



CƠ SỞ DỮ LIỆU

BÀI 3: MÔ HÌNH ERD VÀ CHUẨN HOÁ CƠ SỞ DỮ LIỆU

❑ Kết thúc bài học này bạn có khả năng

- ⊙ Mô hình hoá dữ liệu – Data modeling
- ⊙ Tìm hiểu về mô hình thực thể - liên kết (ERD) và các thành phần của nó
- ⊙ Từ mô hình ERD hướng đến xây dựng lược đồ CSDL quan hệ.
- ⊙ Giới thiệu về chuẩn hoá dữ liệu
- ⊙ Quy trình chuẩn hoá dữ liệu
- ⊙ Xây dựng CSDL mức vật lí

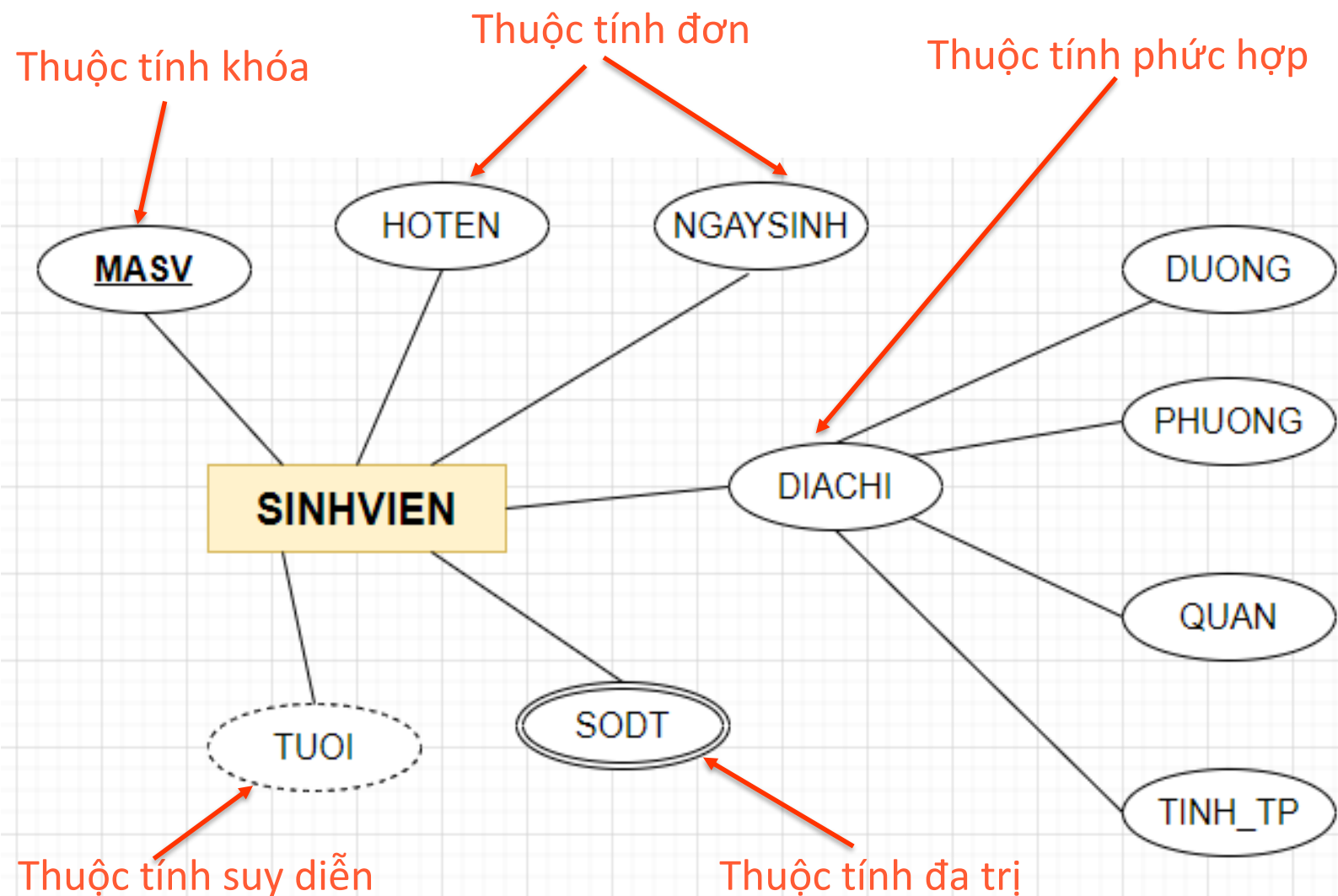


MÔ HÌNH ERD

- Cố gắng tạo ra một biểu diễn của thế giới thực
 - Bỏ qua một số ít sự phức tạp của thế giới thực
 - Sự đơn giản dựa vào một tập nhỏ các ký hiệu
- Cố gắng rút gọn tổ chức dữ liệu thành sự mô tả của các thực thể và các mối liên hệ giữa chúng
 - Sự mô tả của các yêu cầu thông tin mà máy tính có thể sử dụng
- Tiến trình mô hình hóa độc lập với nền tảng phát triển (phần mềm)
- Mô hình được sử dụng để trao đổi giữa người thiết kế CSDL và người sử dụng CSLD đó

