

PHÂN TÍCH THIẾT KẾ HỆ THỐNG

Chương 4: Mô hình hóa dữ liệu

Trương Quốc Định

Khoa Hệ thống thông tin – Trường CNTT&TT – Trường Đại học Cần Thơ

Nội dung

- Khái niệm
- Sơ đồ thực thể liên kết (ERD – Entity Relationship Diagram)
- Thiết kế mô hình dữ liệu mức luận lý
- Thiết kế mô hình dữ liệu mức vật lý

Khái niệm

- **Mô hình hóa dữ liệu** (data modelling) là quá trình tạo **mô hình (cấu trúc)** cho dữ liệu sẽ được lưu trữ trong cơ sở dữ liệu.
- **Mô hình dữ liệu** là một biểu diễn mức khái niệm của các đối tượng dữ liệu, sự kết hợp giữa các đối tượng dữ liệu, các quy tắc.
- Mô hình hóa dữ liệu giúp thể hiện trực quan dữ liệu và các quy luật nghiệp vụ.
- Mô hình dữ liệu cần đảm bảo tính nhất quán trong **quy ước đặt tên, giá trị mặc định, ngữ nghĩa, bảo mật** trong khi vẫn đảm bảo chất lượng của dữ liệu.

Khái niệm

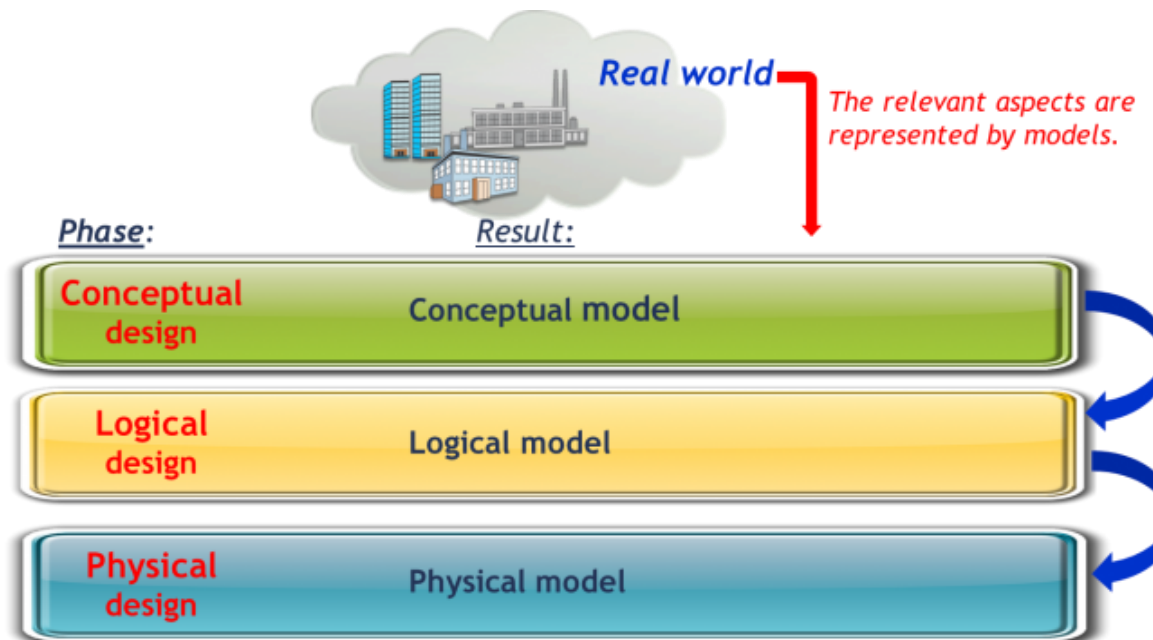
- **Mô hình dữ liệu** (data model) là một mô hình trừu tượng mô tả **tổ chức** dữ liệu, **ngữ nghĩa** dữ liệu và các **ràng buộc toàn vẹn** trên dữ liệu.
 - Nhấn mạnh vào **dữ liệu nào là cần thiết** và **được tổ chức như thế nào** thay vì các hoạt động sẽ được thực hiện trên dữ liệu.



Các mức trừu tượng của mô hình dữ liệu

- Có ba mức trừu tượng:
 - 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)

mỗi mức có một **mục đích cụ thể, khác nhau.**



Các mức trừu tượng của mô hình dữ liệu

- Mô hình dữ liệu **mức quan niệm**
 - Xác định **NHỮNG GÌ** hệ thống 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 mức trừu tượng của mô hình dữ liệu

- 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**.



Các mức trừu tượng của mô hình 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**.



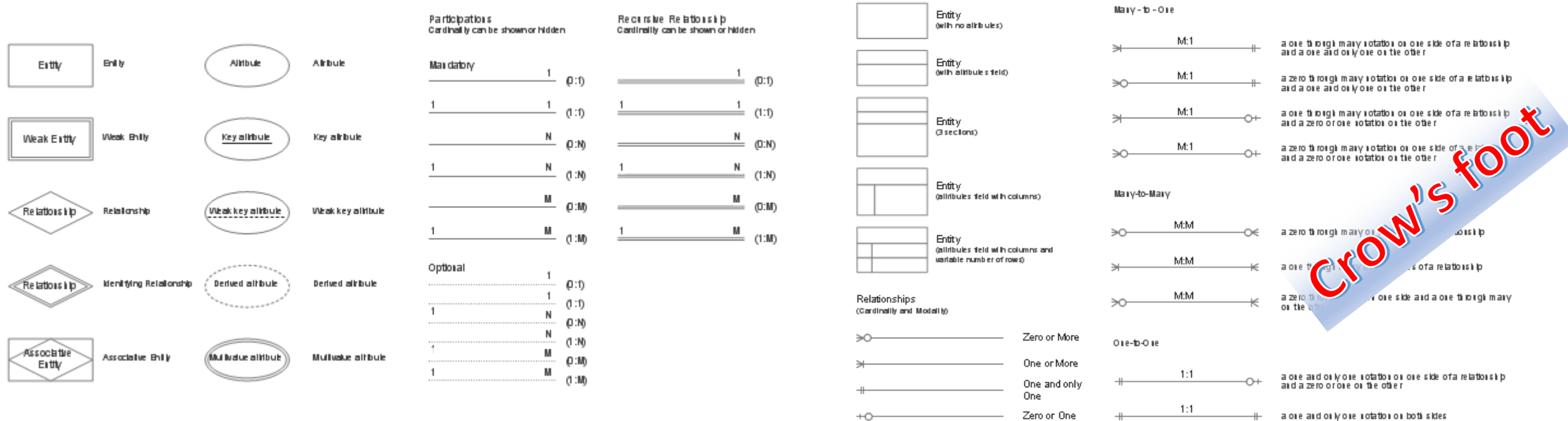
Sơ đồ thực thể liên kết

- Thường được sử dụng để biểu diễn mô hình dữ liệu **mức quan niệm**.
- Bức tranh toàn cảnh biểu diễn cách thức tạo lập, lưu trữ, sử dụng thông tin của một hệ thống nghiệp vụ.
- Hỗ trợ biểu diễn các **quy luật nghiệp vụ** ở mức độ trừu tượng cao.
- Biểu diễn các **liên kết (relationship)** của tập các **thực thể (entity)**.

Tập ký hiệu

- Có 2 tập ký hiệu: Chen, **Crow's foot**.

