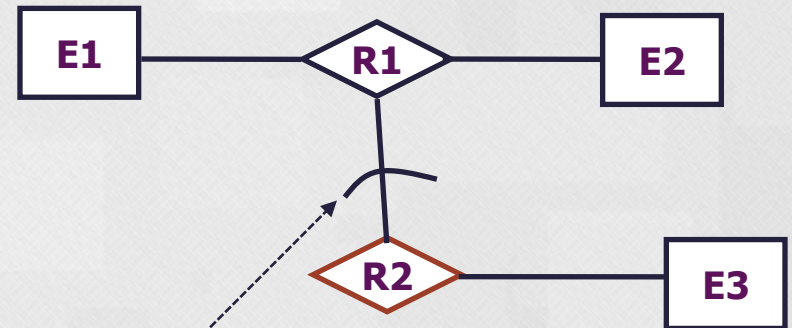
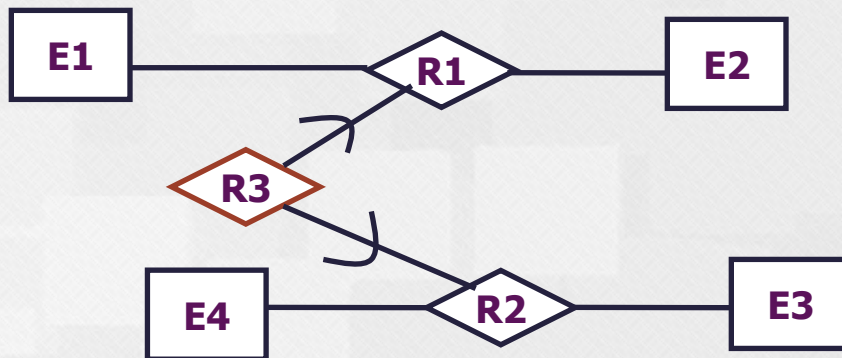


$I_{1CTDH} = \{\text{ĐƠN HÀNG, SẢN PHẨM}\}$

$I_{2CTDH} = \{\text{ĐƠN HÀNG, Số thứ tự}\}$

Mối kết hợp mở rộng

- Là mối kết hợp được định nghĩa trên ít nhất 1 mối kết hợp khác



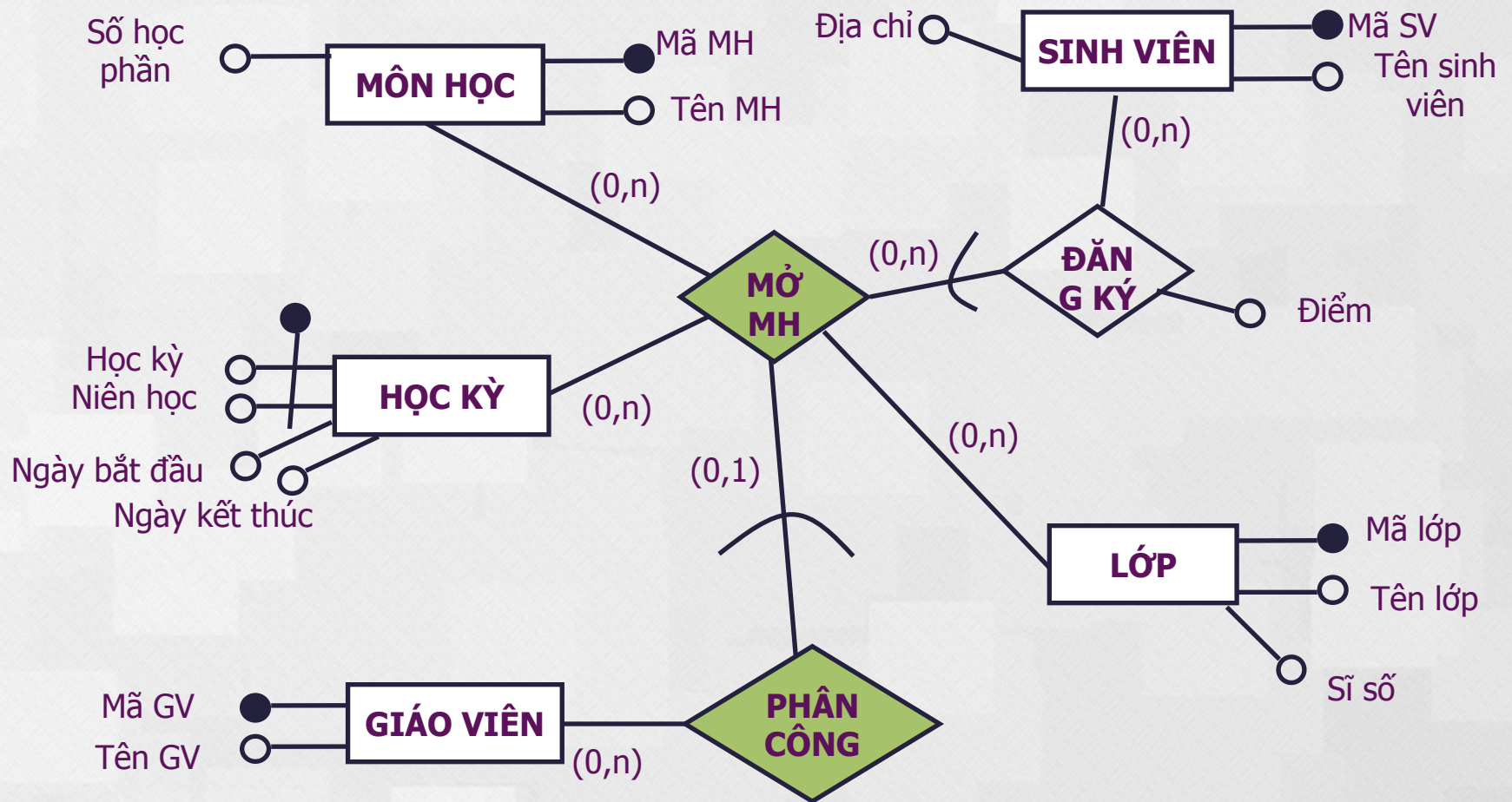
Cung định hướng cho biết
R2 định nghĩa trên R1

Bài tập

○ Bài toán Quản Lý Giáo Vụ

- Lưu trữ các thông tin về sinh viên, môn học, giáo viên, lớp học và các học kỳ trong từng niên khóa
- Lập danh sách mở các môn học cho một lớp trong một học kỳ
- Phân công giảng dạy môn học được mở cho một giáo viên
- Lưu thông tin đăng ký môn học của sinh viên trên môn học được mở
- Ghi nhận điểm kết quả học tập của sinh viên

Bài tập



Nội dung trình bày

- Mô hình thực thể kết hợp nguyên thủy
- Mô hình thực thể kết hợp mở rộng
- Phương pháp phân tích dữ liệu
- Quy tắc mô hình hóa quan niệm dữ liệu
- Tiêu chuẩn chọn lựa khái niệm

Phương pháp phân tích dữ liệu

- Luật căn bản
 - Dùng để tinh chế lược đồ quan niệm
 - Tập hữu hạn các chuyển đổi cần áp dụng cho lược đồ ban đầu để tạo ra lược đồ cuối cùng
 - Được chia thành
 - Luật căn bản trên xuống
 - Luật căn bản dưới lên
- Chiến lược phân tích lược đồ

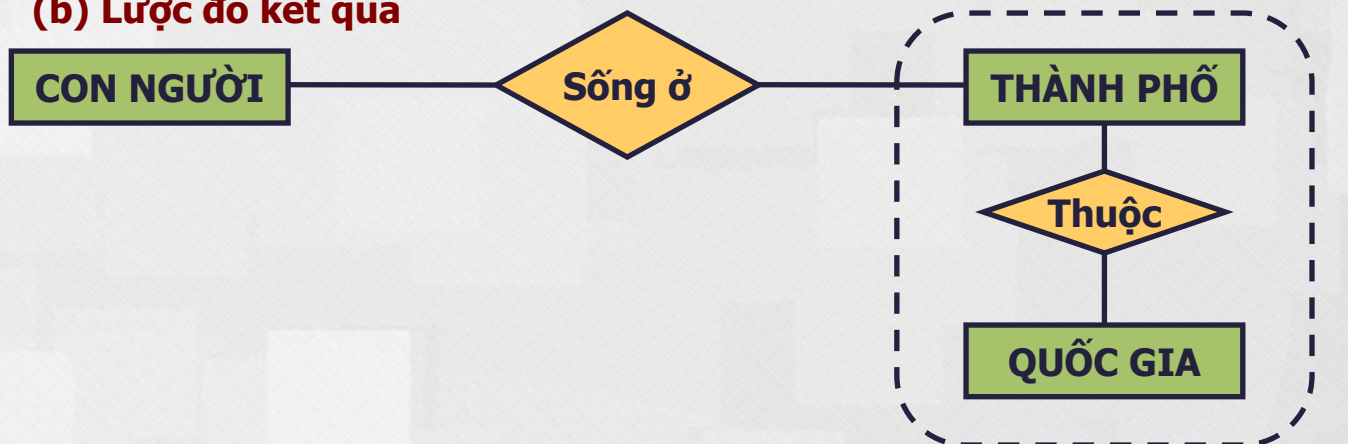
Luật căn bản

○ Xét ví dụ

(a) Lược đồ khởi điểm



(b) Lược đồ kết quả



Luật căn bản (tt)




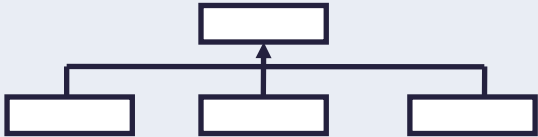



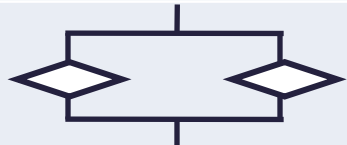








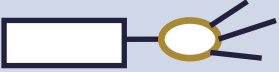



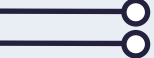
- Lược đồ ở mức quan niệm được tạo ra sau 1 quá trình xử lý lặp đi lặp lại
 - Bắt đầu bằng phiên bản chưa hoàn chỉnh
 - Thực hiện các chuyển đổi
 - Tạo ra phiên bản cuối cùng

Luật căn bản (tt)

