

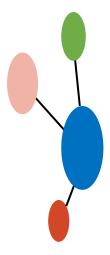
# Chương 2 THÀNH PHẦN DỮ LIỆU

#### Phan Tấn Tài

Khoa Hệ thống thông tin, Trường CNTT-TT Trường Đại học Cần Thơ

#### Nội dung

- Các khái niệm cơ bản
- Mô hình dữ liệu mức quan niệm
- Mô hình dữ liệu mức luận lý
- Mô hình dữ liệu mức vật lý



#### Các khái niệm

- Mô hình dữ liệu (data model) là một mô hình trừu tượng mô tả tổ chức dữ liệu, ngữ nghĩa dữ liệu và các ràng buộc toàn vẹn trên dữ liệu.
  - Nhấn mạnh vào dữ liệu nào là cần thiết và được tổ chức như thế nào.

#### Các mức trừu tượng của mô hình dữ liệu

- Mô hình dữ liệu mức quan niệm (CDM Conceptual Data Model)
- Mô hình dữ liệu mức luận lý (LDM Logical Data Model)
- Mô hình dữ liệu mức vật lý (PDM Physical Data Model)

#### Các khái niệm

- Mô hình dữ liệu mức quan niệm
  - Xác định những dữ liệu cần lưu trữ.
  - Mô hình này thường được tạo bởi chuyên gia lĩnh
     vực và kiến trúc sư dữ liệu (người phân tích, thiết kế).
  - Mục tiêu là để tổ chức, xác định phạm vi dữ liệu cùng các quy luật nghiệp vụ.

#### Các khái niệm

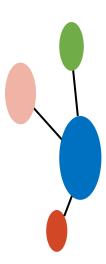
#### Mô hình dữ liệu mức luận lý

- Xác định cách hệ thống sẽ cài đặt dữ liệu mà không quan tâm đến Hệ quản trị CSDL (lưu trữ thế nào, tạo lập ra sao, truy xuất bằng cách nào).
- Mục tiêu là phát triển sơ đồ kỹ thuật cho cấu trúc và quy luật dữ liệu.

#### Mô hình dữ liệu mức vật lý

- Mô tả cách hệ thống sẽ cài đặt dữ liệu bằng cách sử dụng một Hệ quản trị CSDL cụ thể nào đó.
- Mục tiêu là cài đặt CSDL của hệ thống.

- □ Các khái niệm
  - Thực thể, kiểu thực thể và thể hiện của kiểu thực thể
  - Thuộc tính
  - Tổng quát hóa / Chuyên biệt hóa
  - Khóa của thực thể
  - Quan hệ (Relationship)
  - Bản số trong quan hệ
  - Thực thể yếu, khóa của thực thể yếu

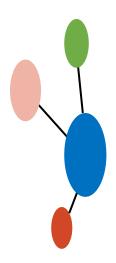


### ■ Một thực thể (Entity):

Là một con người, một nơi, một đối tượng, một sự kiện hay một khái niệm trong môi trường người dùng mà tổ chức đó muốn duy trì dữ liệu.

#### ■ Ví dụ:

- Con người: công nhân, sinh viên,...
- Nơi: cửa hàng, kho, tiểu bang,...
- Đối tượng: hàng hóa, tòa nhà, xe mô tô,...
- Sự kiện: việc bán hàng, sự đăng ký,
- Khái niệm: tài khoản, khóa học

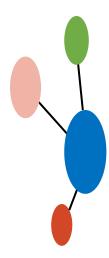


# ☐ Kiểu thực thể (Entity type):

Là một tập hợp các thực thể có cùng đặc điểm và tính chất.

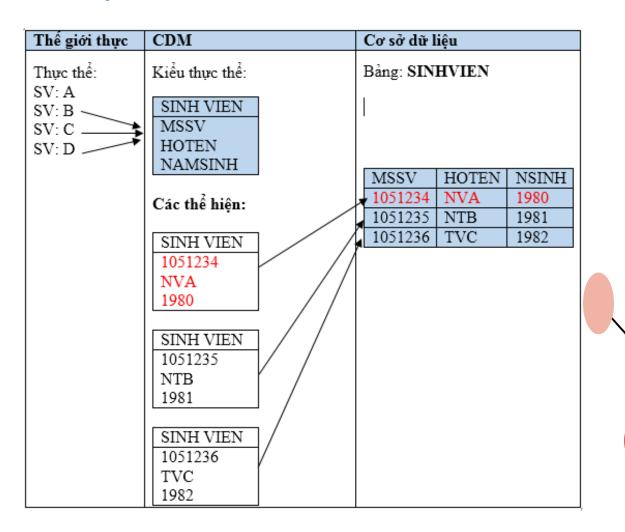
### □ Ký hiệu:





Thể hiện thực thể (Entity instance): Là một hiện thực của một

kiểu thực thể.