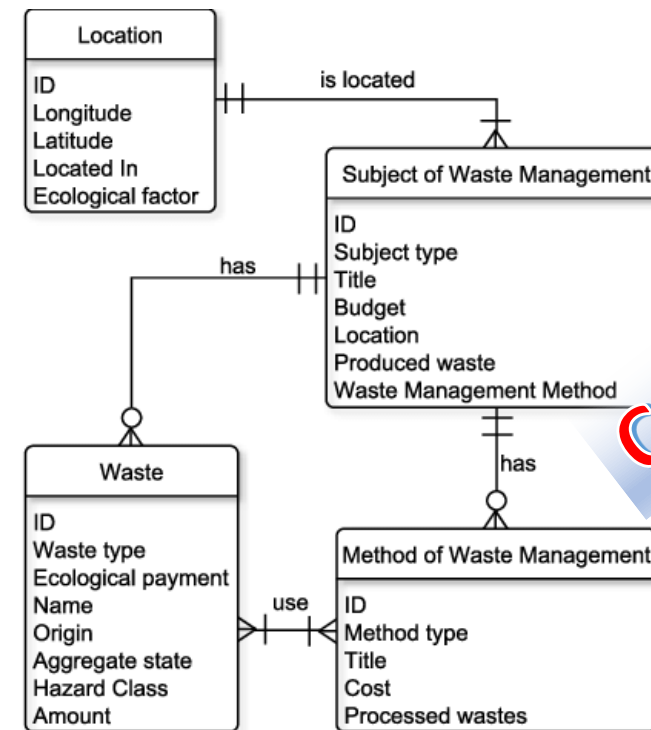
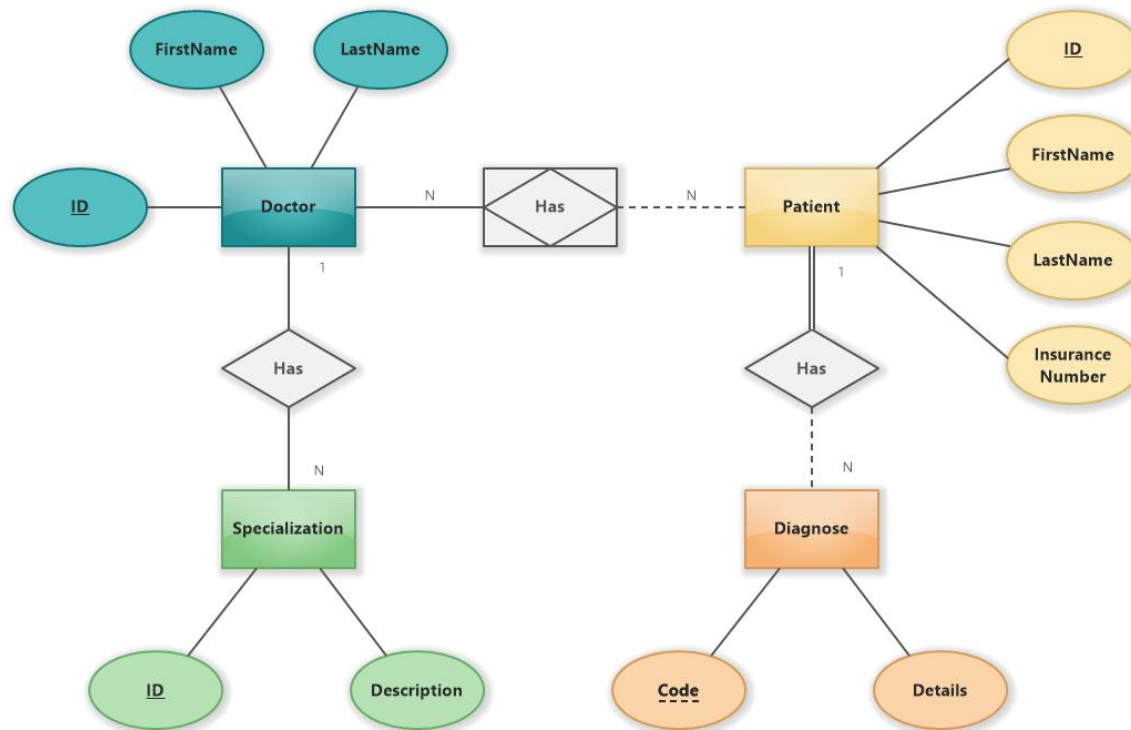
Sử dụng tập
ký hiệu này



Tập ký hiệu

- Có 2 tập ký hiệu: Chen, **Crow's foot**.

Sử dụng tập
ký hiệu này



Crow's foot

Thực thể (Entity)

- Là một phần tử trong mô hình tương ứng với một **lớp đối tượng** thuộc **đơn vị, tổ chức cần mô hình hóa**.
 - Có thể là người (**Sinh viên**), địa điểm (**Huyện**), sự kiện (**Hội nghị**), sự vật (**Sách**), ... có cùng đặc tính.
 - Định danh bằng tên, là **danh từ**.
 - Một thực thể **phải có nhiều thể hiện**.



Thuộc tính của thực thể

- Thuộc tính là một phần tử của mô hình tương ứng với một **đặc tính của một lớp đối tượng, cần mô hình hóa.**
- Định danh bằng tên, là **danh từ**.
 - Các thuộc tính khác nhau phải có tên khác nhau.
- Mỗi thuộc tính sẽ có một kiểu dữ liệu (**quan niệm**).
- Thuộc tính sẽ nhận giá trị từ một tập các giá trị phù hợp nào đó (**miền giá trị của thuộc tính**).
- *Thuộc tính được ghi bên trong, phía dưới tên thực thể.*



Thuộc tính của thực thể

- Ví dụ
 - **Họ tên**: kiểu chuỗi ký tự độ dài thay đổi
 - **Năm sinh**: kiểu số nguyên, [1962, 2004]
 - **Giới tính**: kiểu Yes/No, kiểu chuỗi ký tự.

Thể hiện của thực thể

- Một thể hiện của một thực thể là **một phần tử thuộc một lớp đối tượng** của tổ chức được mô hình hóa.
- Một thể hiện của một thực thể là **tổ hợp hợp lý các giá trị** của các thuộc tính của thực thể đó.

SINH_VIEN
<u>SV_MaSV</u>
SV_HoTen

Thể hiện: SV_MaSV = '1960547', SV_HoTen = 'Nguyễn Văn A'

Khóa của thực thể

CHUC_VU
<u>CV_Ma</u>
CV_Ten

- Là một tổ hợp (**thường là 1**) thuộc tính sao cho **một giá trị của tổ hợp thuộc tính này xác định duy nhất một thể hiện của thực thể.**
- Trong nhiều trường hợp, khóa của thực thể là **thuộc tính chỉ định** của thực thể đó.
 - *Thuộc tính chỉ định là thuộc tính được bổ sung để làm khóa, giá trị của thuộc tính này không có ý nghĩa trong thực tế.*
- Thuộc tính khóa của thực thể **sẽ được gạch dưới, thường được đặt ở vị trí đầu tiên.**

Tổng quát hóa – Chuyên biệt hóa

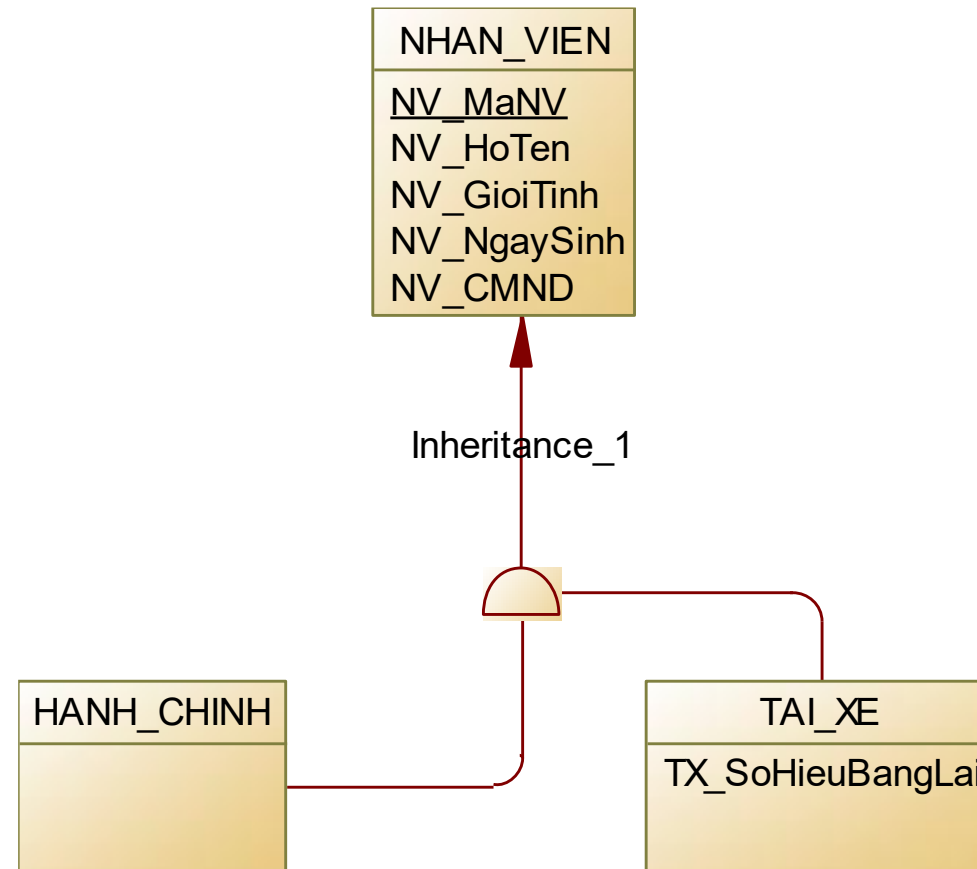
- Tổng quát hóa

- Hai hoặc nhiều thực thể có thể tổng quát hóa thành thực thể ở mức cao hơn nếu chúng có chung một số thuộc tính.

- Chuyên biệt hóa

- Một thực thể có thể được chuyên biệt hóa thành hai hoặc nhiều thực thể ở mức thấp hơn nếu như các thực thể ở mức thấp hơn có thuộc tính khác nhau.
- Thực thể ở mức cao là thực thể tổng quát, các thực thể mức thấp là thực thể chuyên biệt.
- Các thực thể chuyên biệt phải khác nhau về thuộc tính, hoặc liên kết (xét sau) với các thực thể khác.

Tổng quát hóa – Chuyên biệt hóa





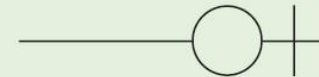
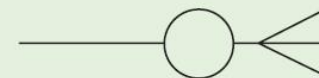
Liên kết

- Là sự kết hợp giữa các thể hiện của 2 thực thể.
- Liên kết được định danh bằng tên, và bắt đầu bằng một **động từ**.
- Thành phần không thể thiếu của liên kết là **cặp bản số** (Cardinality).

Bản số

- Biểu diễn **một thể hiện của thực thể này có thể kết hợp với bao nhiêu thể hiện của thực thể còn lại** đối với liên kết đang xét.
- Bản số bao gồm 2 thành phần
 - **Bản số tối thiểu**: là 0 hoặc 1 – số thể hiện tối thiểu của một thực thể có thể kết hợp với 1 thể hiện của thực thể còn lại.
 - **Bản số tối đa**: là 1 hoặc n – số thể hiện tối đa của một thực thể có thể kết hợp với 1 thể hiện của thực thể còn lại.