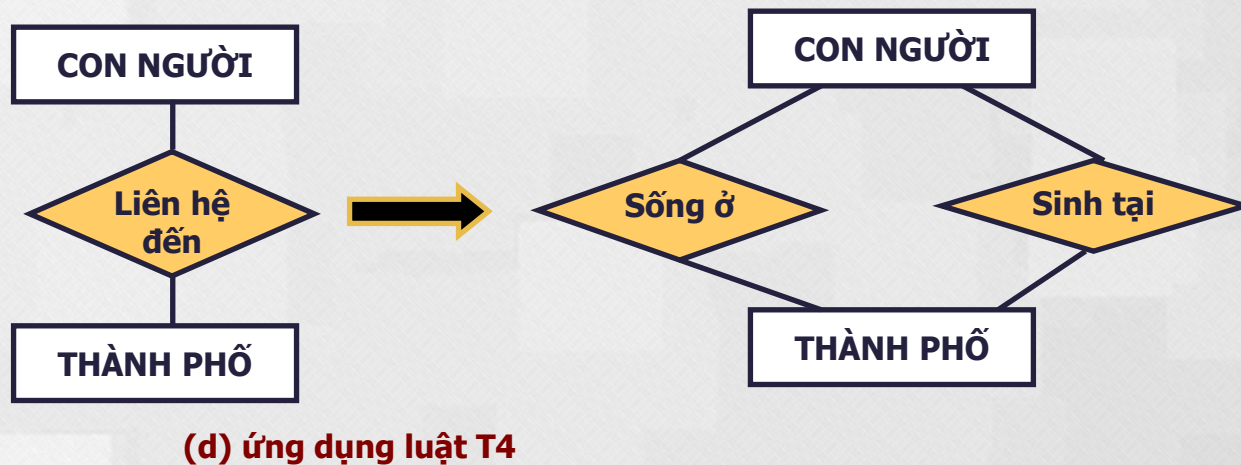
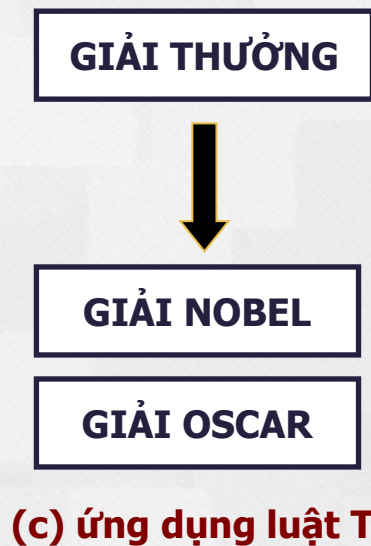
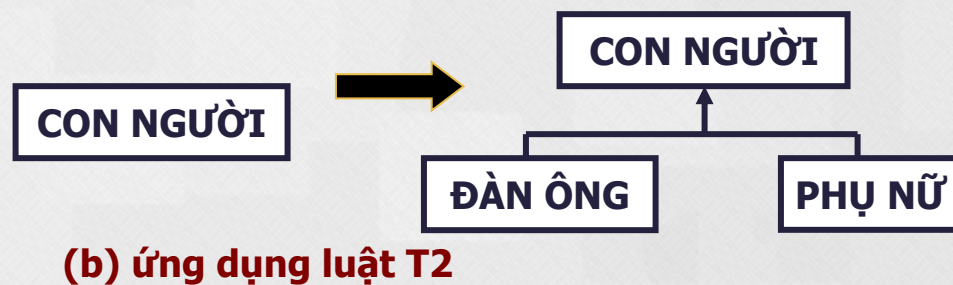
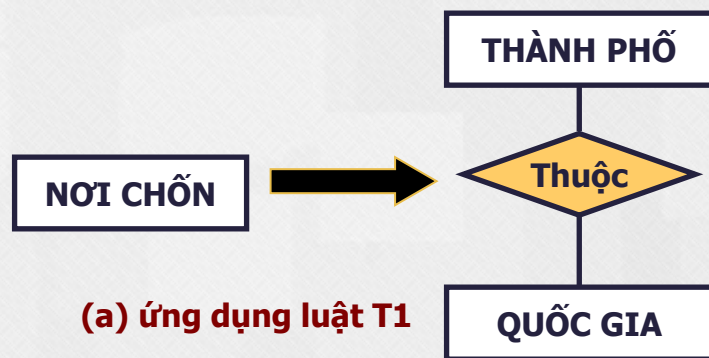
○ Tính chất của chuyển đổi

- Luôn phải có 1 lược đồ khởi điểm, 1 lược đồ kết quả
- Ánh xạ tên giữa 2 lược đồ
- Các khái niệm trong lược đồ kết quả phải kế thừa tất cả các kết nối đã xác định trong lược đồ khởi điểm

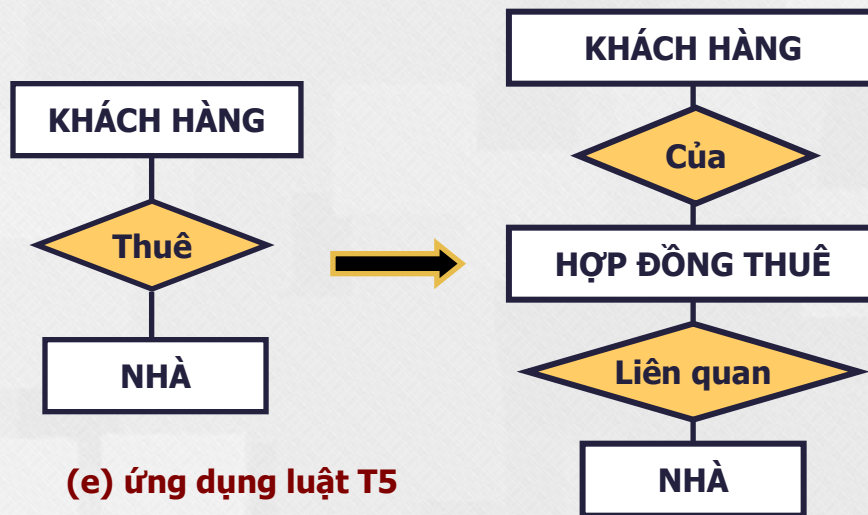
Luật căn bản trên xuống

STT	Luật căn bản	Lược đồ khởi điểm	Lược đồ kết quả
T1	Thực thể → mỗi kết hợp giữa 1 hay nhiều thực thể		
T2	Thực thể → tổng quát hóa Thực thể → tập con		
T3	Thực thể → Các thực thể không có mối quan hệ		
T4	Mối kết hợp → Mỗi kết hợp song song		
T5	Mối kết hợp → Thực thể và các mối kết hợp		
T6	Phát triển thêm thuộc tính	 	 
T7	Phát triển thêm thuộc tính kết hợp	 	 
T8	Tinh chế thuộc tính		 

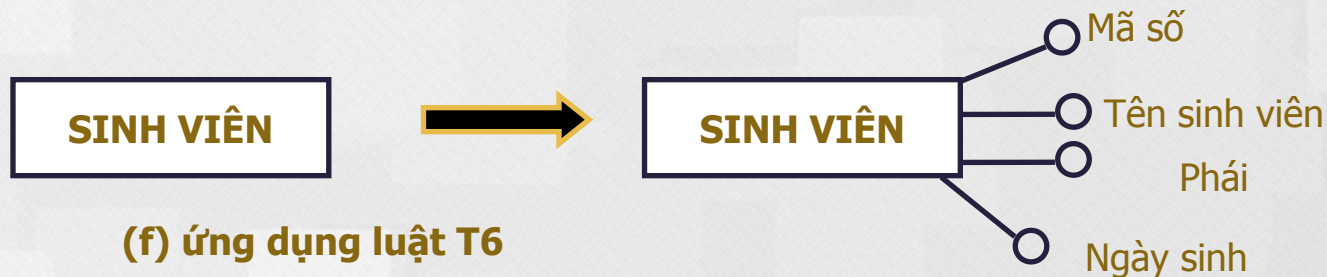
Ví dụ



Ví dụ (tt)



(e) ứng dụng luật T5

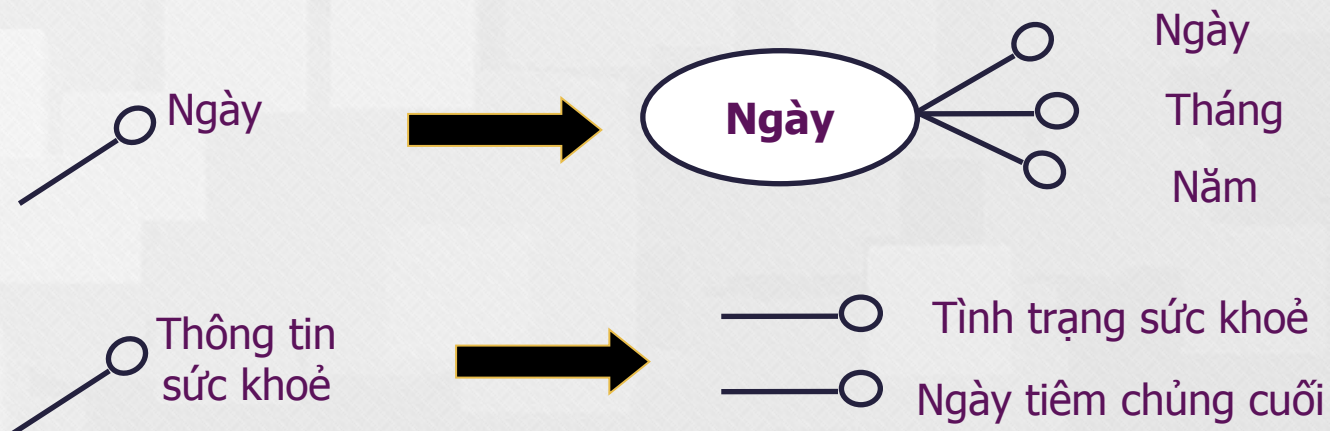


(f) ứng dụng luật T6

Ví dụ (tt)




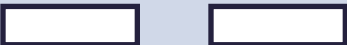
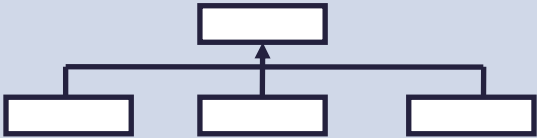



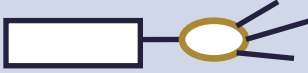


(g) ứng dụng luật T7

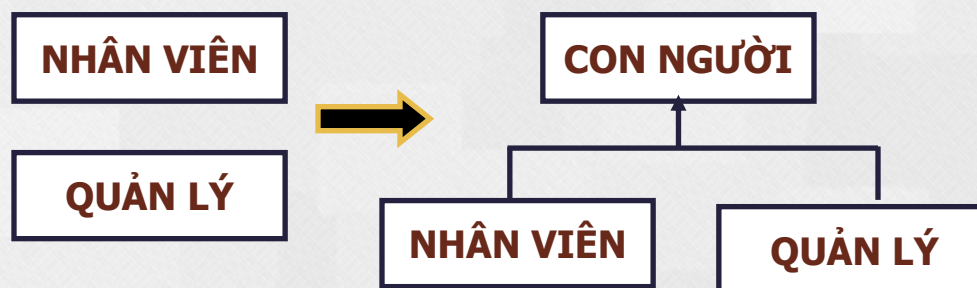


(h) ứng dụng luật T8

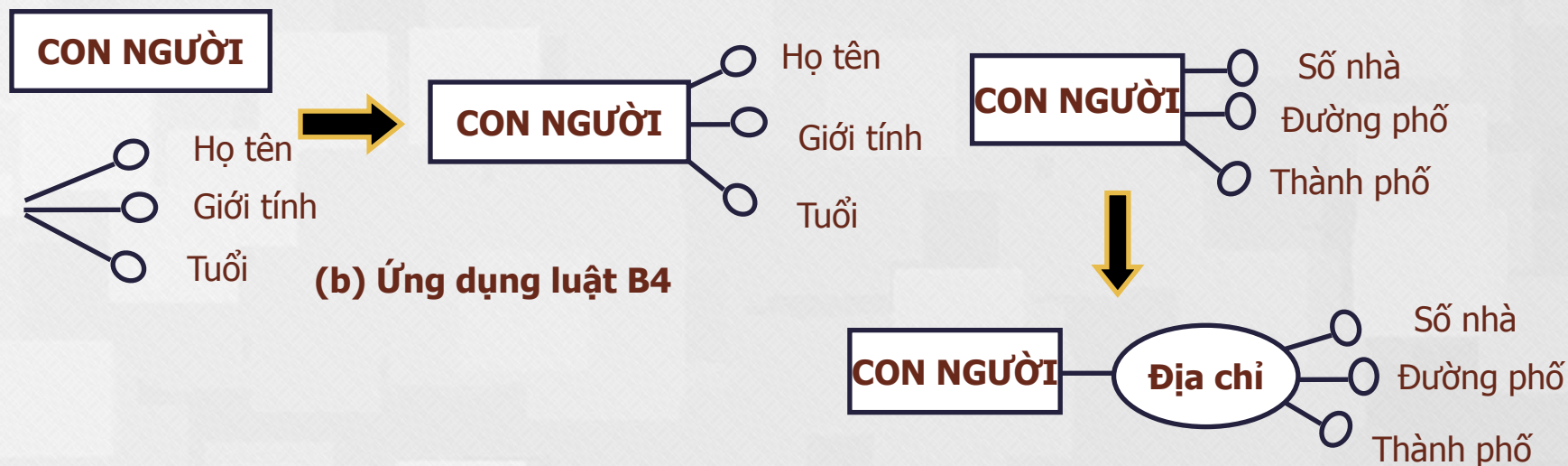
Luật căn bản dưới lên

STT	Luật căn bản	Lược đồ khởi điểm	Lược đồ kết quả
B1	Giai đoạn tạo thực thể		
B2	Giai đoạn tạo mối kết hợp		
B3	Giai đoạn tổng quát hóa (tập con)		
B4	Cấu trúc các thuộc tính		
B5	Cấu trúc các thuộc tính kết hợp		

Ví dụ (tt)