- ☐ Thuộc tính (Attribute): sẽ mô tả một tính chất hay một đặc điểm của một kiểu thực thể
- Miền trị của thuộc tính (Attribute domain): là một tập hợp các giá trị có thể gán được cho một thuộc tính
- □ Ví dụ: thuộc tính giới tính, có thể có một trong hại

giá trị "nam" hay "nữ"

Các thuộc tính



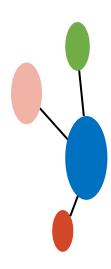
### ☐ Khóa của thực thể (Entity key):

Là một thuộc tính hay một tập hợp các thuộc tính mà giá trị của nó xác định duy nhất một thể hiện của một kiểu thực thể.

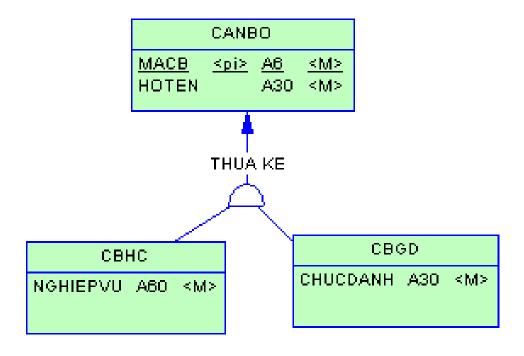
> STT MAHP HK NH TIETBD SOTIET MAPHONG

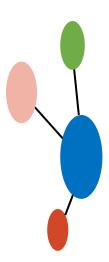






- ☐ Tổng quát hóa (generalization)
- ☐ Chuyên biệt hóa (specialization)





□ Mối quan hệ (Relationship):

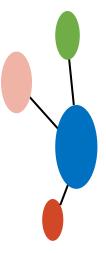
Là một sự kết hợp giữa các thể hiện của một hay nhiều kiểu thực thể mà nó liên quan với nhau.

Bậc của quan hệ (Degree):

Là số kiểu thực thể tham gia vào quan hệ

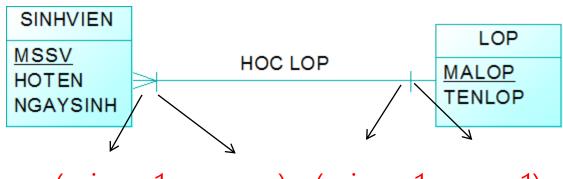
Quan hệ một ngôi hay quan hệ đệ qui (Unary Relationship/ Recursive Relationship)

- □ Các dạng quan hệ (Relationship types)
  - Quan hệ một một
  - Quan hệ một nhiều
  - Quan hệ nhiều nhiều
  - Các quan hệ đặc biệt
    - Quan hệ đệ qui
    - Quan hệ phụ thuộc hàm



### ■ Bản số (Cadinality)

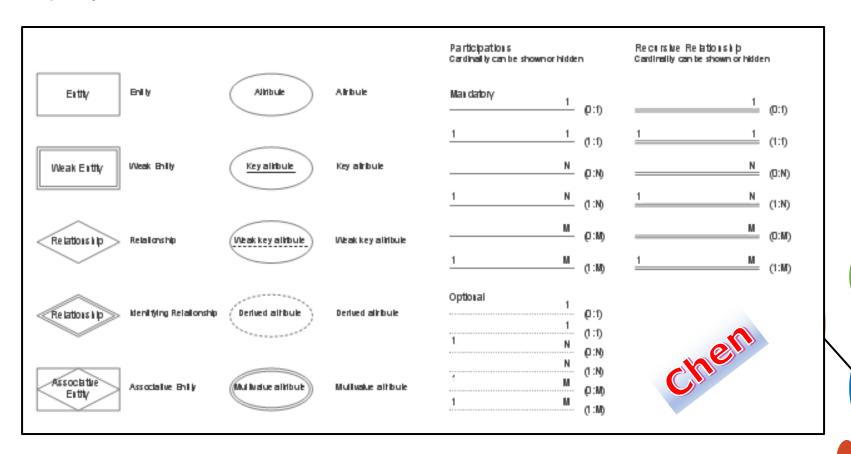
- Bản số của kiểu thực thể A trong mối quan hệ với kiểu thực thể B là số thể hiện của kiểu thực thể B có thể kết hợp với một thể hiện của kiểu thực thể A.
- Mỗi bản số: có hai giá trị là bản số tối thiểu (min) và bản số tối đa (max)
- Bản số min = {0, 1} và Bản số max = {1, n ≥2}