Bản số - Ký hiệu

Symbol	Meaning
	One—Mandatory
	Many—Mandatory
	One—Optional
	Many—Optional

Tối thiểu — (1, 1)

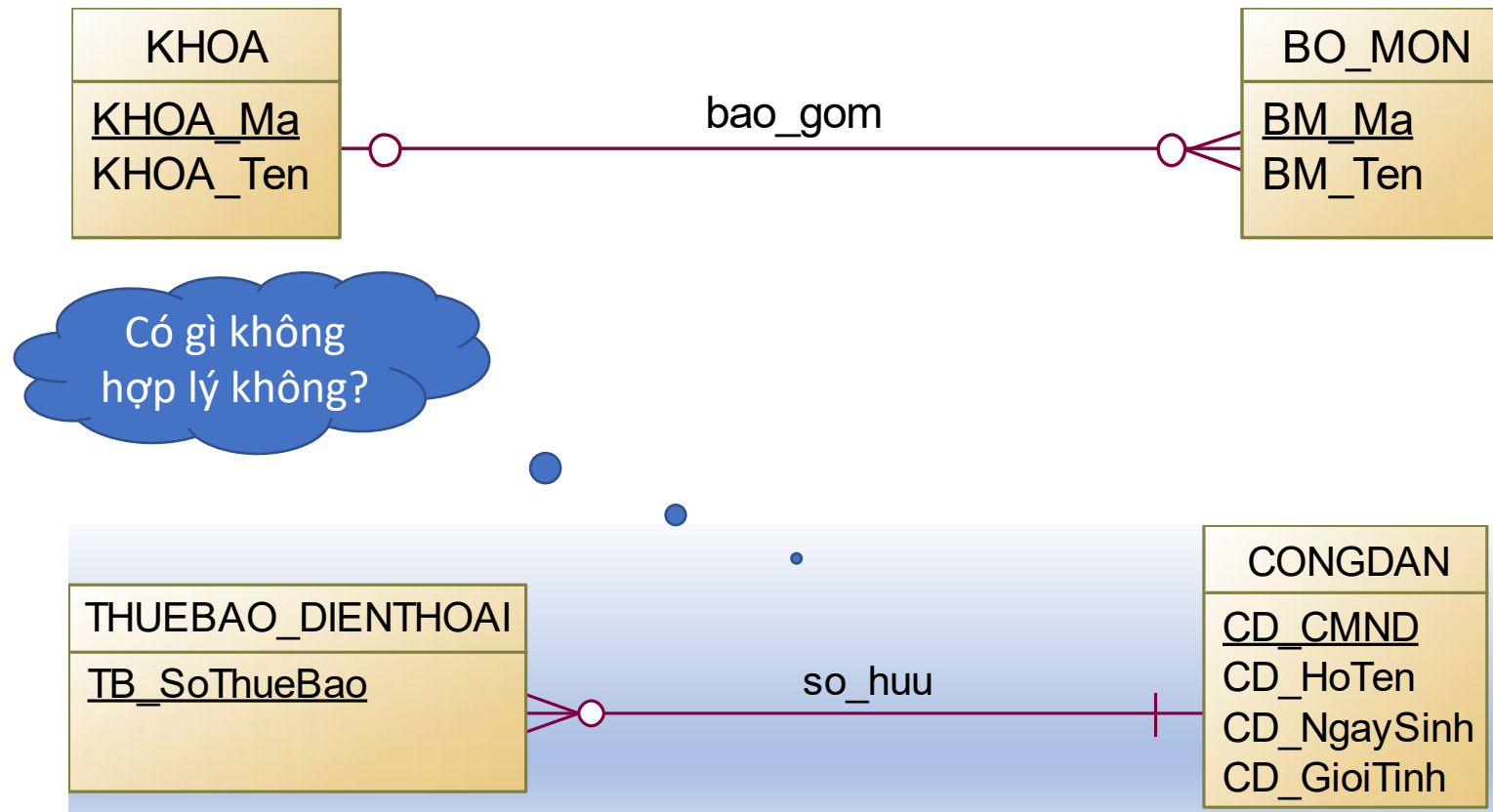
Tối đa — (1, n)

(1, n)

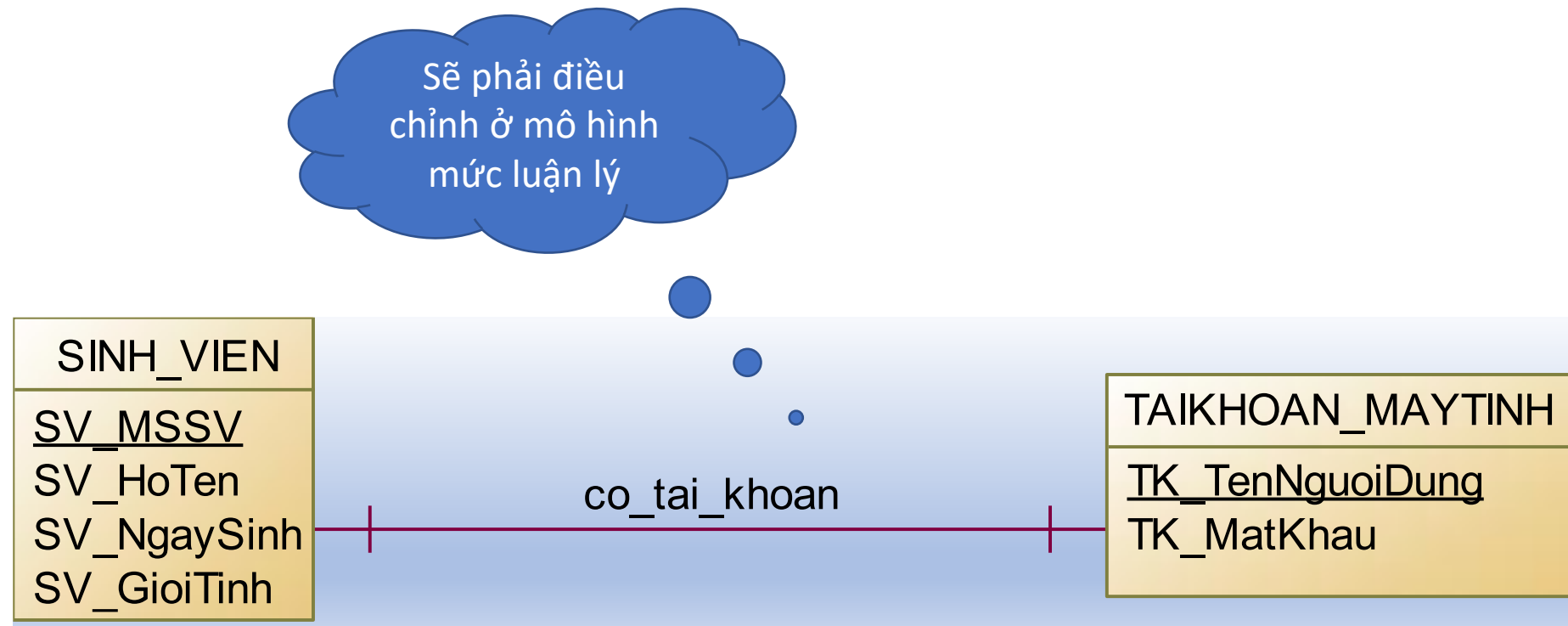
(0, 1)

(0, n)

Bản số - Ví dụ

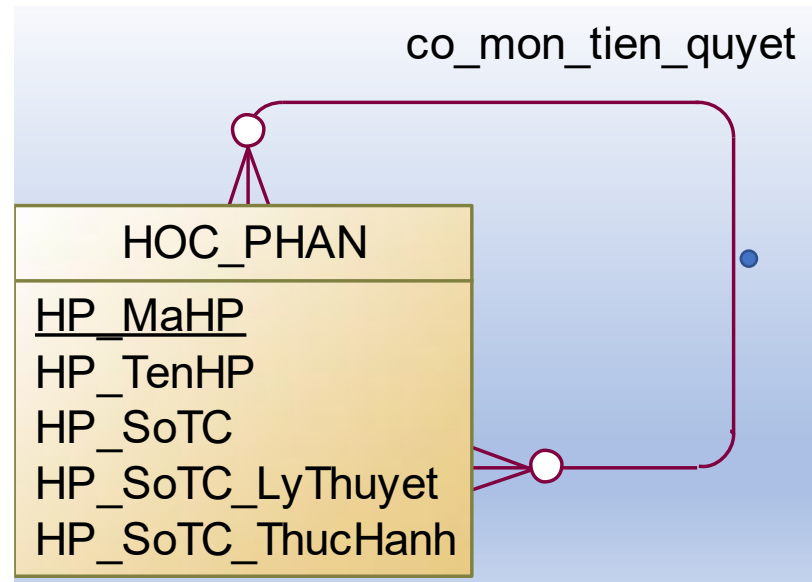


Bản số - Ví dụ



Liên kết tự thân

- Là sự kết hợp giữa các thể hiện của cùng 1 thực thể.