(a) Ứng dụng luật B3



(b) Ứng dụng luật B4

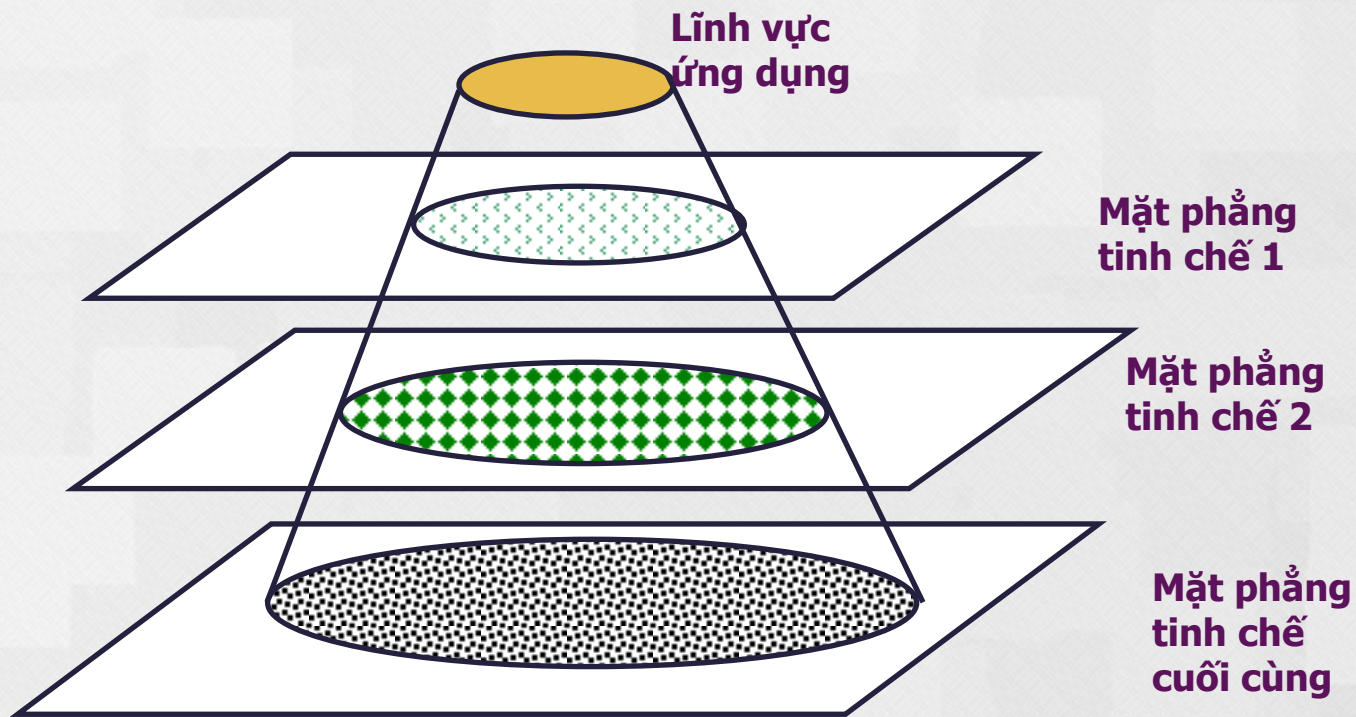


(c) Ứng dụng luật B5

Phương pháp phân tích dữ liệu

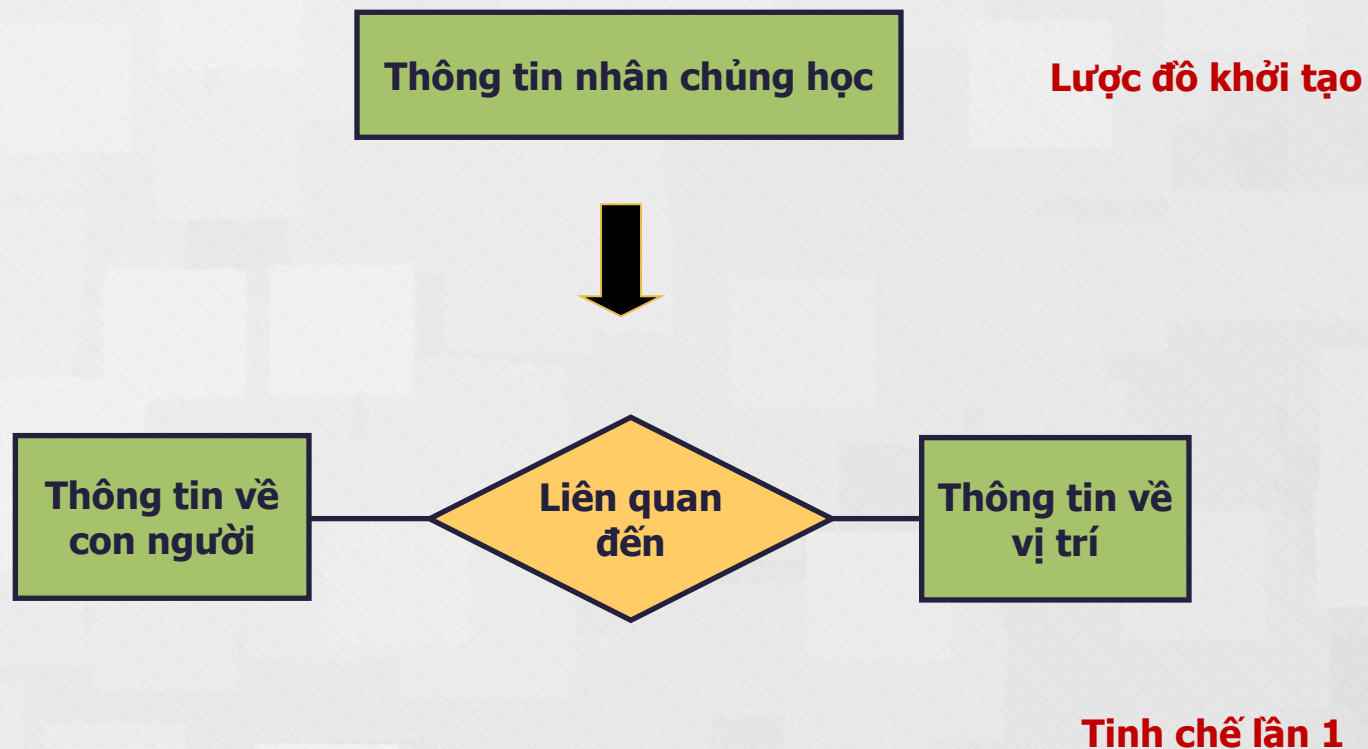
- Luật căn bản
- Chiến lược phân tích lược đồ
 - Trên xuống (top-down)
 - Dưới lên (bottom-up)
 - Phối hợp (mixed)
 - Trong ra ngoài (inside-out)