(min = 1, max = n) (min = 1, max = 1)

# Tập ký hiệu

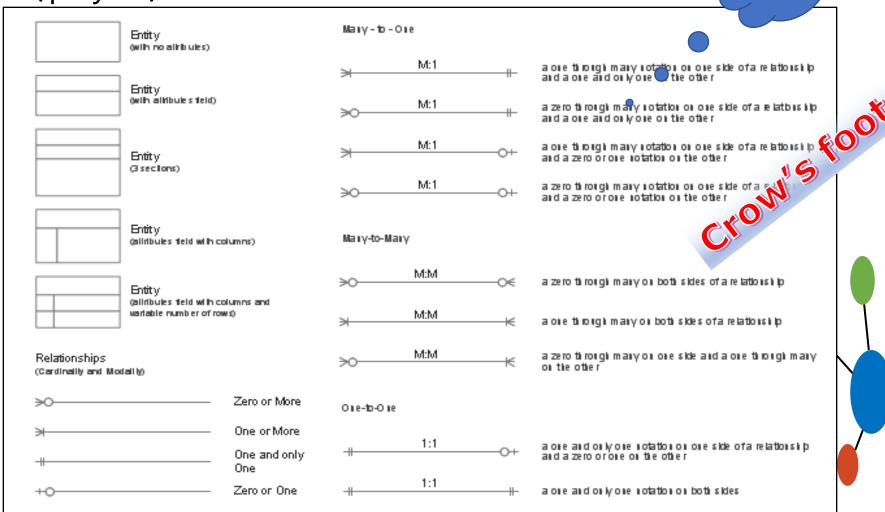
• Tập ký hiệu Chen.



### Tập ký hiệu

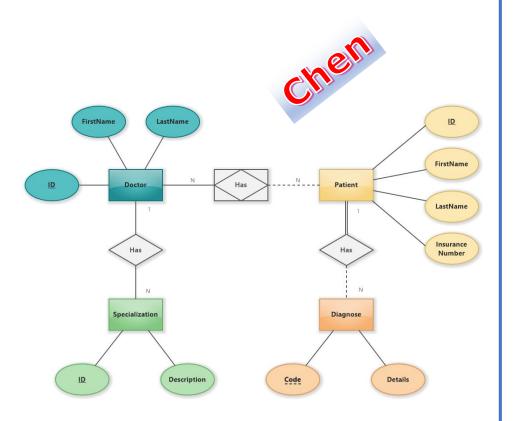
Sử dựng tấp Ký hiệu này

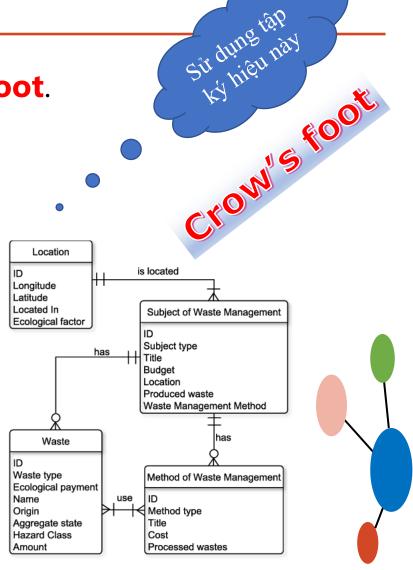
Tập ký hiệu Crow's foot.



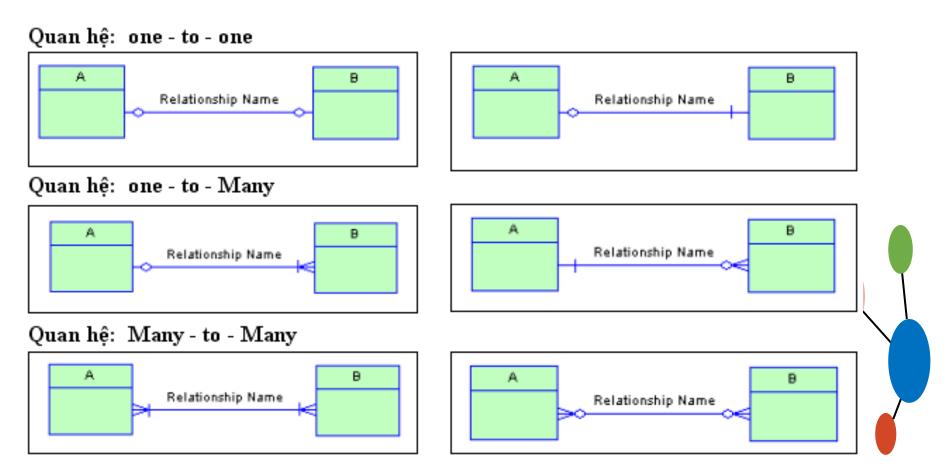
# Tập ký hiệu

Có 2 tập ký hiệu: Chen, Crow's foot.