Phát biểu ý nghĩa
của liên kết

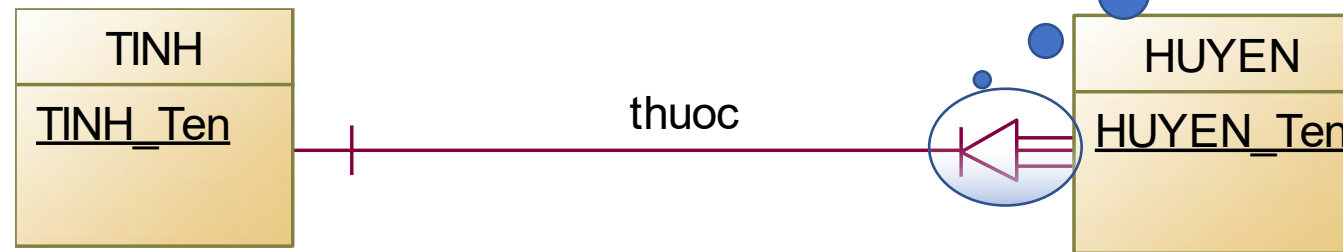
- Thiết kế mô hình cho **cây gia phả**, chỉ biểu diễn **mối quan hệ cha – con**.

Thực thể yếu (Weak entity)

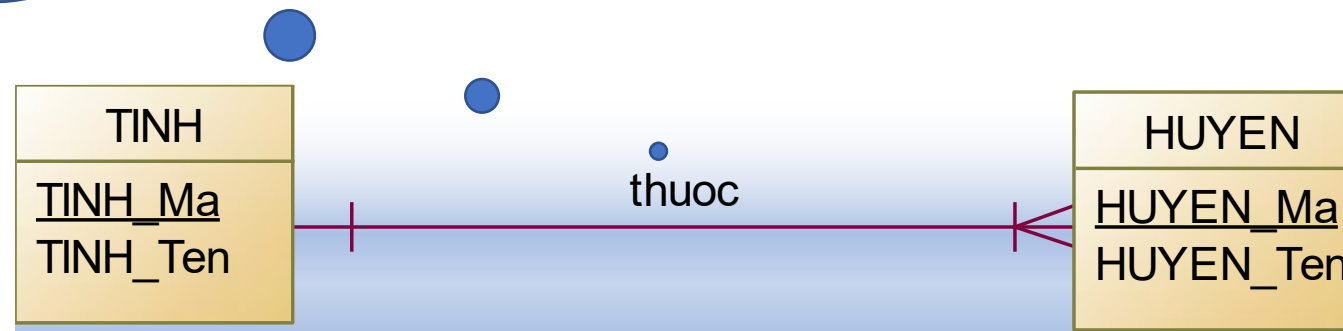
- Giá trị của thuộc tính khóa của thực thể yếu chưa đủ để xác định các thể hiện khác nhau của thực thể đó (nghĩa là giá trị thuộc tính khóa có thể trùng lại).
- Thực thể yếu có 3 cách dùng.
- **Cách 1**
 - Thực thể yếu phải phụ thuộc vào các thực thể khác.
 - Khóa của thực thể yếu sẽ kết hợp với khóa của các thực thể khác.

Thực thể yếu (Weak entity)

Bản số: (1, n), phụ thuộc



Cách này đúng không?



Liên kết có thuộc tính

- Mô hình dữ liệu biểu diễn **Hóa đơn mua những Sản phẩm (Gạo, Sữa, Muối ...)** gì?



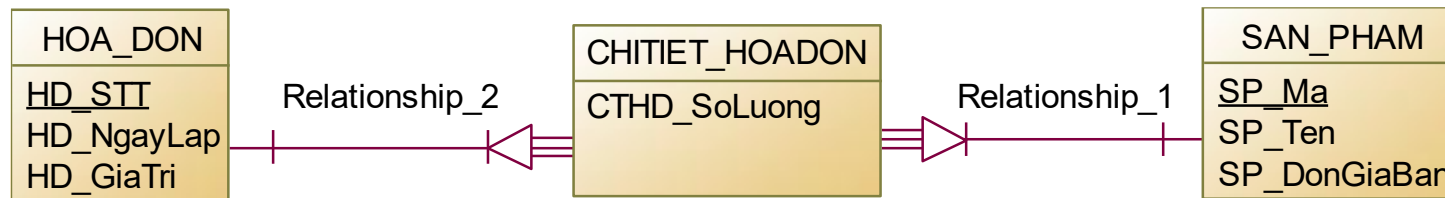
Liên kết có thuộc tính

- Mô hình dữ liệu biểu diễn **Hóa đơn mua Sản phẩm (Gạo, Sữa, Muối ...)** gì với số lượng là bao nhiêu?
 - Hóa đơn HD00012654 mua 2kg gạo ST25, 1kg đường, 1 chai dầu ăn.
 - Hóa đơn HD00012666 mua 3kg gạo ST25, 2 chai dầu ăn.
 - *1 hóa đơn có thể mua nhiều sản phẩm với số lượng mỗi sản phẩm khác nhau.*
 - *1 sản phẩm có thể được mua ở các hóa đơn khác nhau với số lượng khác nhau.*
- ⇒ số lượng thay đổi theo sản phẩm và hóa đơn ⇒ thuộc tính của liên kết giữa hóa đơn và sản phẩm.

Liên kết có thuộc tính

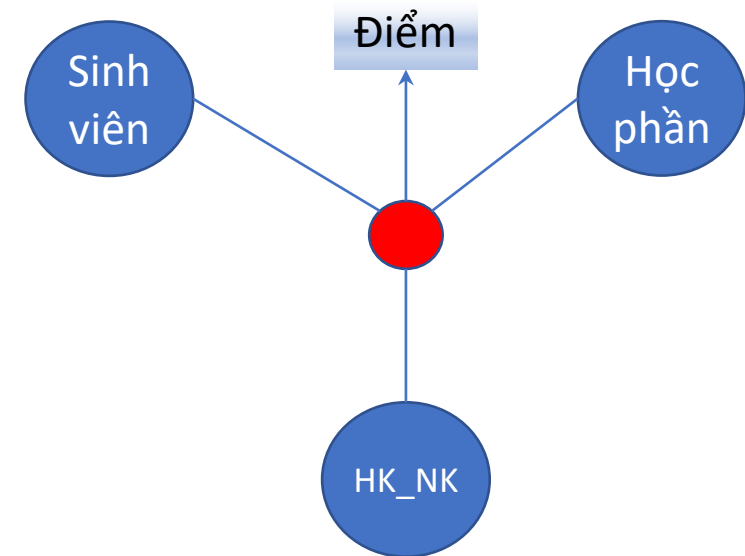
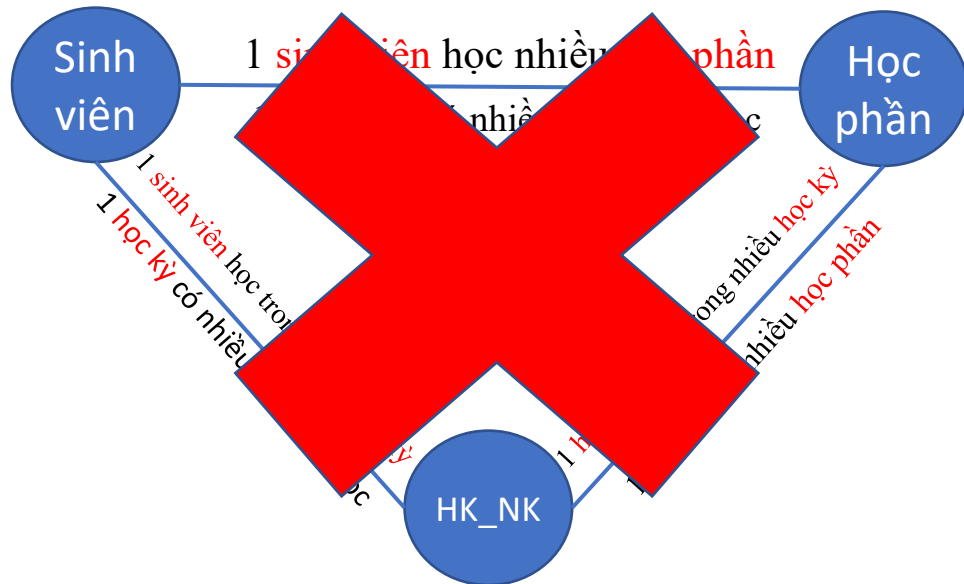
- Tập ký hiệu Crow's foot không cho phép biểu diễn liên kết có thuộc tính.
- Liên kết có thuộc tính sẽ được chuyển thành (cách sử dụng thứ 2 của thực thể yếu):
 - Thực thể yếu với thuộc tính là thuộc tính của liên kết.
 - Thực thể yếu phụ thuộc vào 2 thực thể liên quan đến liên kết đang xét.