Chiến lược trên xuống



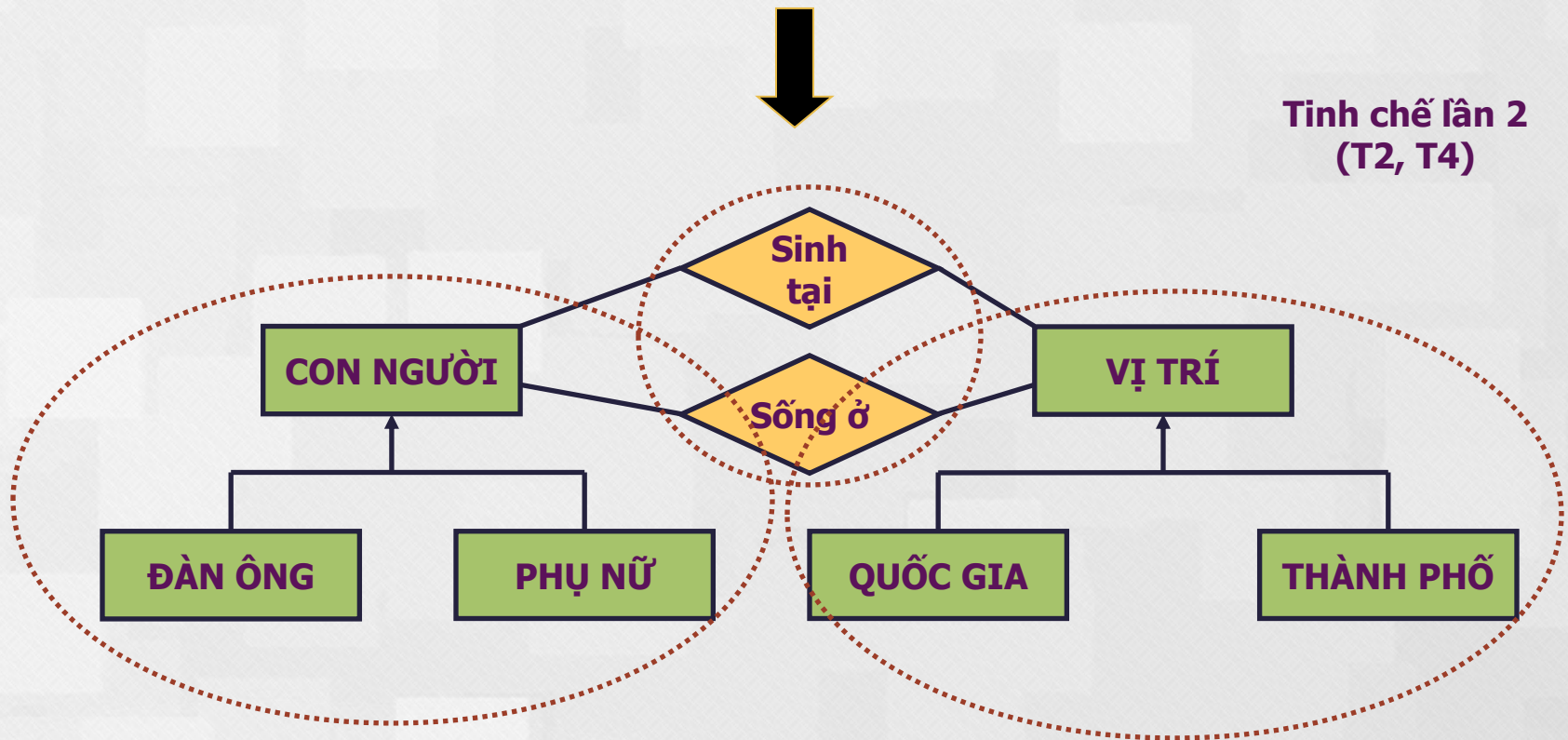
Ví dụ

○ Ứng dụng Quản lý nhân chủng học



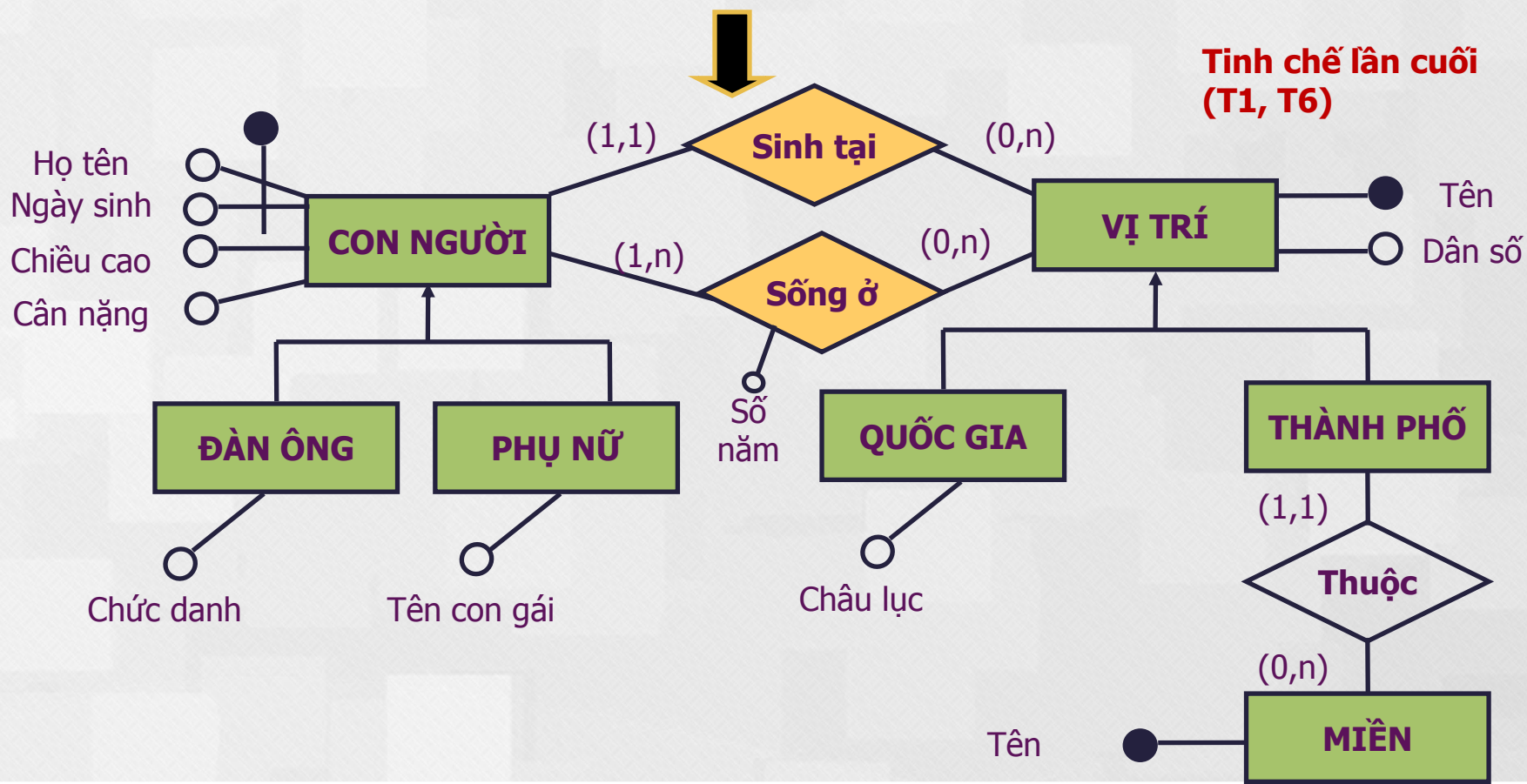
Ví dụ

Ứng dụng Quản lý nhân chủng học



Ví dụ

Ứng dụng Quản lý nhân chủng học

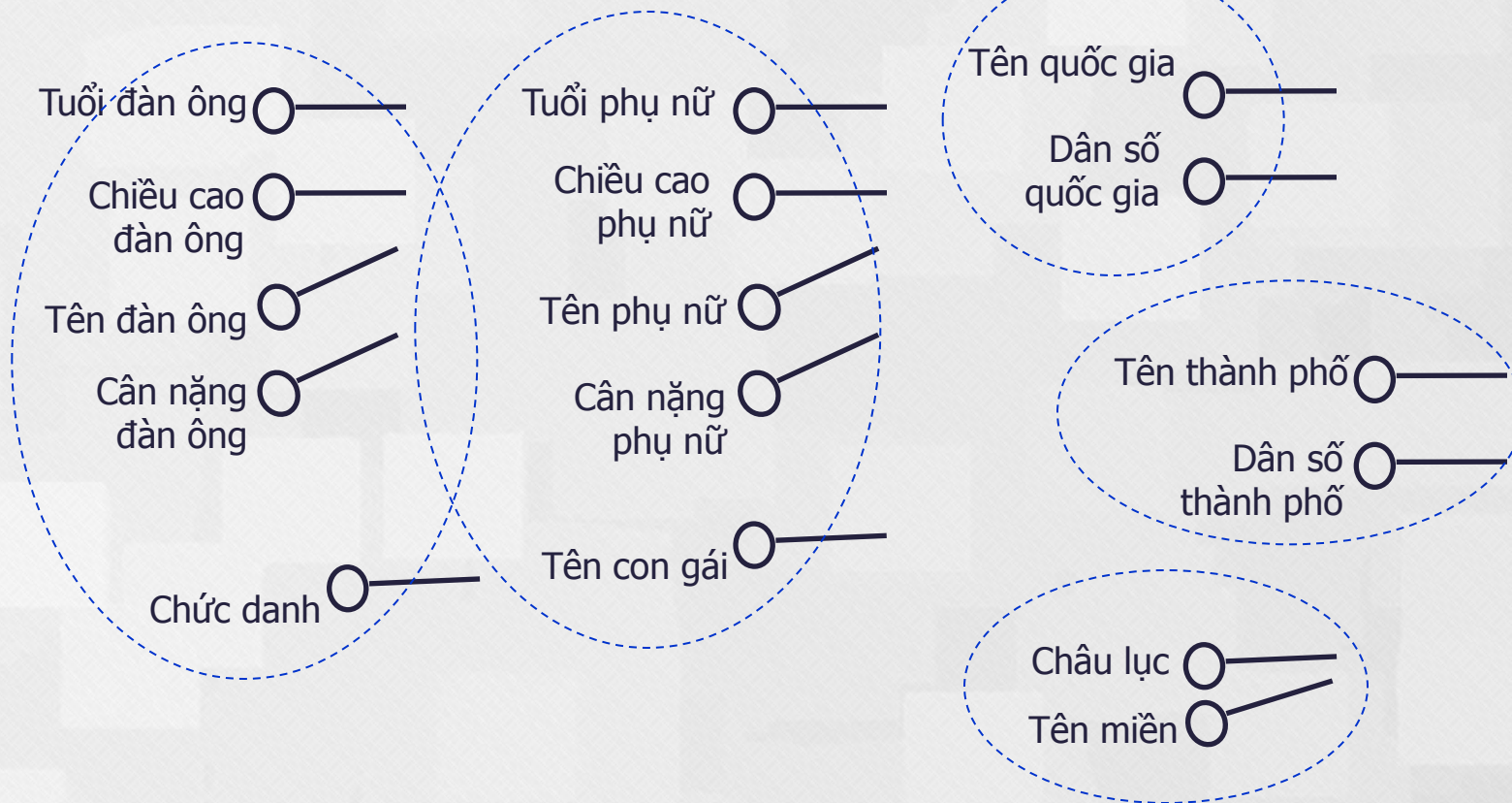


Chiến lược dưới lên



Ví dụ

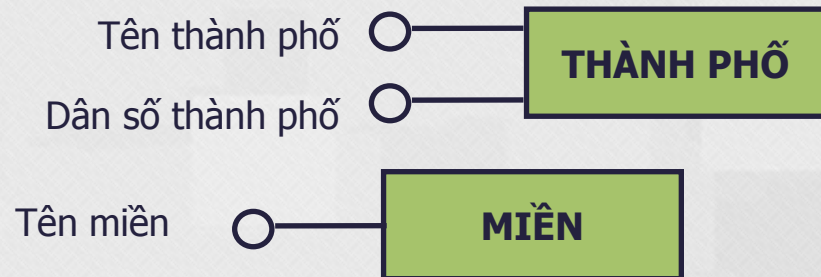
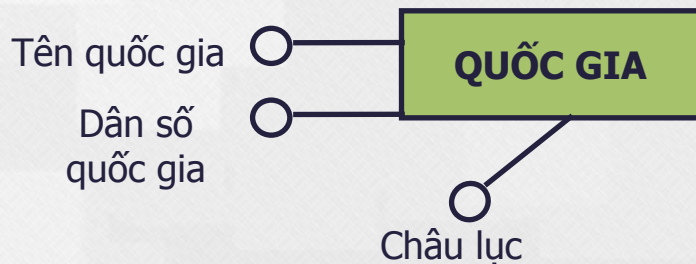
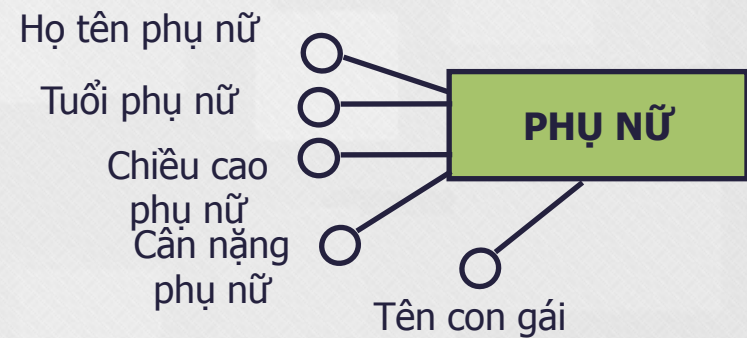
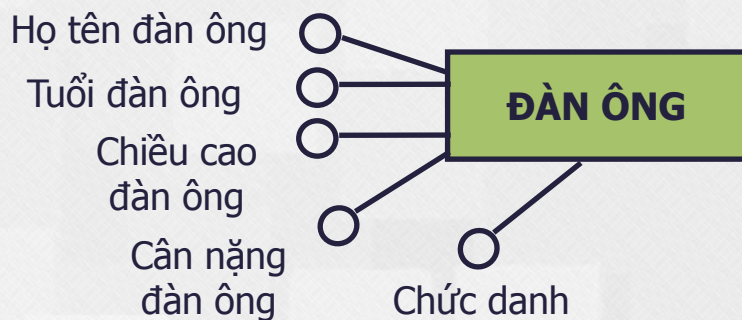
Ứng dụng Quản lý nhân chủng học