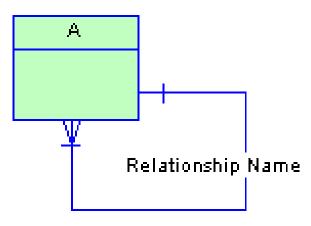


### Các dạng quan hệ (Relationship types)

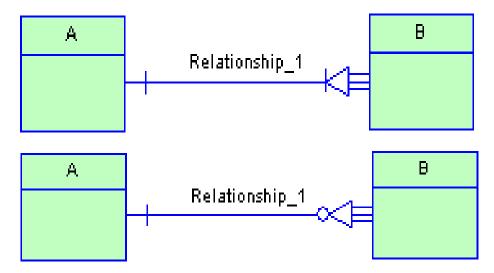


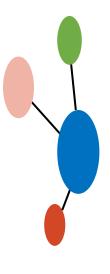
#### □ Các kiểu quan hệ (Relationship types)

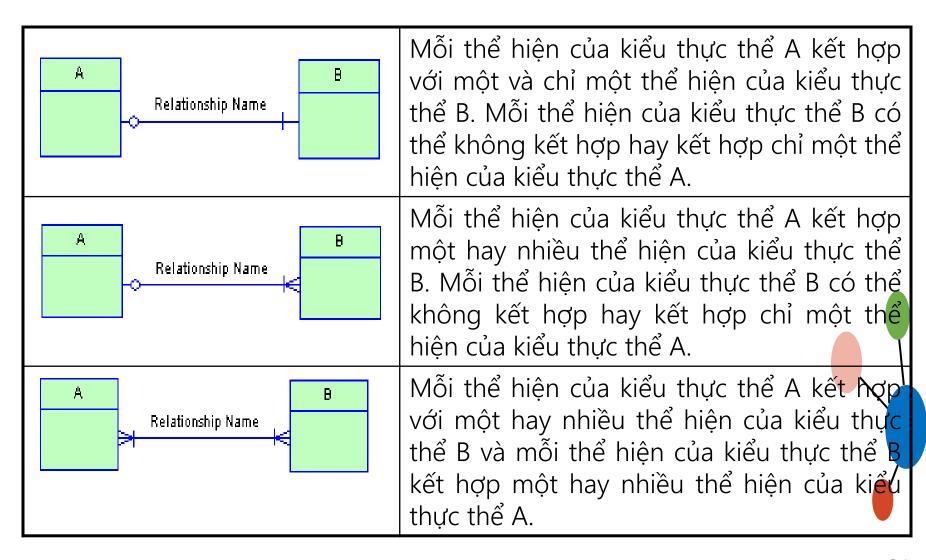
Quan hệ đệ qui

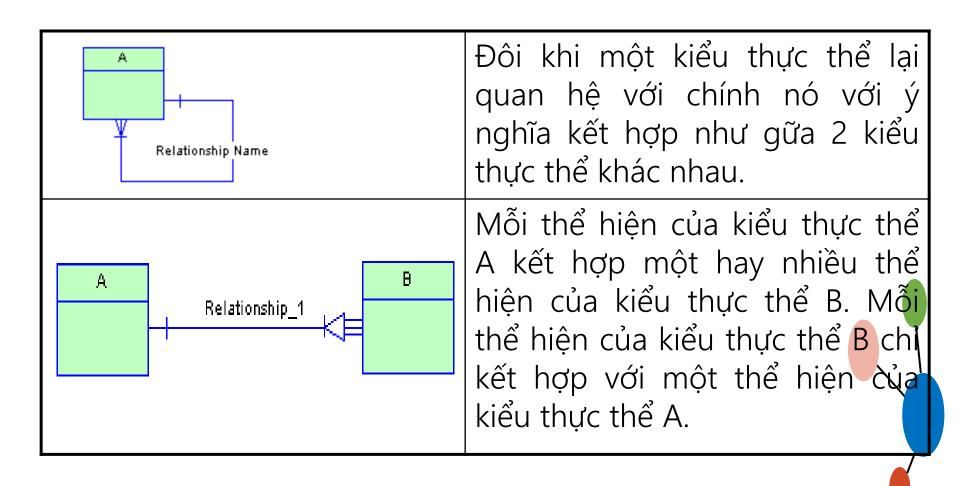


Quan hệ phụ thuộc hàm









☐ Kiểu thực thể yếu (weak entity):

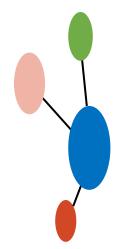
Là kiểu thực thể có phụ thuộc hàm vào các kiểu thực thể khác.