Liên kết có thuộc tính



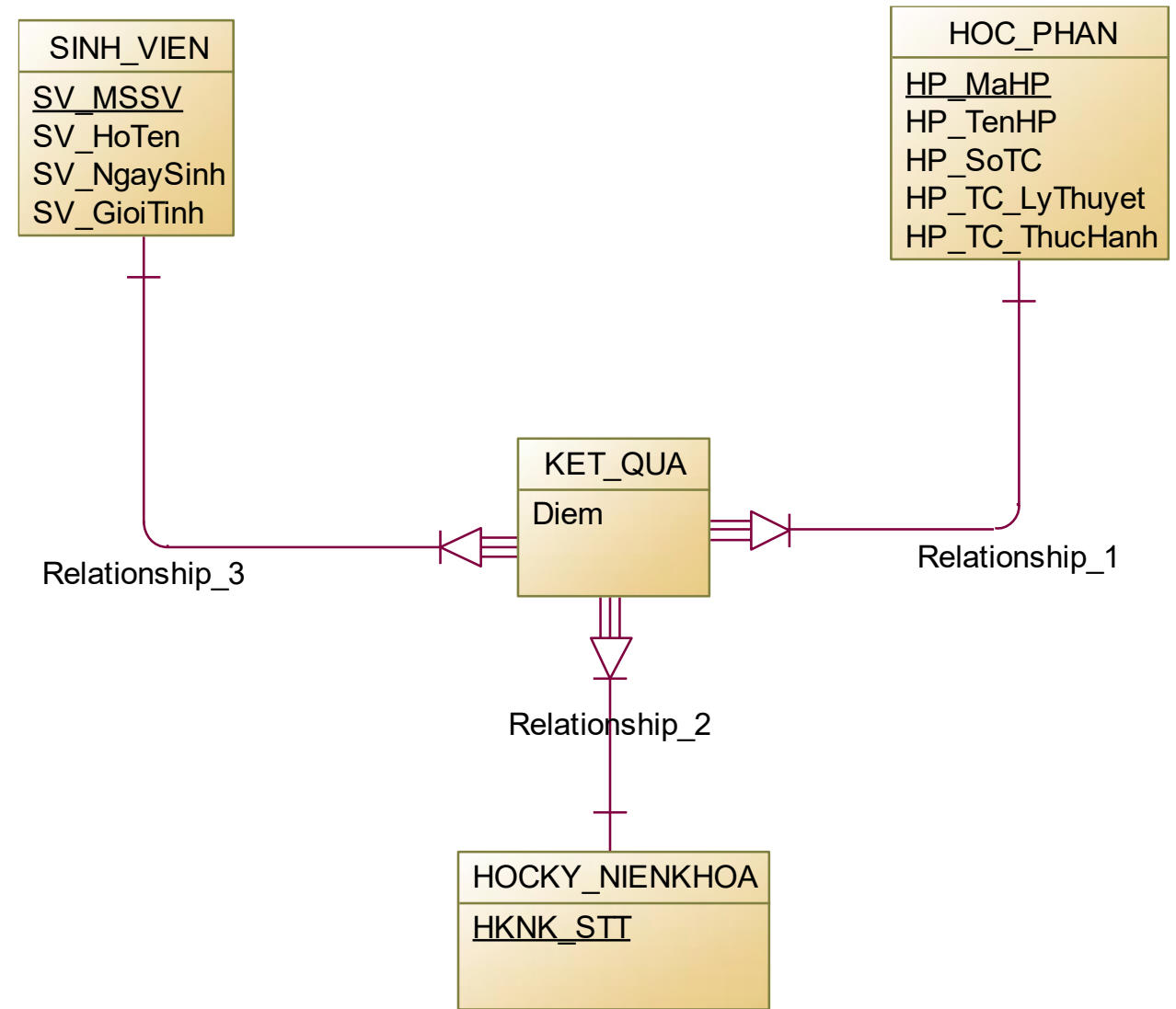
Liên kết nhiều chiều (liên kết nhiều hơn 2 thực thể)

- Giả sử chúng ta cần quản lý dữ liệu điểm của sinh viên, yêu cầu quản lý là cần biết Sinh viên nào học Học phần gì tại Học kỳ niên khóa bao nhiêu và đạt mấy điểm.



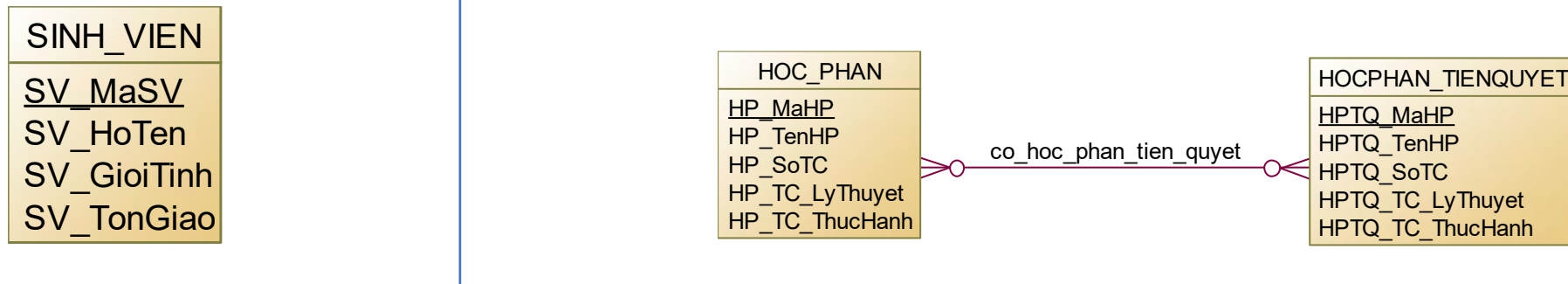
Liên kết nhiều chiều (liên kết nhiều hơn 2 thực thể)

- Liên kết nhiều chiều sẽ được chuyển thành (cách sử dụng thứ 3 của thực thể yếu)
- Thực thể yếu không có thuộc tính.
- Thực thể yếu phụ thuộc vào các thực thể liên quan đến liên kết đang xét.



Bài tập ứng dụng 1

- Tìm những điểm không hợp lý trong các sơ đồ sau



- Phân tích mô hình dữ liệu mức quan niệm cho hệ thống sau
 - <https://drive.google.com/file/d/1snXsFAsUfBQBhHP2fyuZtOBkq7hk1aMQ/view?usp=sharing>

Thiết kế mô hình dữ liệu mức luận lý

- Mô hình mức luận lý giúp 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).
 - Trong phạm vi học phần, **mô hình dữ liệu quan hệ** (Relational data model) được sử dụng để biểu diễn dữ liệu mức luận lý.
- ⇒ **chuyển mô hình Thực thể - Liên kết sang mô hình Quan hệ**

Quy tắc 1

- Thực thể chuyển thành quan hệ có cùng tên.
- Thuộc tính của thực thể chuyển thành thuộc tính của quan hệ.
- Khóa của thực thể là khóa chính của quan hệ.

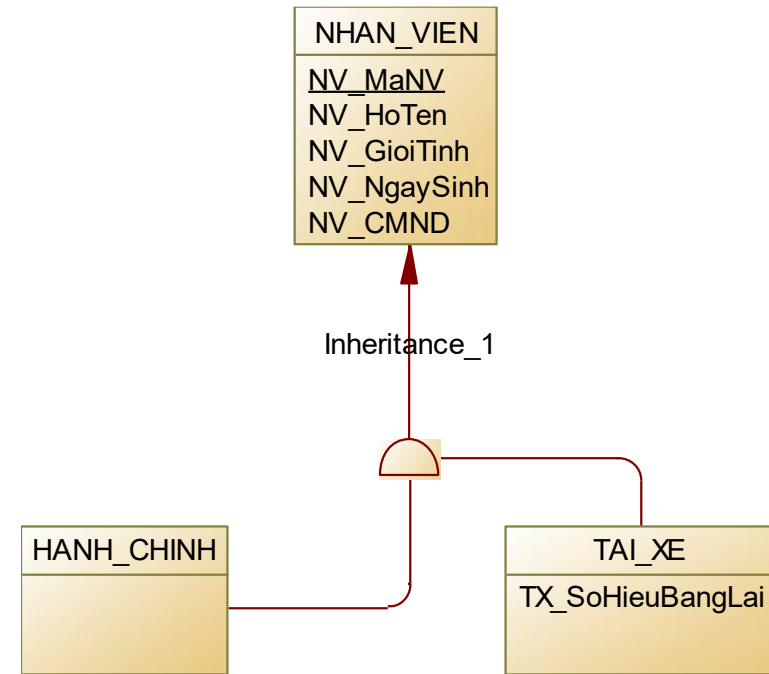
SINH_VIEN
<u>SV_MaSV</u>
SV_HoTen

⇒

SINH_VIEN (**SV_MaSV**, SV_HoTen)

Quy tắc 2

- Nếu mô hình có tồn tại **tổng quát hóa – chuyên biệt hóa** thì lựa chọn 1 trong các cách sau:
 - Cách 1
 - Không “quan tâm” đến thực thể tổng quát.
 - Các thực thể chuyên biệt trở thành quan hệ với tất cả các **thuộc tính chung** (thuộc tính của thực thể tổng quát) + **thuộc tính riêng** (thuộc tính của mình).
 - **Khóa chính** của các quan hệ là **thuộc tính khóa** của thực thể tổng quát.