Thu thập các đặc trưng của các thành phần trong hệ thống

Ví dụ

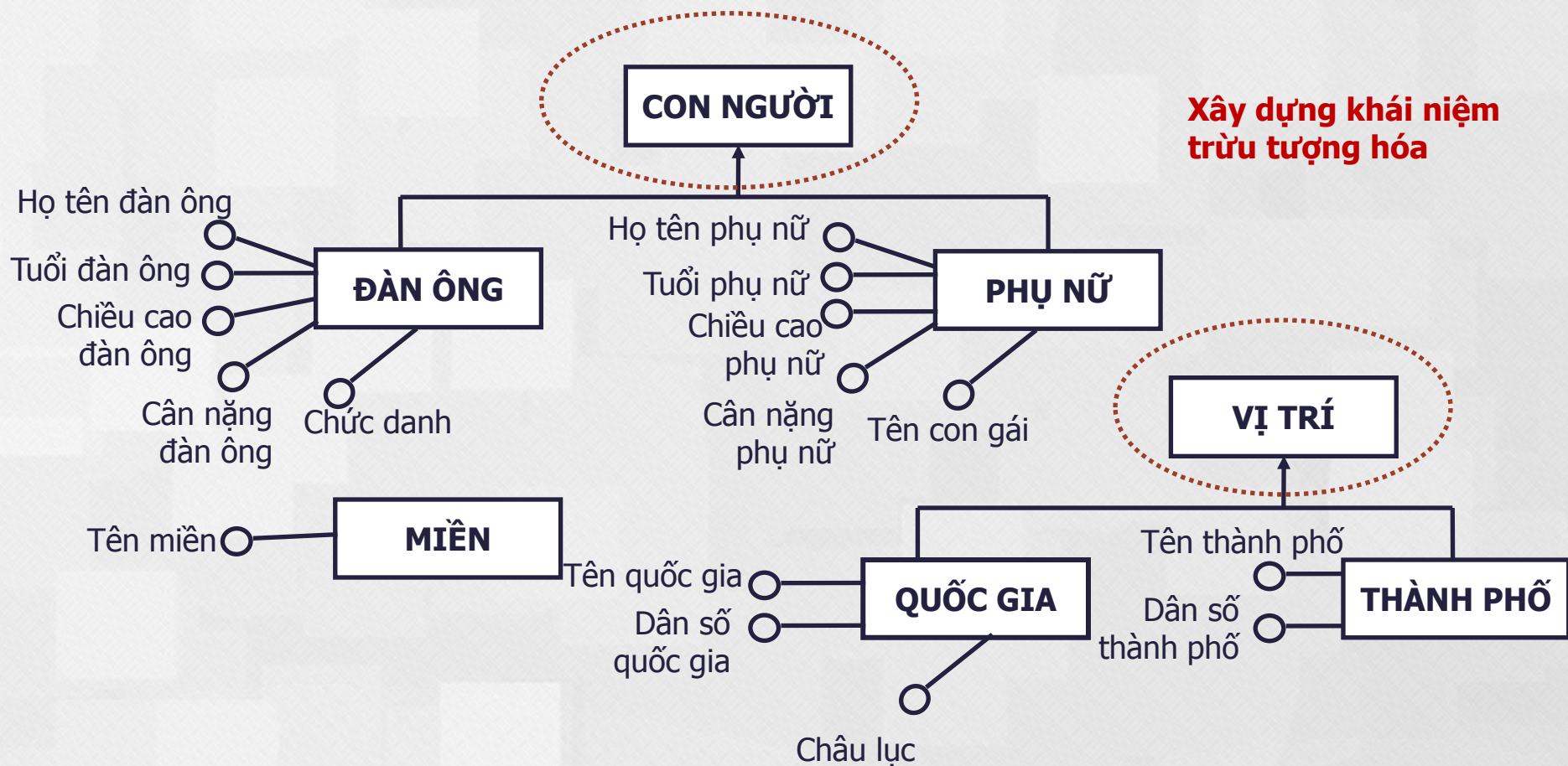
Ứng dụng Quản lý nhân chủng học



Kết hợp các đặc trưng để hình thành thực thể

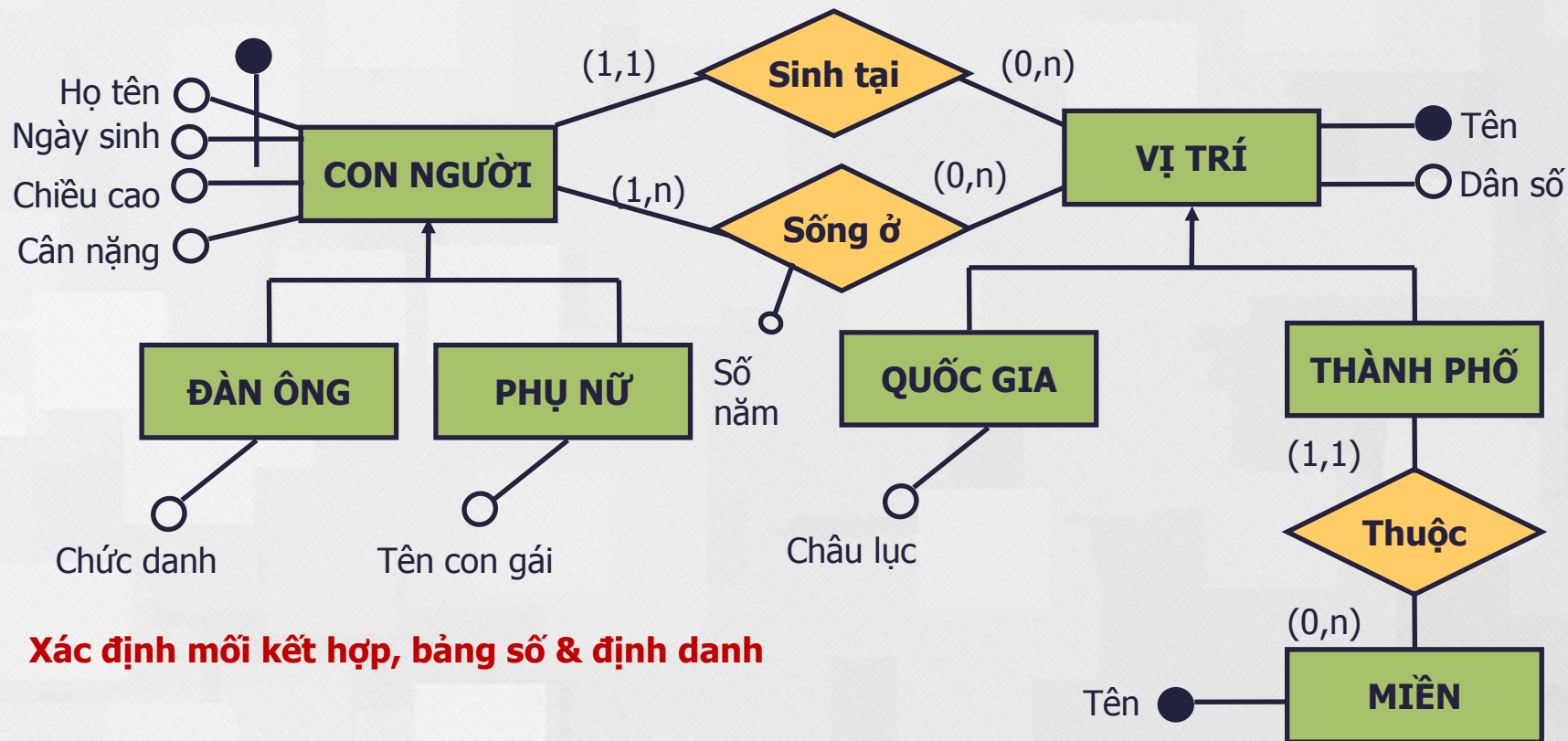
Ví dụ

Ứng dụng Quản lý nhân chủng học



Ví dụ

Ứng dụng Quản lý nhân chủng học



Chiến lược trong ra ngoài



Chọn lọc các khái niệm quan trọng nhất

Các thực thể quan trọng và nổi bật

Lược đồ khởi điểm

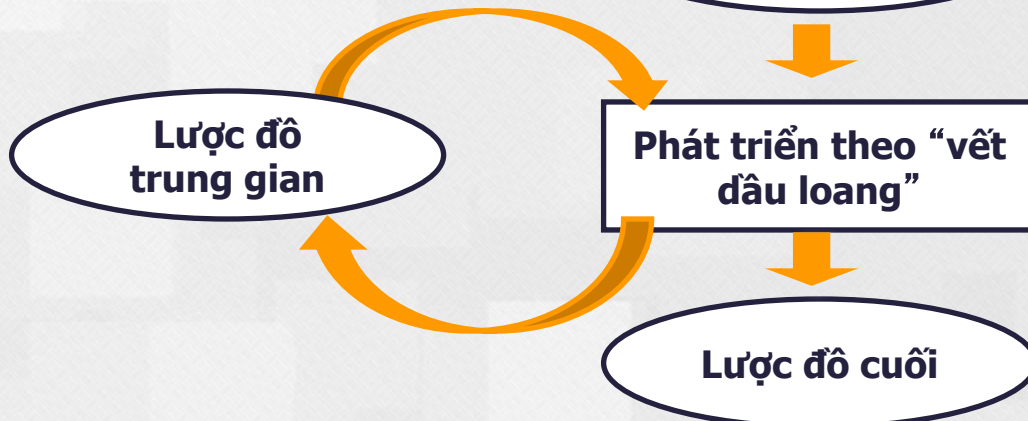


Phát triển theo “vết dầu loang”