☐ Khóa của kiểu thực thể yếu:

Là khóa của các kiểu thực thể mà nó có phụ thuộc hàm và các thuộc tính khóa của nó (nếu có).

#### Các bước xây dựng CDM

- B1: Xác định các kiểu thực thể
- B2: Xác định các mối quan hệ và bản số
- B3: Xác định các thuộc tính và miền trị
- B4: Xác định các khóa ứng viên và khóa chính
- B5: Xác định tổng quát hóa/
   chuyên biệt hóa các thực thể (nếu có)
- B6: Vẽ sơ đồ E-R.
- B7: Kiểm tra lại ERD



#### Các bước xây dựng CDM - Ví dụ

Nhà học C1 có nhiều tầng, mỗi tầng có nhiều phòng và trong mỗi phòng được trang bị một số loại thiết bị cần thiết. Mỗi tầng có số thứ tự tầng và diễn giải. Mỗi phòng có số thứ tự phòng, sức chứa của phòng và được trang bị một số loại thiết bị cần thiết (như bàn, ghế, quạt, máy chiếu,...).

Mỗi <u>loại thiết bị</u> được đánh một <u>mã loại thiết bị</u> và có một <u>diễn giải cho loại thiết bị</u> đó. Một loại thiết bị có thể được trang bị cho một phòng hay không được trang bị cho phòng nào, một phòng được trang bị ít nhất một loại thiết bị nào đó.

25

#### Các bước xây dựng CDM - Ví dụ

Cách đánh số thứ tự phòng như sau:

- Phòng 1 ở tầng 1 được đánh stt là 101, phòng 2 ở tầng 1 được đánh stt là 102, phòng 3 ở tầng 1 được đánh stt là 103,
- Phòng 1 ở tầng 2 được đánh stt là 201, phòng 2 ở tầng
  2 được đánh stt là 202, phòng 3 ở tầng 2 được đánh stt là 203,

• ...

#### Các bước xây dựng CDM – Ví dụ

1. Xác định các kiểu thực thể



2. Xác định các quan hệ và bản số



3. Xác định các thuộc tính



#### Các bước xây dựng CDM – Ví dụ

- 4. Xác định khóa của thực thể
  - Thực thể TANG có khóa là STT\_TANG
  - Thực thể PHONG có khóa là STT\_TANG và STT\_PHONG
  - Thực thể LOAI\_THIETBI có khóa là MA\_TB
- 5. Tổng quát hóa/chuyên biệt hóa (không có)
- 6. Vẽ ERD



7. Kiểm tra lại ER-D

Quy tắc 1: Mỗi kiểu thực thể trong mô hình thực thể - quan hệ được chuyển thành một quan hệ (bảng).

Ví dụ: Kiểu thực thể NHANVIEN



Mô hình dữ liệu mức luận lý (LDM)

NHANVIEN(**MANV**, HOTEN, GIOITINH, NGAYSINH, DIACHI, QUEQUAN

Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt hóa khi chuyển sang LDM có thể sử dụng 1 trong ba cách sau:

#### Cách 1:

- Xóa đi các kiểu thực thể chuyên biệt hóa, chỉ sinh ra đúng một quan hệ với tên là tên của thực thể tổng quát.
- Thuộc tính của quan hệ là tất cả các thuộc tính của kiểu thực thể tổng quát hóa và các thuộc tính của các kiểu thực thể chuyên biệt hóa. Các thuộc tính của kiểu thực thể chuyên biệ có thể nhận giá trị rỗng.
- Khóa chính của quan hệ là khóa chính của kiểu thực thể tổng quát hóa.

Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt hóa khi chuyển sang LDM có thể sử dụng 1 trong ba cách sau:

#### Cách 2:

- Xóa thực thể tổng quát hóa. Các kiểu thực thể chuyên biệt được chuyển thành các quan hệ tương ứng trong LDM.
- Các thuộc tính của mỗi quan hệ bao gồm các thuộc tính của kiểu thực thể tổng quát hóa và các thuộc tính riêng biệt tương ứng.
- Khóa chính của mỗi quan hệ là khóa chính của kiểu thực thể tổng quát hóa.