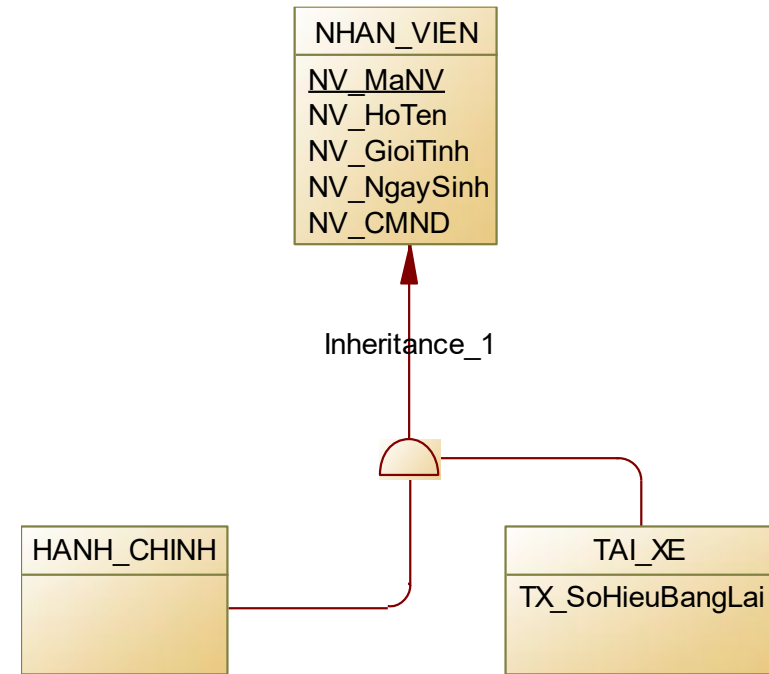


HANH_CHINH (NV_MaNV, NV_HoTen, NV_GioiTinh, NV_NgaySinh, NV_CMND)

TAI_XE (NV_MaNV, NV_HoTen, NV_GioiTinh, NV_NgaySinh, NV_CMND, TX_SoHieuBangLai)

Quy tắc 2

- Nếu mô hình có tồn tại **tổng quát hóa** – **chuyên biệt hóa** thì lựa chọn 1 trong các cách sau:
 - Cách 2
 - Chuyển thực thể tổng quát thành quan hệ.
 - Các thực thể chuyên biệt trở thành quan hệ với tất cả các thuộc tính riêng (thuộc tính của mình).
 - Khóa chính của các quan hệ là thuộc tính khóa của thực thể tổng quát.



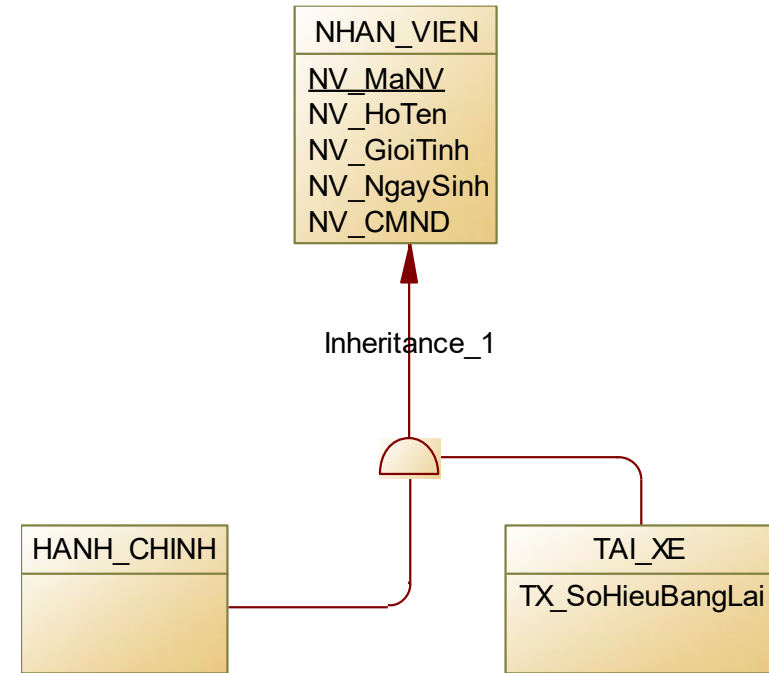
NHAN_VIEN (NV_MaNV, NV_HoTen, NV_GioiTinh, NV_NgaySinh, NV_CMND)

HANH_CHINH (NV_MaNV)

TAI_XE (NV_MaNV, TX_SoHieuBangLai)

Quy tắc 2

- Nếu mô hình có tồn tại **tổng quát hóa – chuyên biệt hóa** thì lựa chọn 1 trong các cách sau:
 - Cách 3
 - Xây dựng một quan hệ duy nhất, quan hệ này có tên của thực thể tổng quát, bao gồm tất cả các thuộc tính của thực thể tổng quát và tất cả các thuộc tính của các thực thể chuyên biệt.
 - Bổ sung một thuộc tính để phân biệt các thể hiện của các thực thể chuyên biệt.
 - *Khóa là thuộc tính khóa của thực thể tổng quát.*



NHAN_VIEN (NV_MaNV, NV_HoTen, NV_GioiTinh, NV_NgaySinh, NV_CMND, TX_SoHieuBangLai, NV_HanhChinh)

Quy tắc 3

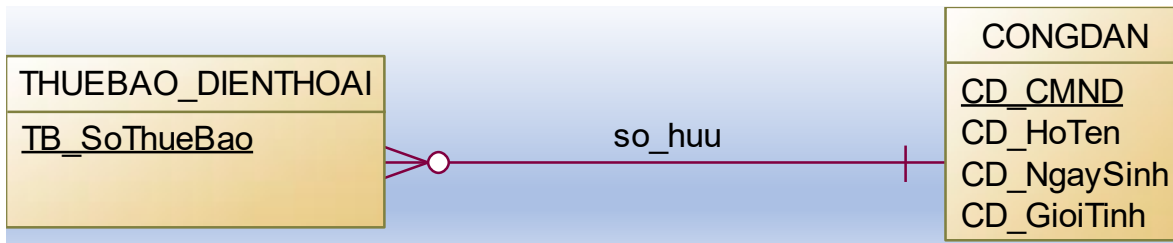
- Nếu là **thực thể yếu** thì thực thể này sẽ chuyển thành **quan hệ với tất cả các thuộc tính**.
- **Khóa chính của quan hệ** sẽ là **thuộc tính khóa của thực thể yếu và các thuộc tính khóa của các thực thể mà nó phụ thuộc vào**.



⇒ TINH (**TINH_Ten**)
⇒ HUYEN (**HUYEN_Ten**, TINH_Ten)

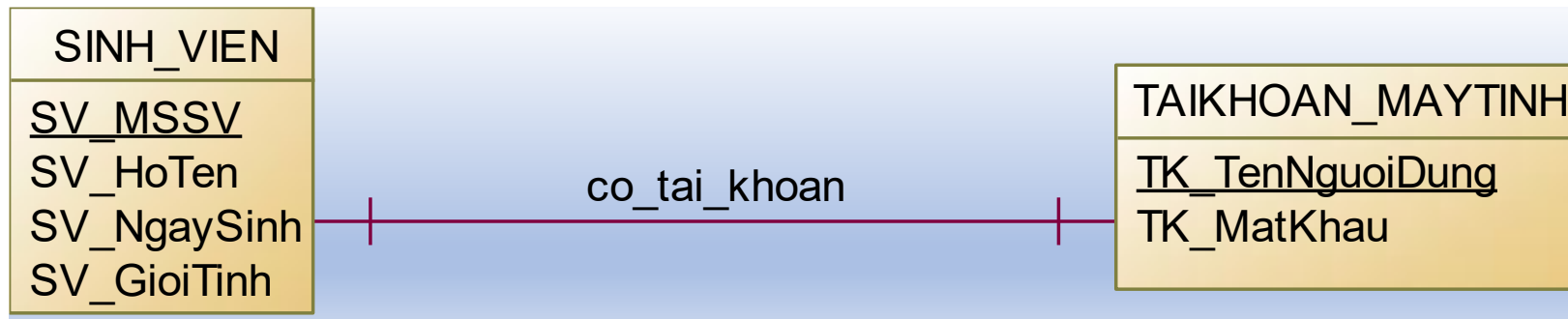
Quy tắc 4

- Với mỗi **phụ thuộc hàm mạnh** (1 bản số của liên kết là 1, 1)
 - Thực thể 2 là thực thể **gắn với bản số (1, 1)** của liên kết.
 - Thực thể 2 chuyển thành quan hệ.
 - *Chuyển thuộc tính khoá của thực thể 2 về làm khóa ngoại cho quan hệ tạo thành từ thực thể 1.*



- ⇒ THUEBAO_DIENHONAI (TB_SoThueBao, CD_CMND)
- ⇒ CONGDAN (CD_CMND, CD_HoTen, CD_NgaySinh, CD_GioiTinh)

Quy tắc 4

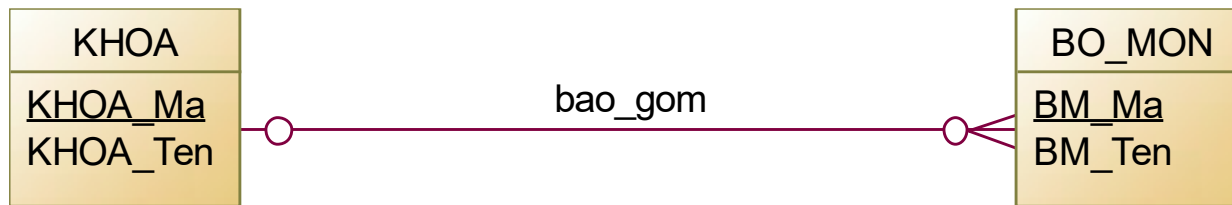


Có gì sai?