Phát triển thêm các khái niệm có liên quan đến khái niệm khởi điểm

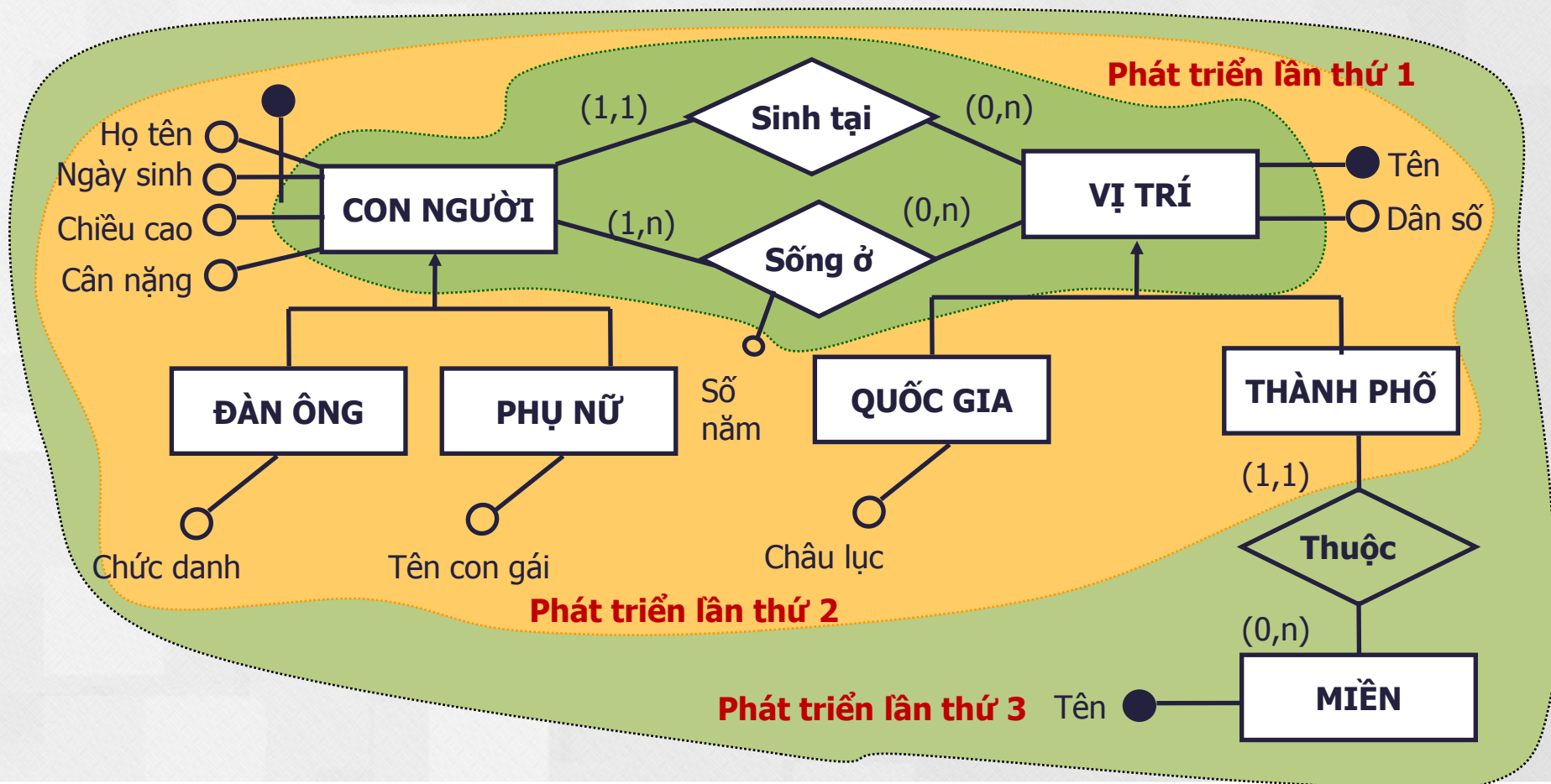
Lược đồ trung gian

Lược đồ cuối



Ví dụ

Ứng dụng Quản lý nhân chủng học

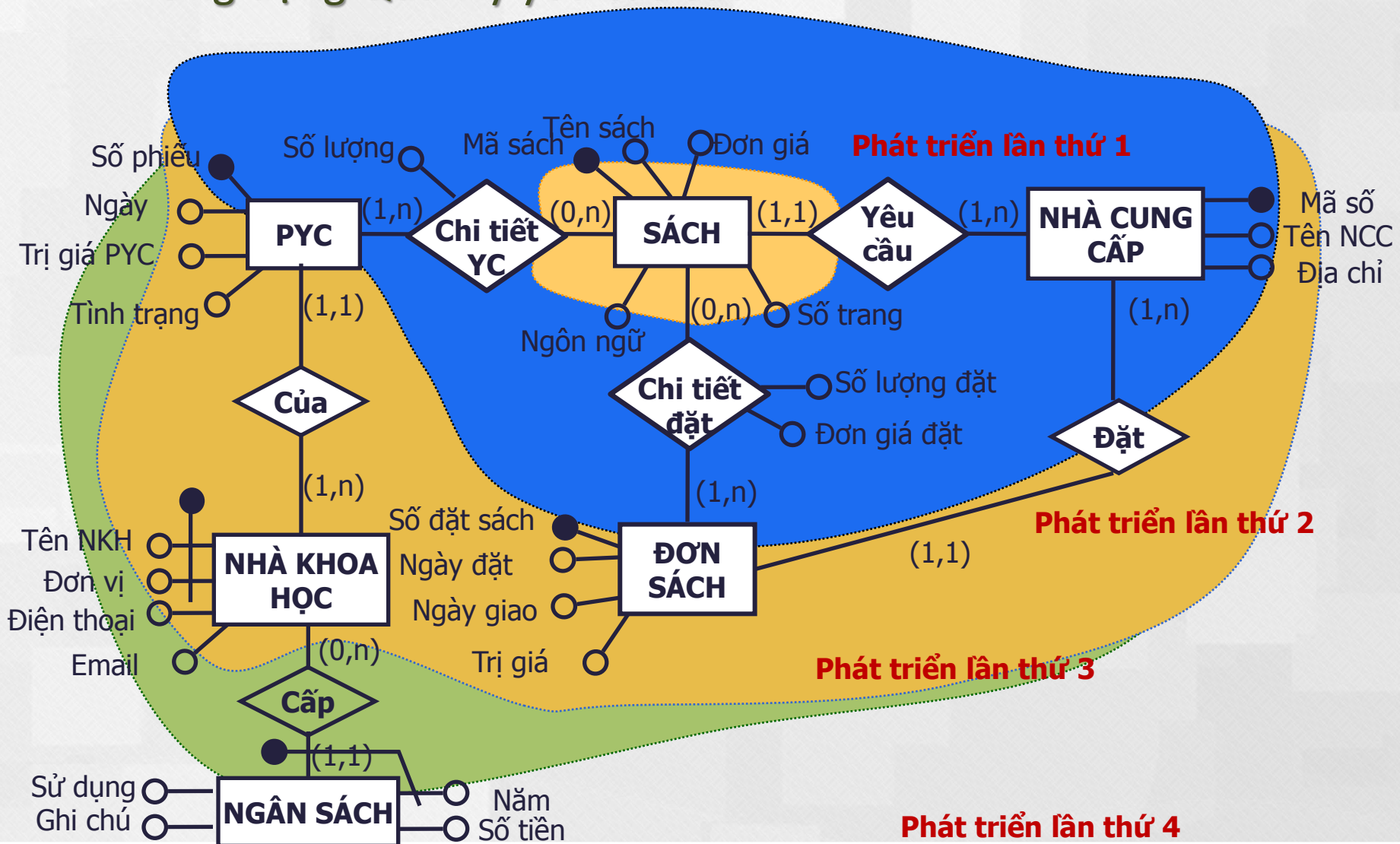


Ví dụ

- **Quản lý yêu cầu sách của nhà khoa học (NKH)**
 - Sau khi nhận được phiếu yêu cầu (PYC), NKH điền vào PYC sách cần mua và gửi lại cho nhân viên nghiệp vụ. Nhân viên này tiếp nhận PYC và lưu lại chờ ngày xử lý.
- **Đến thời điểm hết hạn quy định nộp. Nhân viên nghiệp vụ tập hợp tất cả các PYC và xử lý:**
 - Kiểm tra PYC có sách nào sách nào không thuộc danh mục sách có thể đặt hay không? hoặc PYC có tổng trị giá có không vượt quá số ngân sách được cấp cho NKH hay không?
 - Nếu một trong hai điều kiện trên không thoả thì nhân viên sẽ thông báo cho NKH để điều chỉnh.
 - Nếu cả hai điều kiện đều thoả thì nhân viên sẽ phân loại các sách cần đặt trên tất cả các PYC theo từng nhà cung ứng (NCU).
 - Lập đơn đặt sách gửi đến cho từng NCU và thông báo cho NKH ngày dự kiến nhận sách.