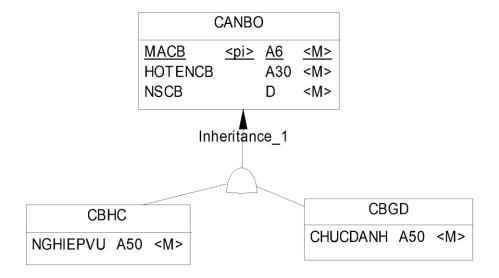
Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt hóa khi chuyển sang LDM có thể sử dụng 1 trong ba cách sau:

#### Cách 3:

- Mỗi kiểu thực thể tổng quát hóa hay chuyên biệt hóa đều được chuyển thành một quan hệ tương ứng trong LDM.
- Các thuộc tính của quan hệ tổng quát hóa gồm các thuộc tính khóa chính và các thuộc tính còn lại của nó.
- Các thuộc tính của các quan hệ chuyên biệt bao gồm các thuộc tính khóa của kiểu thực thể tổng quát và các thuộc tính riêng biệt tương ứng của nó.

Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt hóa khi chuyển sang LDM

#### Cách 1:

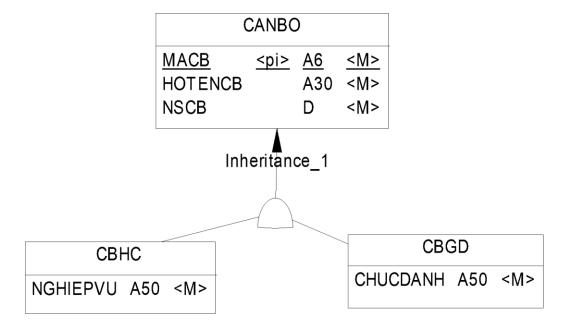




CANBO (MACB, HOTENCB, NSCB, NGHIEPVU, CHUCDANH)

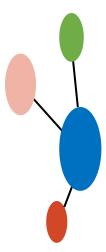
Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt hóa khi chuyển sang LDM

#### Cách 1:



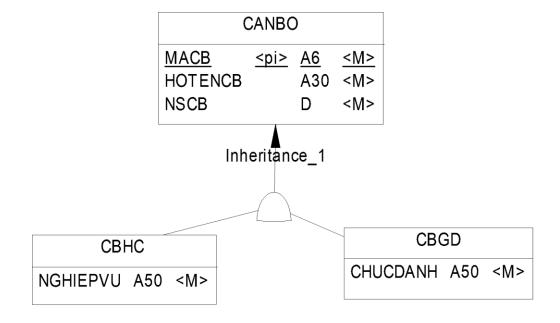


CANBO (MACB, HOTENCB, NSCB, NGHIEPVU, CHUCDANH)



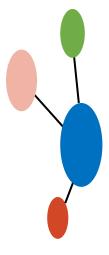
Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt hóa khi chuyển sang LDM

#### Cách 2:



#### LDM:

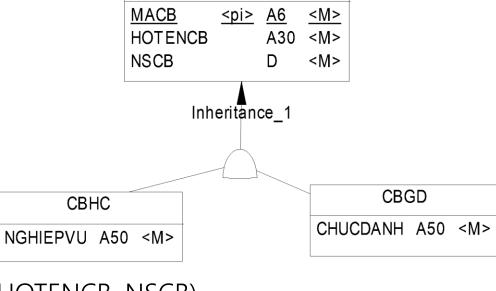
- CBHC (MACB, HOTENCB, NSCB, NGHIEPVU)
- CBGD (MACB, HOTENCB, NSCB, CHUCDANH)



Quy tắc 2: Các kiểu thực thể có tổng quát hóa/chuyên biệt

hóa khi chuyển sang LDM

Cách 3:



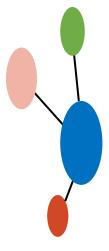
**CANBO** 

#### LDM:

- + CANBO (MACB, HOTENCB, NSCB)
- + CBHC (MACB, NGHIEPVU)
- + CBGD (MACB, CHUCDANH)

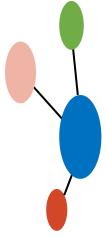
Các tham chiếu

- + CBHC(MACB) → CANBO(MACB)
- + CBGD(MACB) → CANBO(MACB)

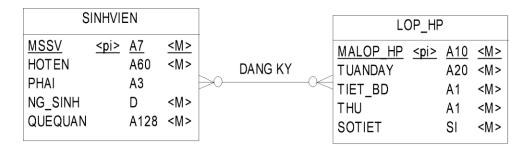


Quy tắc 3: Một quan hệ nhiều – nhiều (Many – to - Many) sẽ được chuyển thành một quan hệ (bảng) trong mô hình quan hệ.