- ⇒ SINH_VIEN (SV_MSSV, SV_HoTen, SV_NgaySinh, SV_GioiTinh, *TK_TenNguoiDung*)
- ⇒ TAIKHOAN_MAYTINH (TK_TenNguoiDung, TK_MatKhai, *SV_MSSV*)

Quy tắc 5

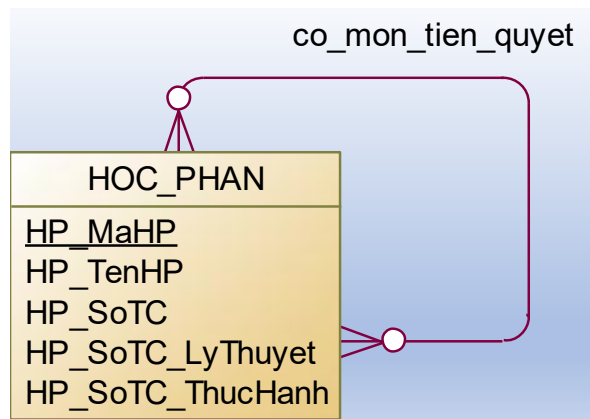
- Với mỗi **phụ thuộc hàm yếu** (1 bản số của liên kết là 0, 1) thì:
 - Thực hiện theo quy tắc 4.*
 - Hoặc **tạo quan hệ mới** (tên quan hệ là tên của liên kết) **gồm 2 thuộc tính là 2 thuộc tính khóa của 2 thực thể tương ứng**. **Khóa chính của quan hệ mới là 2 thuộc tính này.**



- ⇒ KHOA (**KHOA_Ma**, KHOA_Ten)
- ⇒ BO_MON (**BM_Ma**, BM_Ten)
- ⇒ BAO_GOM (**KHOA_Ma**, **BM_Ma**)

Quy tắc 6

- Liên kết có **hai bản số đều** thuộc dạng $(1, n)$ hoặc $(0, n)$.
- *2 thực thể chuyển thành 2 quan hệ theo quy tắc 1.*
- Thêm **quan hệ mới gồm 2 thuộc tính khóa** của hai thực thể tương ứng.
- *Khóa chính của quan hệ mới là tổ hợp hai thuộc tính khóa.*



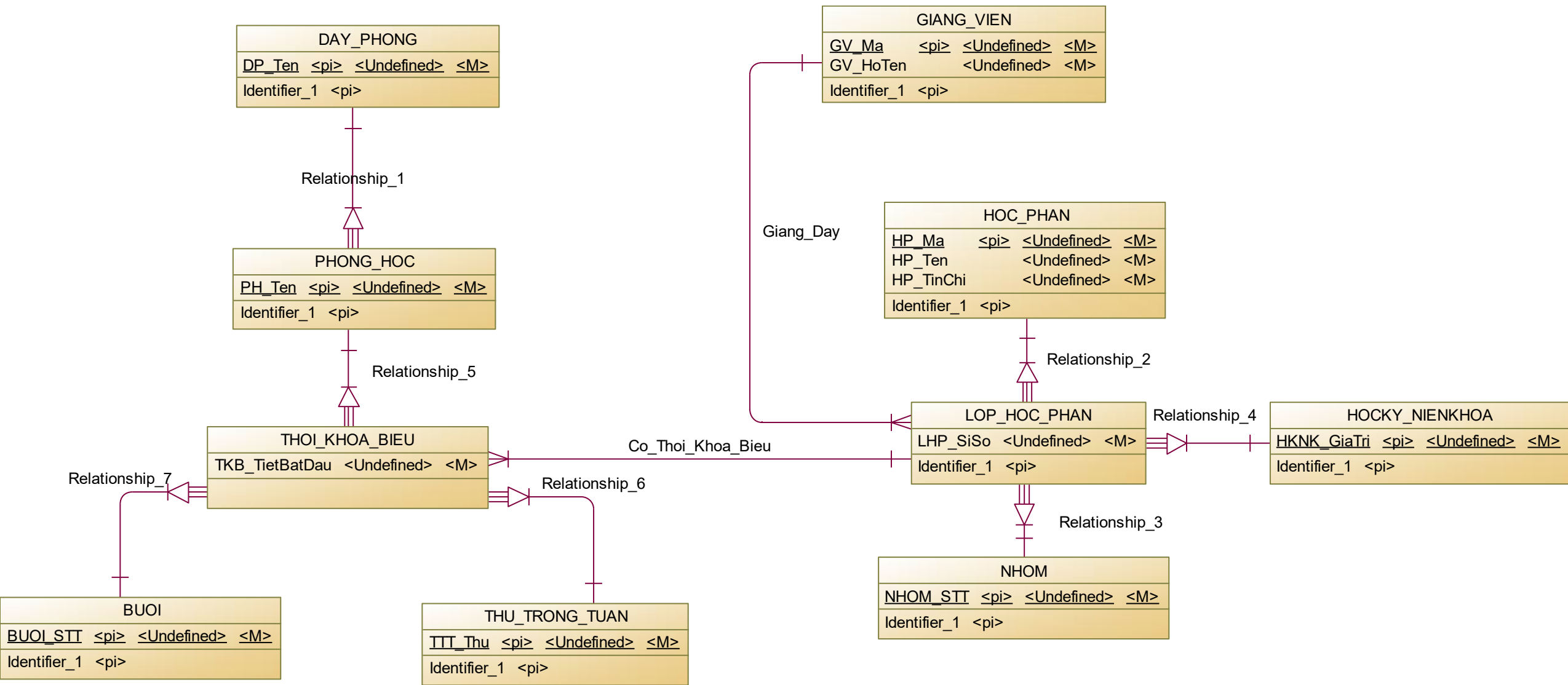
⇒ HOC_PHAN (**HP_MaHP**, HP_TenHP, HP_SoTC, HP_SoTC_LyThuyet, HP_SoTC_ThucHanh)
 ⇒ CO_MON_TIEN_QUYET (**HP_MaHP, HP_MaHP_TienQuyet**)

Bài tập ứng dụng 2

- Thiết kế sơ đồ E-R để biểu diễn dữ liệu TKB giảng dạy như minh họa.

Thứ	Mã HP	Ký hiệu	Tên Học phần	Lớp học phần	Sĩ Số	Tiết Học	Phòng	Tuần Dạy 123456789012345
Thời Khóa Biểu Giảng Dạy								
2	CT296	03	Phân tích và thiết kế hệ thống thông tin	CT29603	37	123-----	401/TS	123456789012345
2	CT109	01	Phân tích và thiết kế hệ thống thông tin	CT10901	35	-----678-----	401/D1	123456789012345
3	CT308H	M02	Thương mại điện tử	DI18V7F2	40	123-----	CLC1DI	123456789012345
3	CT109	02	Phân tích và thiết kế hệ thống thông tin	CT10902	35	-----678-----	404/D1	123456789012345
4	CT308H	M01	Thương mại điện tử	DI18V7F1	40	123-----	CLC1DI	123456789012345
4	CT109	03	Phân tích và thiết kế hệ thống thông tin	CT10903	35	-----678-----	305/D1	123456789012345
5	CT296	01	Phân tích và thiết kế hệ thống thông tin	CT29601	37	123-----	202/TS	123456789012345
5	CT296	04	Phân tích và thiết kế hệ thống thông tin	CT29604	37	-----678-----	109/C1	123456789012345
6	CT296	02	Phân tích và thiết kế hệ thống thông tin	CT29602	37	123-----	403/TS	123456789012345
6	CT296	05	Phân tích và thiết kế hệ thống thông tin	CT29605	37	-----678-----	404/D1	123456789012345

- Chuyển sơ đồ E-R thiết kế sang mô hình quan hệ.