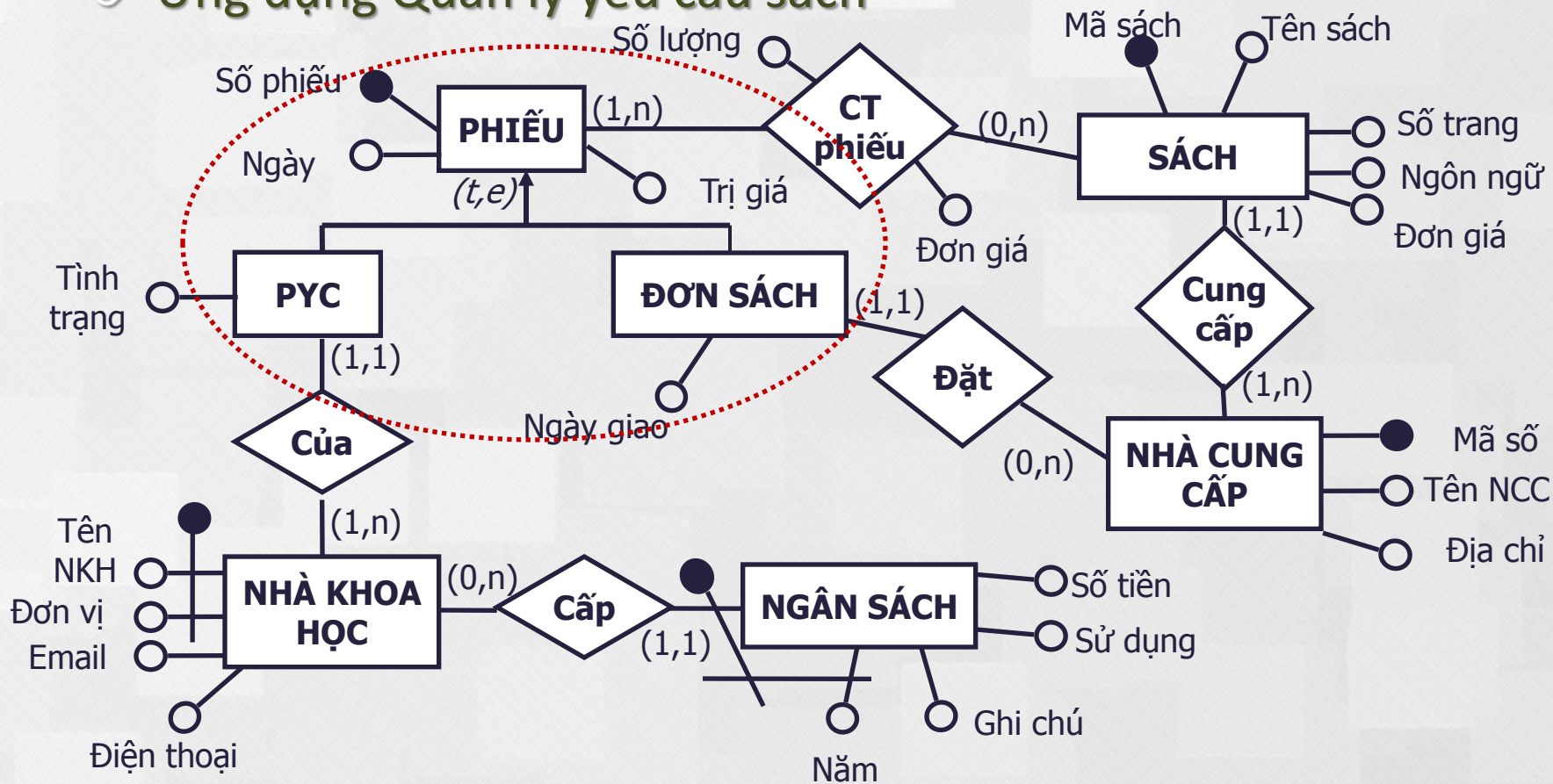
Ví dụ

Ứng dụng Quản lý yêu cầu sách



Ví dụ

Ứng dụng Quản lý yêu cầu sách

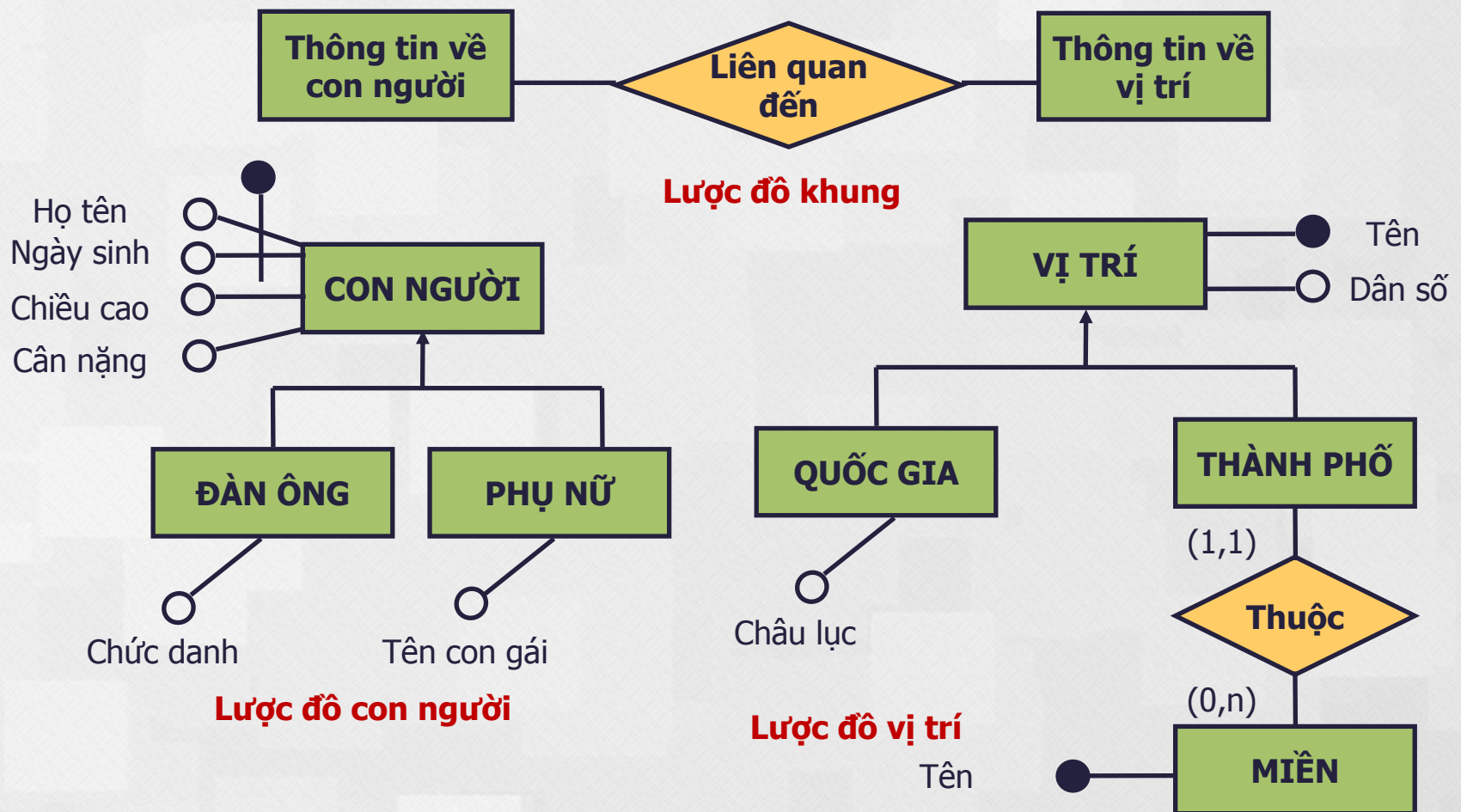


Chiến lược phối hợp



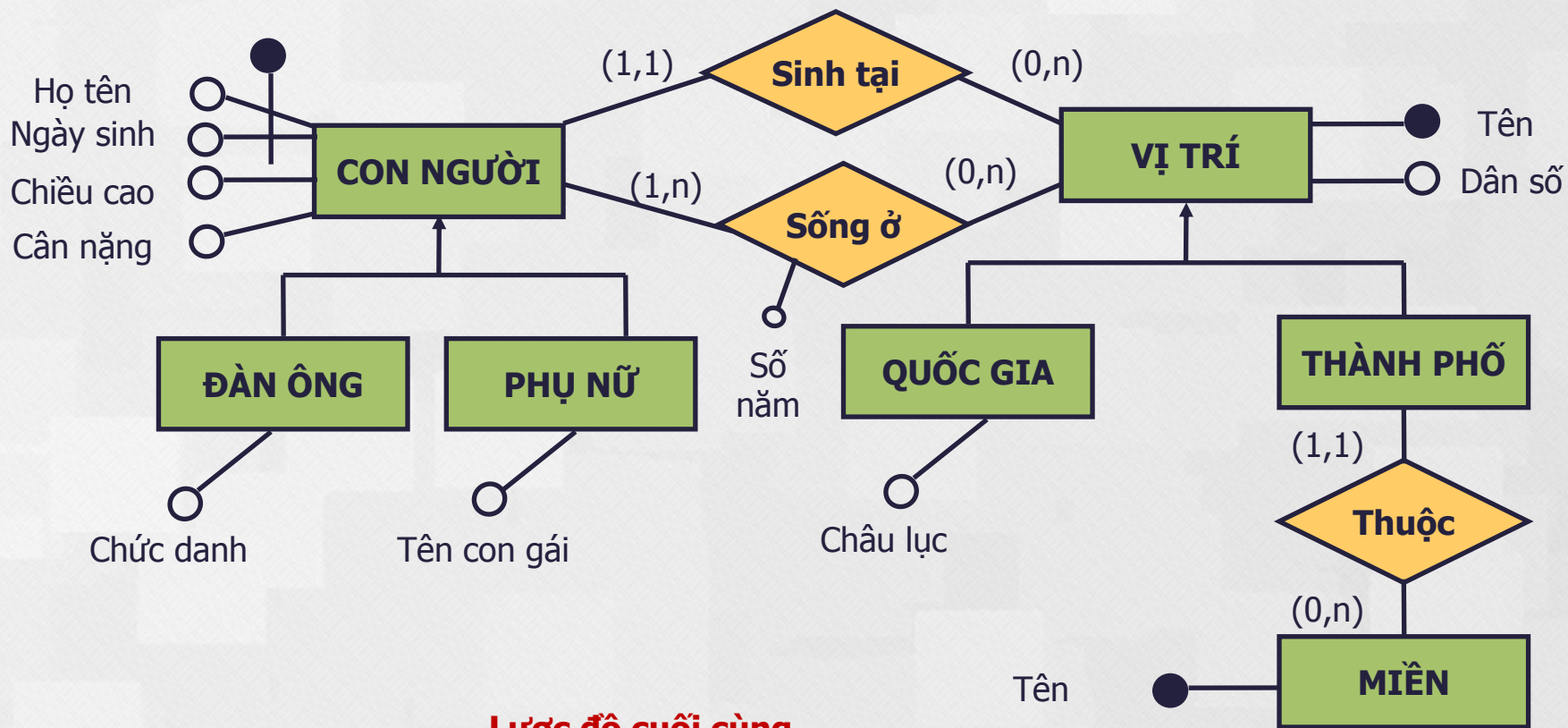
Ví dụ

Ứng dụng Quản lý nhân chủng học



Ví dụ

Ứng dụng Quản lý nhân chủng học



So sánh các chiến lược

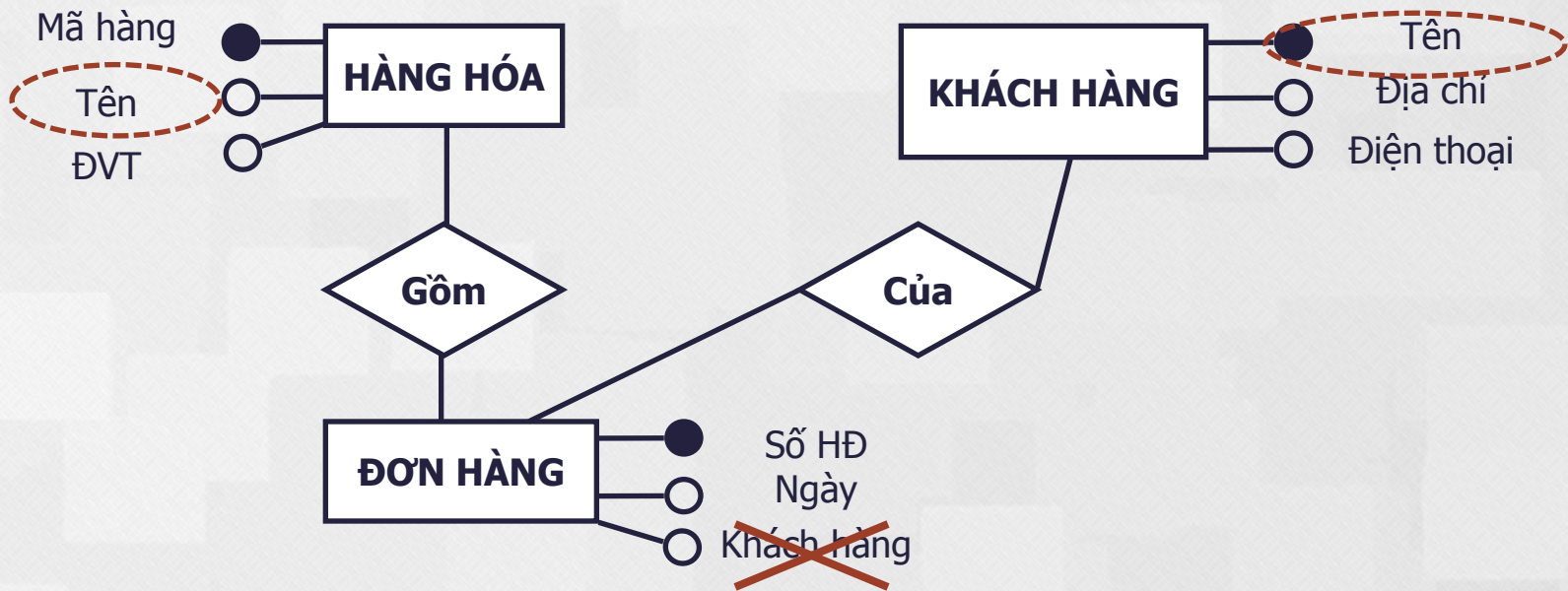
Chiến lược	Mô tả	Ưu điểm	Khuyết điểm
Trên xuống	Các khái niệm từng bước được tinh chế	Không có các hiệu ứng lề không mong muốn	Đòi hỏi phân tích viên phải giỏi với mức trừu tượng hóa cao lúc khởi điểm
Dưới lên	Các khái niệm được xây dựng từ các thành phần cơ bản	<ul style="list-style-type: none">- Dễ dàng cho các ứng dụng có tính chất cục bộ- Phân tích không phải chịu gánh nặng quá lớn lúc bắt đầu	Khi cần thiết có thể phải xây dựng lại cấu trúc trong quá trình tinh chế ứng dụng luật cơ bản
Trong ngoài	Các khái niệm được xây dựng theo cách tiếp cận “vết dầu loang”	<ul style="list-style-type: none">- Dễ dàng phát hiện ra các khái niệm mới liên quan đến các khái niệm đã có- Phân tích không phải chịu gánh nặng quá lớn lúc bắt đầu	Hình ảnh toàn bộ của ứng dụng chỉ được xây dựng vào giai đoạn cuối cùng
Phối hợp	Phân tích từ trên xuống các yêu cầu, tích hợp từ dưới lên, sử dụng lược đồ khung	Tiếp cận theo cách “chia để trị” để giảm độ phức tạp	Đòi hỏi quyết định quan trọng về lược đồ khung tại thời điểm bắt đầu của quá trình thiết kế

Nội dung trình bày

- Mô hình thực thể kết hợp nguyên thủy
- Mô hình thực thể kết hợp mở rộng
- Phương pháp phân tích dữ liệu
- Quy tắc mô hình hóa quan niệm dữ liệu
- Tiêu chuẩn chọn lựa khái niệm

Qui tắc 1

Mọi thuộc tính chỉ mô tả đặc trưng cho đúng thực thể mang thuộc tính ấy



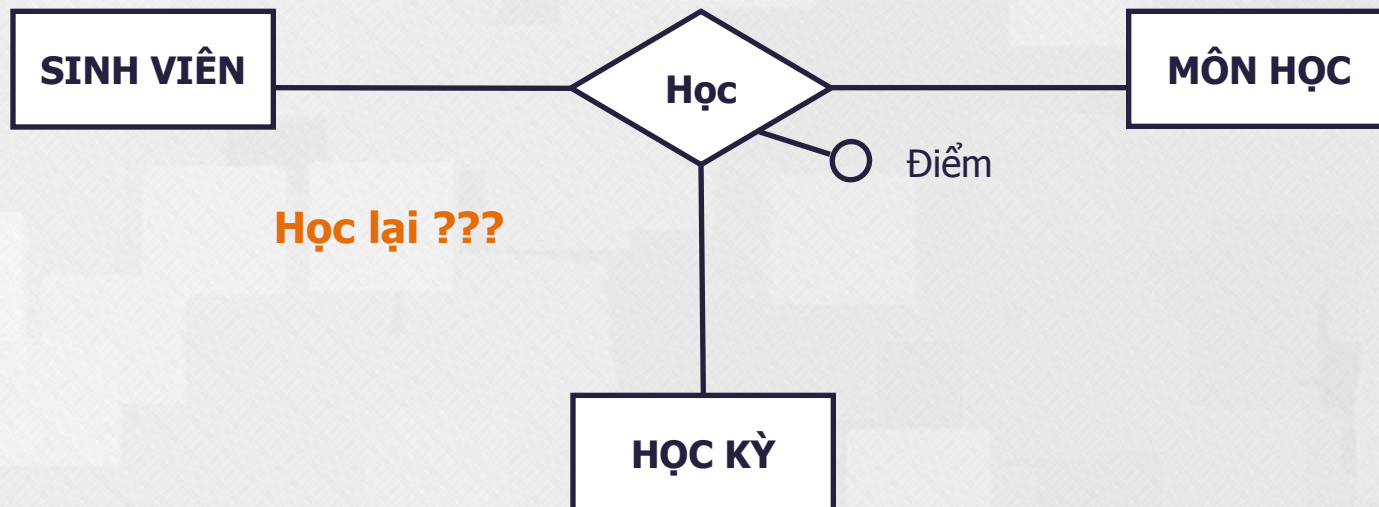
Qui tắc 2

- Nếu có đặc trưng phụ thuộc vào nhiều thực thể thì đó là đặc trưng của mối kết hợp định nghĩa trên các thực thể đó



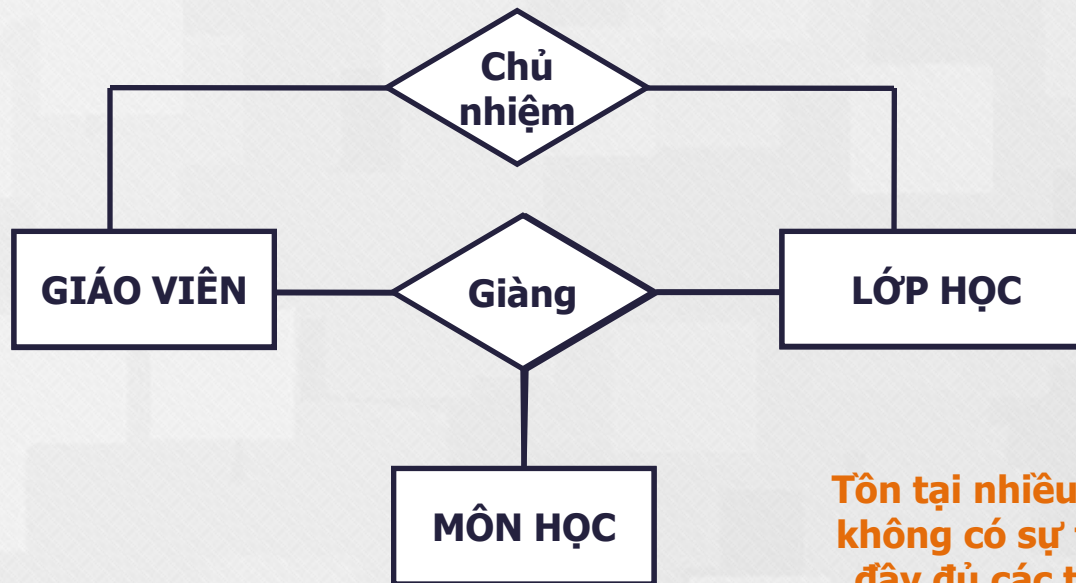
Qui tắc 3

- Các thực thể cùng liên quan đến 1 mối kết hợp thì một tổ hợp thể hiện của các thực thể đó phải là thể hiện duy nhất của mối kết hợp