- Quan hệ này có tên là tên của quan hệ, các thuộc tính là các thuộc tính khóa của các kiểu thực thể có liên quan.
- Khóa chính của bảng là tổ hợp tất cả các thuộc tính.
- Các tham chiểu tương ứng cũng được tạo



Quy tắc 3: Một quan hệ nhiều – nhiều (Many – to - Many) sẽ được chuyển thành một bảng trong mô hình quan hệ.



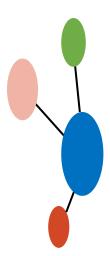
Mô hình quan hệ tương ứng là:

SINHVIEN (<u>MSSV</u>, HOTEN, PHAI, NG\_SINH, QUEQUAN) LOP\_HP (<u>MALOP\_HP</u>, TUANDAY, TIET\_BD, THU, SOTIET) DANGKY (<u>MSSV, MALOP\_HP</u>)

Các tham chiếu

DANGKY(MSSV) -----> SINHVIEN(MSSV)

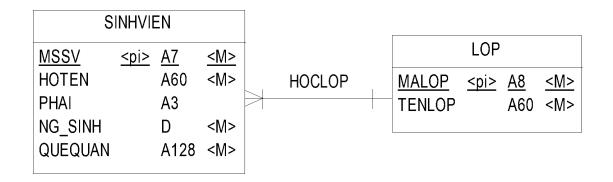
DANGKY(MALOP\_HP) ----> LOP\_HP(MALOP\_HP)



Quy tắc 4: mỗi quan hệ One-Many sẽ chuyển thành một tham chiếu trong mô hình quan hệ

- Một quan hệ One-Many sẽ trở thành một tham chiếu "bảng con đến bảng cha".
  - Nghĩa là, khóa chính của kiểu thực thể bên 1 sẽ được đưa về quan hệ (bảng) tương ứng với kiểu thực thể bên nhiều để đóng vai trò khóa ngoại.
- Mỗi quan hệ One-Many dạng phụ thuộc hàm: khóa chính của thực thể bên 1 vừa đóng vai trò khóa ngoại, vừa là thành phần của khóa chính bên nhiều.

Quy tắc 4: mỗi quan hệ One-Many sẽ chuyển thành một tham chiếu trong mô hình quan hệ.



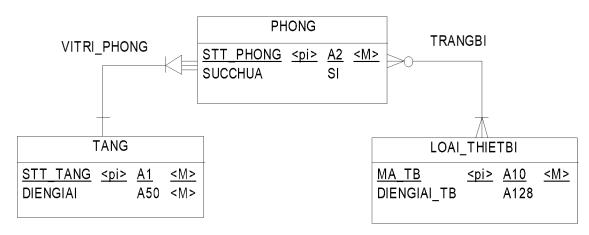
Mô hình quan hệ tương ứng là

SINHVIEN (<u>MSSV</u>, HOTEN, PHAI, NG\_SINH, QUEQUAN, MALOP) LOP (<u>MALOP</u>, TENLOP)

Tham chiếu sinh ra:

Fk1: SINHVIEN(MALOP) → LOP(MALOP)

#### Quy tắc 4:



#### LDM / Mô hình quan hệ

TANG (STT\_TANG, DIENGIAI)

PHONG (STT\_PHONG, STT\_TANG, SUCCHUA)

LOAI\_THIETBI (MA\_TB, DIENGIAI\_TB)

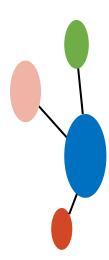
TRANGBI (STT\_PHONG, STT\_TANG, MA\_TB)

#### Các tham chiếu

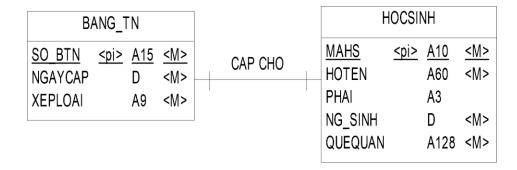
PHONG(STT\_TANG) → TANG(STT\_TANG)

TRANGBI(STT\_PHONG, STT\_TANG)→PHONG(STT\_PHONG, STT\_TANG)

 $TRANGBI(MA_TB) \rightarrow LOAI_THIETBI(MA_TB)$ 



Quy tắc 5: Chuyển quan hệ One-One thành quan hệ One-Many và áp dụng quy tắc 4.