Thiết kế mô hình dữ liệu mức vật lý

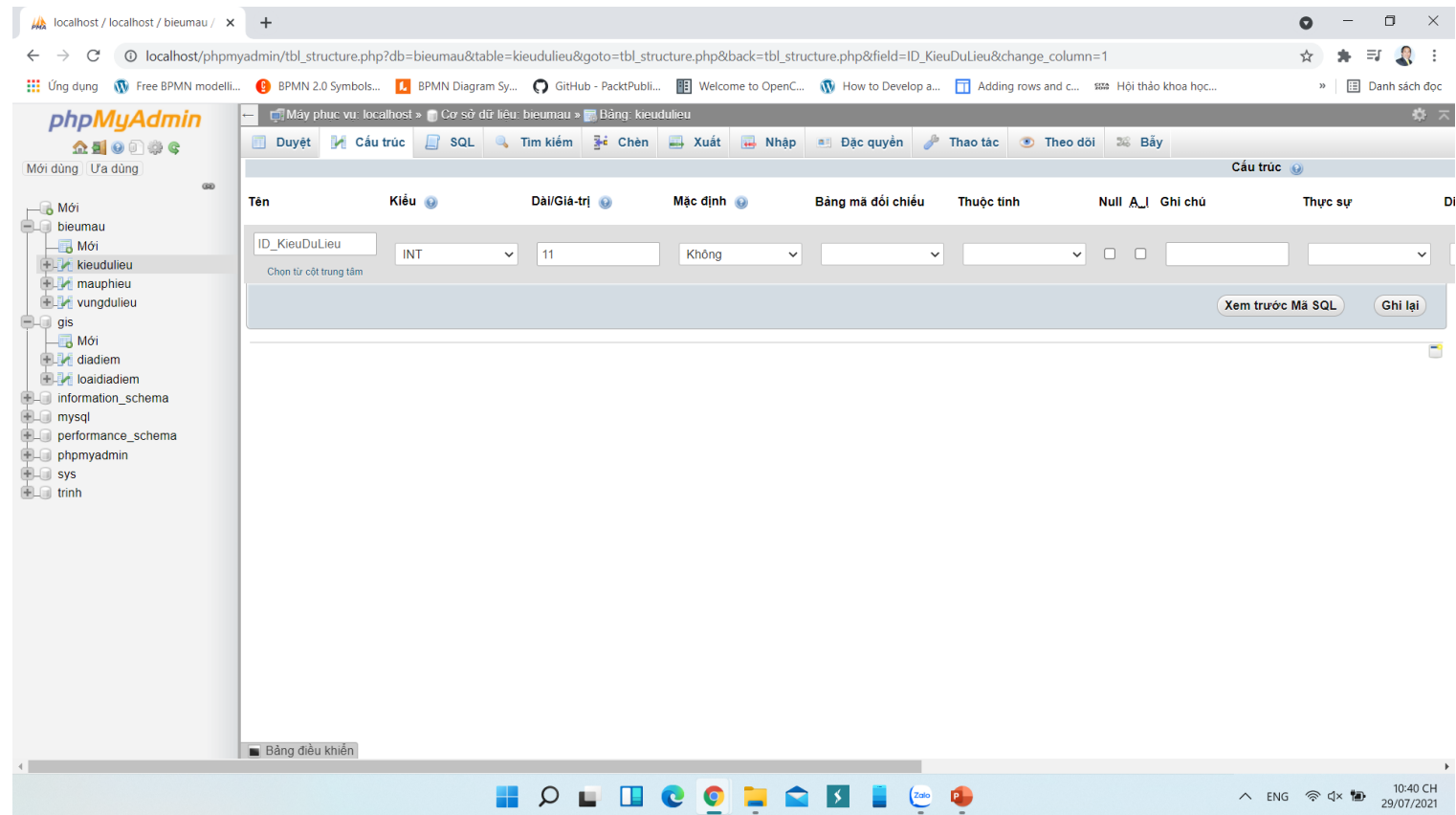
- Cấu trúc CSDL dựa trên lược đồ CSDL quan hệ thu được ở bước thiết kế mô hình mức luận lý.
- Lựa chọn hệ quản trị CSDL phù hợp.

Rank			DBMS	Database Model	Score		
Jul 2021	Jun 2021	Jul 2020			Jul 2021	Jun 2021	Jul 2020
1.	1.	1.	Oracle +	Relational, Multi-model i	1262.66	-8.28	-77.59
2.	2.	2.	MySQL +	Relational, Multi-model i	1228.38	+0.52	-40.13
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model i	981.95	-9.12	-77.77
4.	4.	4.	PostgreSQL +	Relational, Multi-model i	577.15	+8.64	+50.15
5.	5.	5.	MongoDB +	Document, Multi-model i	496.16	+7.95	+52.68
6.	↑ 7.	↑ 8.	Redis +	Key-value, Multi-model i	168.31	+3.06	+18.26
7.	↓ 6.	↓ 6.	IBM Db2	Relational, Multi-model i	165.15	-1.88	+1.99
8.	8.	↓ 7.	Elasticsearch +	Search engine, Multi-model i	155.76	+1.05	+4.17
9.	9.	9.	SQLite +	Relational	130.20	-0.33	+2.75
10.	↑ 11.	10.	Cassandra +	Wide column	114.00	-0.11	-7.08

Chi tiết xem tại https://en.wikipedia.org/wiki/Comparison_of_relational_database_management_systems

Thiết kế mô hình dữ liệu mức vật lý

- Chọn **định dạng lưu trữ** cho **mỗi thuộc tính** của mô hình cơ sở dữ liệu vật lý.
 - 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 ...



Thiết kế mô hình dữ liệu mức vật lý

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

Phân loại dữ liệu

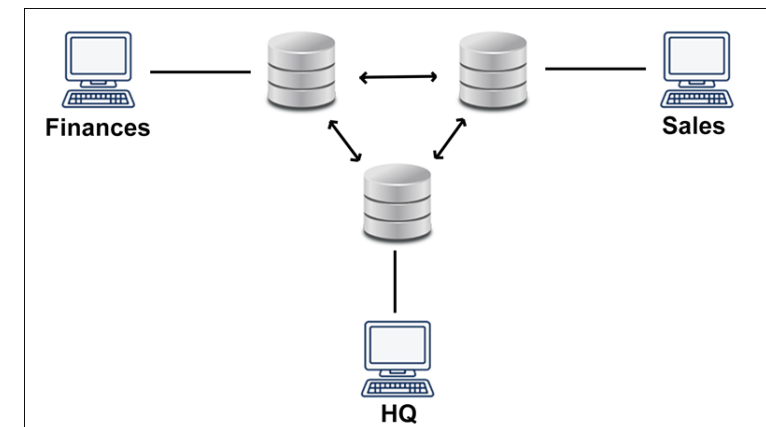
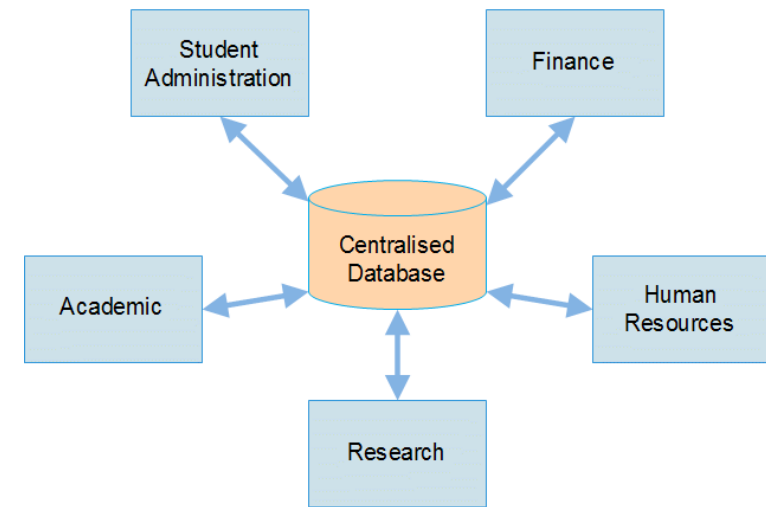
- Dữ liệu thường trực

- Dữ liệu này thường không biến đổi hay nói cách khác là rất ít biến đổi theo thời gian.
- Trong hệ thống thông tin chúng tồn tại vĩnh cửu và không thay đổi giá trị.
- 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 có được xóa không? Tại sao?*

Phân loại dữ liệu

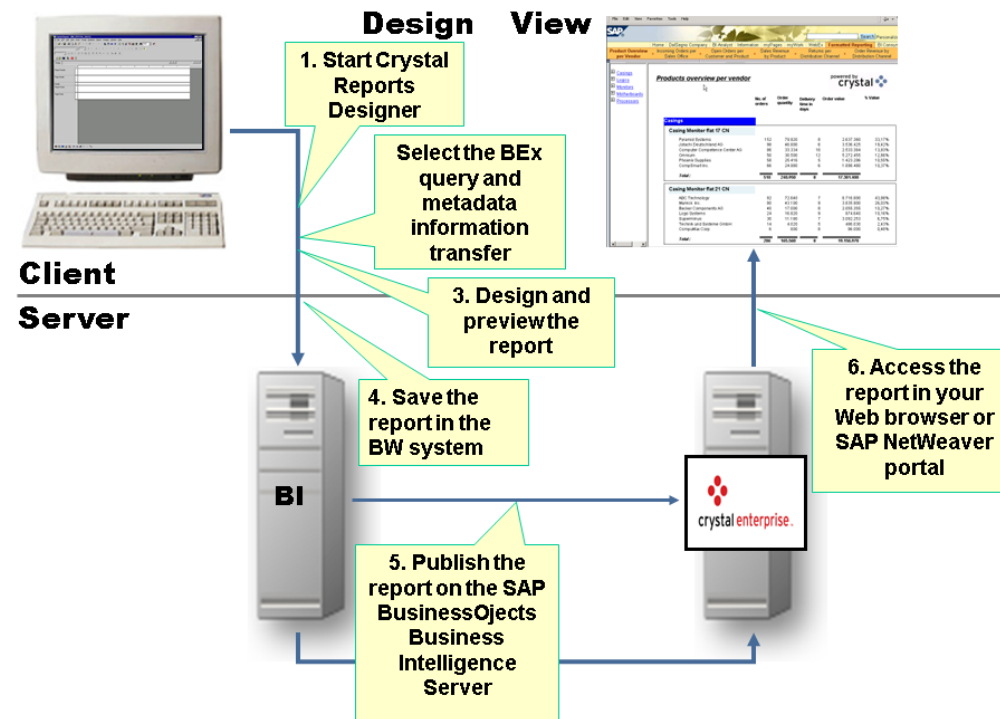
- Dữ liệu gốc

- 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?



Phân loại dữ liệu

- **Dữ liệu tạm thời**
 - Chỉ dùng trong một thời gian nào đó, khi không cần có thể xóa đi.



Mô tả bảng dữ liệu

TT	Tên	Kiểu	Kích thước	Số số lẻ	Miền giá trị	Trị mặc nhiên	MIN	MAX	Khóa chính	Duy nhất	NOT NULL	RB TV Luận Lý	RBTV Khóa ngoại	Diễn giải
1	CD_STT	Int			$[1, \infty)$		1		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_TrinhDoVanHoa	Int			$[1, 12]$						X			Trình độ văn hóa