Qui tắc 4

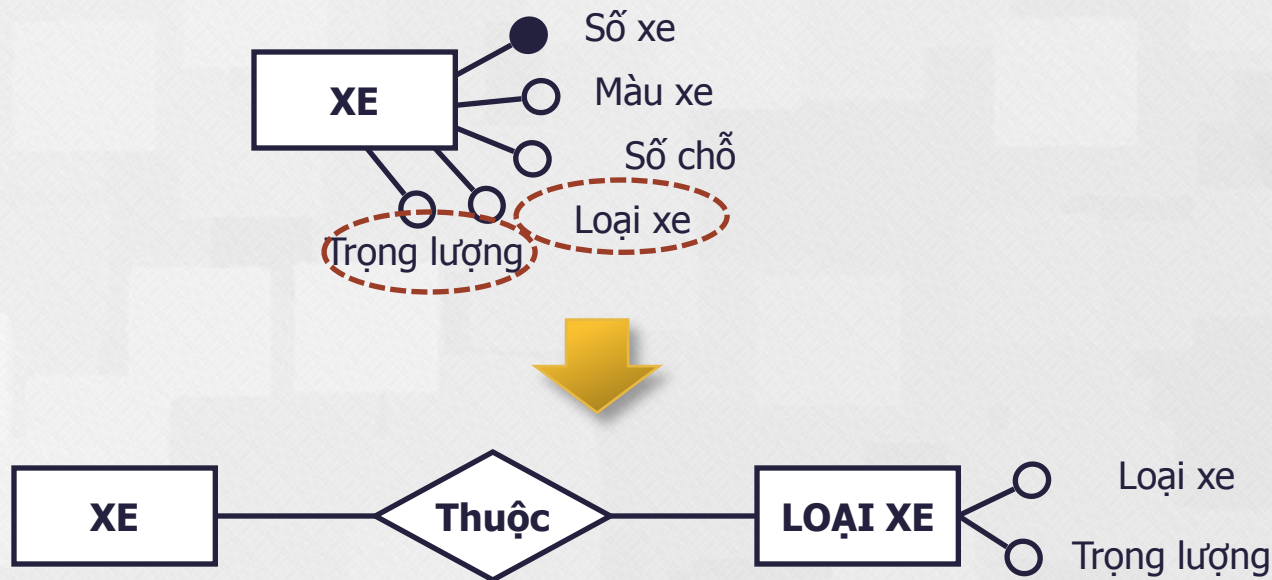
- Các nhánh nối với mỗi kết hợp phải là nhánh bắt buộc, nếu không phải ta nên tách thành nhiều mối kết hợp



Tồn tại nhiều thể hiện
không có sự tham gia
đầy đủ các thực thể

Qui tắc 5

- Nếu có 1 đặc trưng phụ thuộc vào 1 thuộc tính của thực thể thì tồn tại thực thể ẩn → cần được định nghĩa bổ sung**



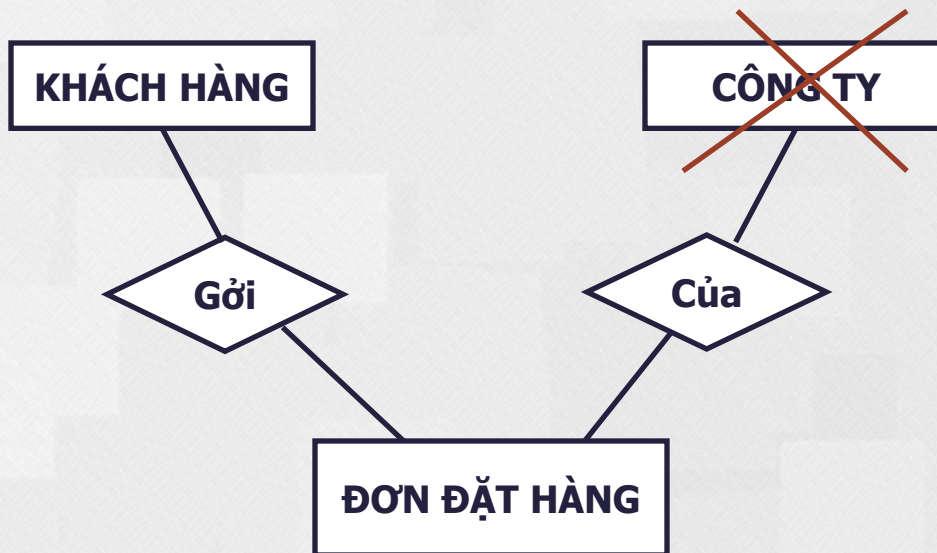
Nội dung trình bày

- Mô hình thực thể kết hợp nguyên thủy
- Mô hình thực thể kết hợp mở rộng
- Phương pháp phân tích dữ liệu
- Quy tắc mô hình hóa quan niệm dữ liệu
- Tiêu chuẩn chọn lựa khái niệm

Thực thể hay không là thực thể?

Trong 1 số trường hợp, các khái niệm cần biểu diễn chỉ có 1 thể hiện, nếu không có nhu cầu mở rộng về sau thì không nên xem là thực thể

Khách hàng gửi đơn đặt hàng cho công ty ...



Chỉ có một thể hiện
CÔNG TY trong suốt
chu trình hệ thống

Thực thể hay không là thực thể?

Đối tượng quan tâm không có cấu trúc đặc trưng (chỉ có 1 thuộc tính) thì cẩn thận khi quyết định đó là 1 thực thể

NGK thuộc một loại và có một hiệu nào đó (ví dụ như Tribeco) ...



Nếu Loại nước không được đề cập tới các đặc trưng khác thì nên là thuộc tính

Thực thể hay thuộc tính?

- **Thực thể:** xác định 1 số đặc trưng cơ bản như thuộc tính, mối kết hợp, tổng quát hóa
- **Thuộc tính:** cấu trúc nguyên tố, đơn giản, không có các đặc trưng khác

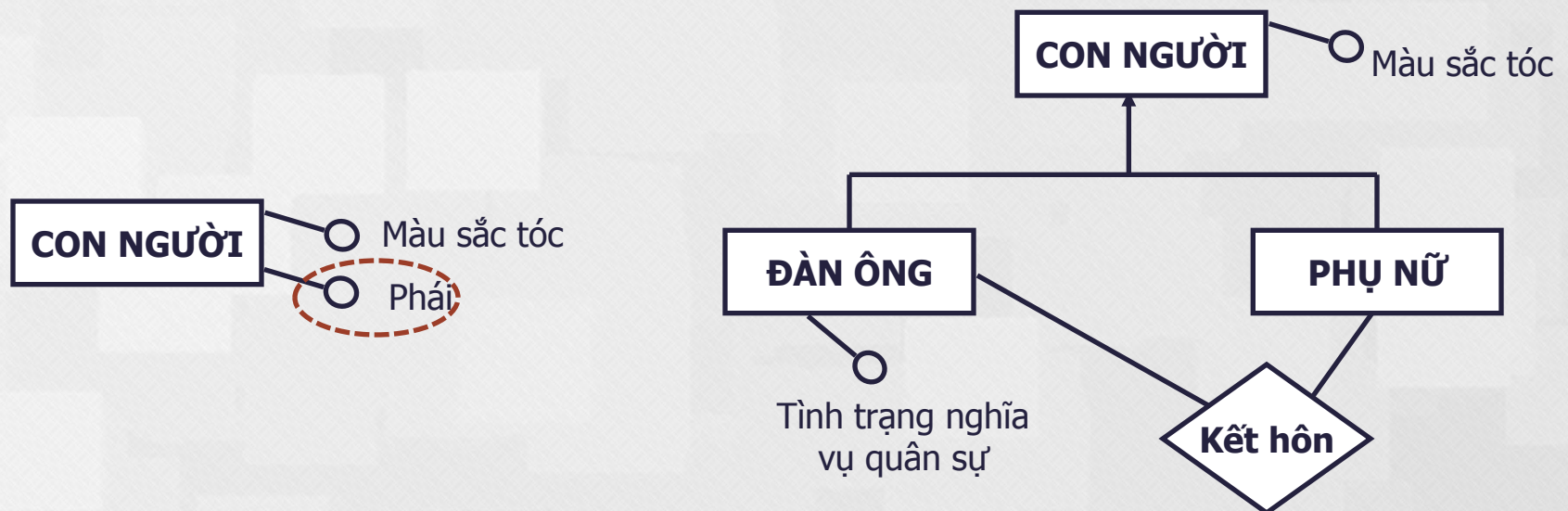
Mỗi xe hơi đặc trưng bởi mã số xe, loại xe, hãng sản xuất, số chỗ và màu sắc ...)



Tổng quát hóa hay thuộc tính?

- **Tổng quát hóa: một số đặc trưng sẽ được liên kết ở cấp thấp hơn**
- **Thuộc tính: trường hợp ngược lại**

Mỗi con người có họ tên, giới tính, ngày sinh và màu tóc. Riêng trường hợp nam giới thì có thêm thông tin năm hoàn thành nghĩa vụ quân sự ...



Thuộc tính kết hợp hay đơn?

- **Thuộc tính kết hợp: một số đặc trưng sẽ được liên kết ở cấp thấp hơn**
- **Thuộc tính đơn: trường hợp ngược lại**

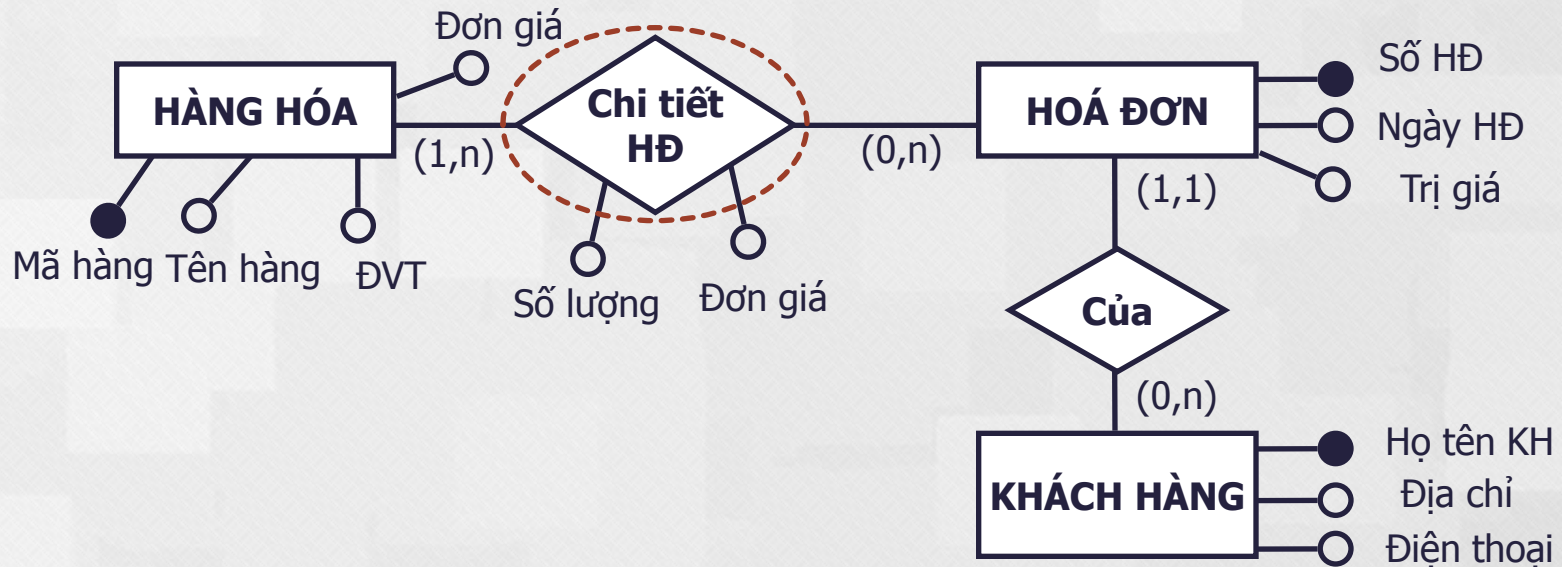
Mỗi sinh viên lưu trữ các thông tin như họ tên, địa chỉ (ghi rõ số nhà, đường, quận, thành phố), ngày sinh, nơi sinh ...



Mối kết hợp hay thực thể?

- Thực thể: khái niệm quan tâm có 1 số đặc trưng (mối kết hợp, định danh...)

Khi bán hàng, nhà cung cấp sẽ lập hóa đơn chứa các thông tin như số phiếu, ngày lập phiếu, tổng số tiền. Trong hóa đơn gồm nhiều chi tiết hóa đơn, mỗi chi tiết gồm mã số hàng hóa, số lượng, đơn giá và thành tiền ...



Mối kết hợp hay thực thể?