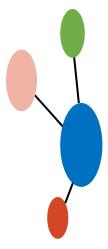
LDM / Mô hình quan hệ

HOCSINH(<u>MAHS</u>, HOTEN, PHAI, NG\_SINH, QUEQUAN)

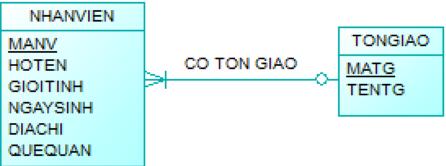
BANG\_TN(SO\_BTN, NGAYCAP, XEPLOAI, MAHS)

Tham chiếu:

BANG\_TN(MAHS) → HOCSINH(MAHS)



Quy tắc 6: Một bản số của quan hệ có dạng là Optional-One thì quan hệ sẽ được chuyển thành bảng..



#### Cách 1:

- NHANVIEN (MANV, HOTNE, GIOITINH, NGAYSINH, DIACHI, QUEQUAN)
- TONGIAO (MATG, TENTG)
- COTONGIAO (MANV, MATG)

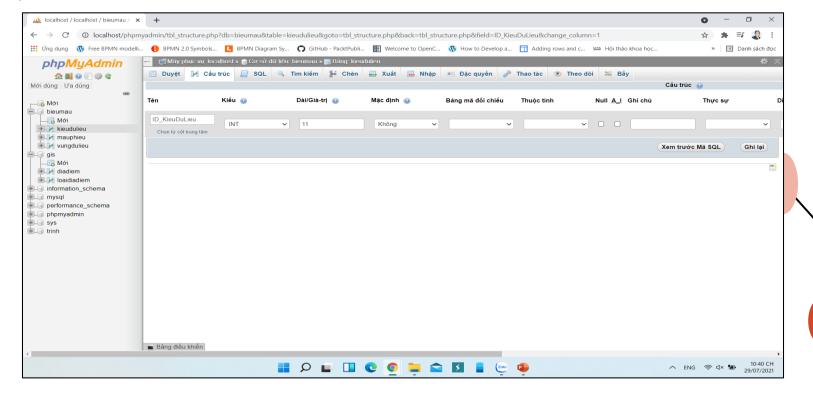
#### Cách 2:

- NHANVIEN (<u>MANV</u>, HOTNE, GIOITINH, NGAYSINH, DIACHI, QUEQUAN, <u>MATG</u>)
- TONGIAO (MATG, TENTG)

Lưu ý: MATG được phép NULL.

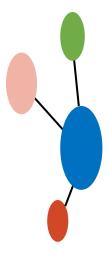


 Chọn định dạng lưu trữ cho mỗi thuộc tính của mô hình cơ sở dữ liệu vật lý, gồm: Kiểu dữ liệu, không gian lưu trữ dữ liệu tối thiểu, tối đa, chiều dài, số lượng chữ số thập phân, giá trị tối thiểu, giá trị tối đa, ...



TT	Tên	Kiểu	Kích thước	Số số lẻ	Miền giá trị	Trị mặc nhiên	MIN	MAX	Khó a chín h	Duy nhất	NO T NU LL	RB TV Luận Lý	RBTV Khóa ngoại	Diễn giải
1	CD_STT	Int			[1, ∞)		1		X	X	X			Kiểu số tự tăng dùng làm khóa chính
2	CD_CMND	Varchar	13							X	X			Số chứng minh nhân dân
3	CD_HoTen	Varchar	30								X			Họ tên công dân
4	CD_GioiTinh	Varchar	10								X			Giới tính, lưu nam hoặc nữ
5	CD_NgaySinh	Date									X	> 18 tuổi		Ngày sinh
6	CD_SoDienThoai	Varchar	11											Số điện thoại di động
7	CD_TrinhDoVan Hoa	Int			[1, 12]						X			Trình độ văn hóa

- Phân loại dữ liệu
  - Dữ liệu thường trực (dữ liệu cơ sở, dữ liệu danh mục).
  - Dữ liệu gốc.
  - Dữ liệu tạm thời.



- Dữ liệu thường trực
  - Là dữ liệu không thay đổi hay rất ít thay đổi theo thời gian.
  - Loại dữ liệu này chỉ có thể bổ sung (thêm bộ hay thêm thuộc tính), không nên thay đổi giá trị, đặc biệt là không được xóa.
- Trong hệ thống quản lý đào tạo, dữ liệu về học phần có là dữ liệu thường trực không? Nếu một học phần nào đó không còn giảng dạy nữa thì cá được xóa không? Tại sao?

#### Dữ liệu gốc

- Là dữ liệu lịch sử lưu những giá trị xảy ra theo thời gian, không gian.
- Dữ liệu này phát sinh với tốc độ nhanh nên phải có cách thức lưu trữ hoặc xử lý thích hợp.
- Cách thức lưu trữ
  - Tổ chức tập trung hay phân tán?
- Dữ liệu tạm thời
  - Chỉ dùng trong một thời gian nào đó, khi không cần có thể xóa đi.