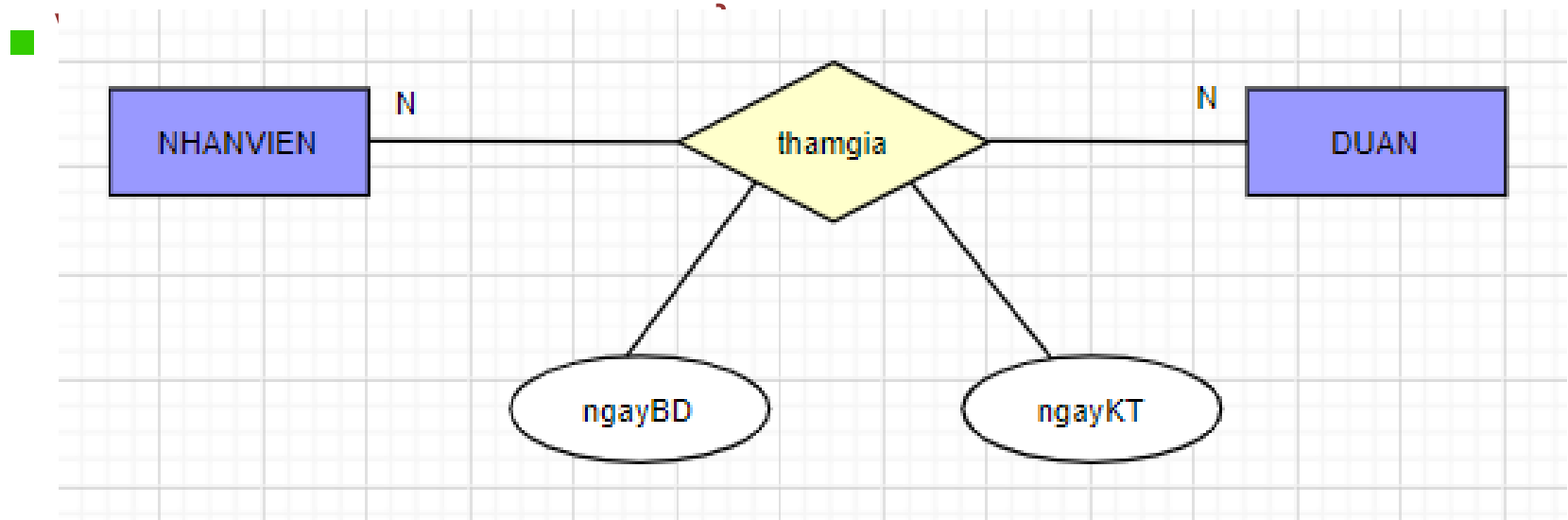
- **Thuộc tính đơn**
Không thể tách nhỏ ra được
- **Thuộc tính kết hợp**
Có thể tách thành nhiều thuộc tính nhỏ hơn
- **Thuộc tính đơn trị**
Có giá trị duy nhất cho một thực thể (vd: số CMND)
- **Thuộc tính đa trị**
Có nhiều giá trị khác nhau ở cùng 1 thực thể (vd: số đt)
- **Thuộc tính suy diễn**
Giá trị của nó được suy ra từ thuộc tính khác (vd: tuổi)
- **Mỗi thực thể đều được phân biệt bởi thuộc tính khóa**

TÌM HIỂU CÁC LOẠI THUỘC TÍNH



- Mỗi quan hệ: Là sự liên kết giữa 2 hay nhiều tập thực thể
- Có các loại mối quan hệ như sau:
 - Quan hệ 1-1 (một - một)
 - Quan hệ 1-n (một - nhiều)
 - Quan hệ n-n (nhiều - nhiều)
- Ví dụ: giữa tập thực thể NHÂNVIEN và PHONGBAN có các mối quan hệ:
 - Một nhân viên thuộc một phòng ban nào đó
 - Một phòng ban có một nhân viên làm trưởng phòng

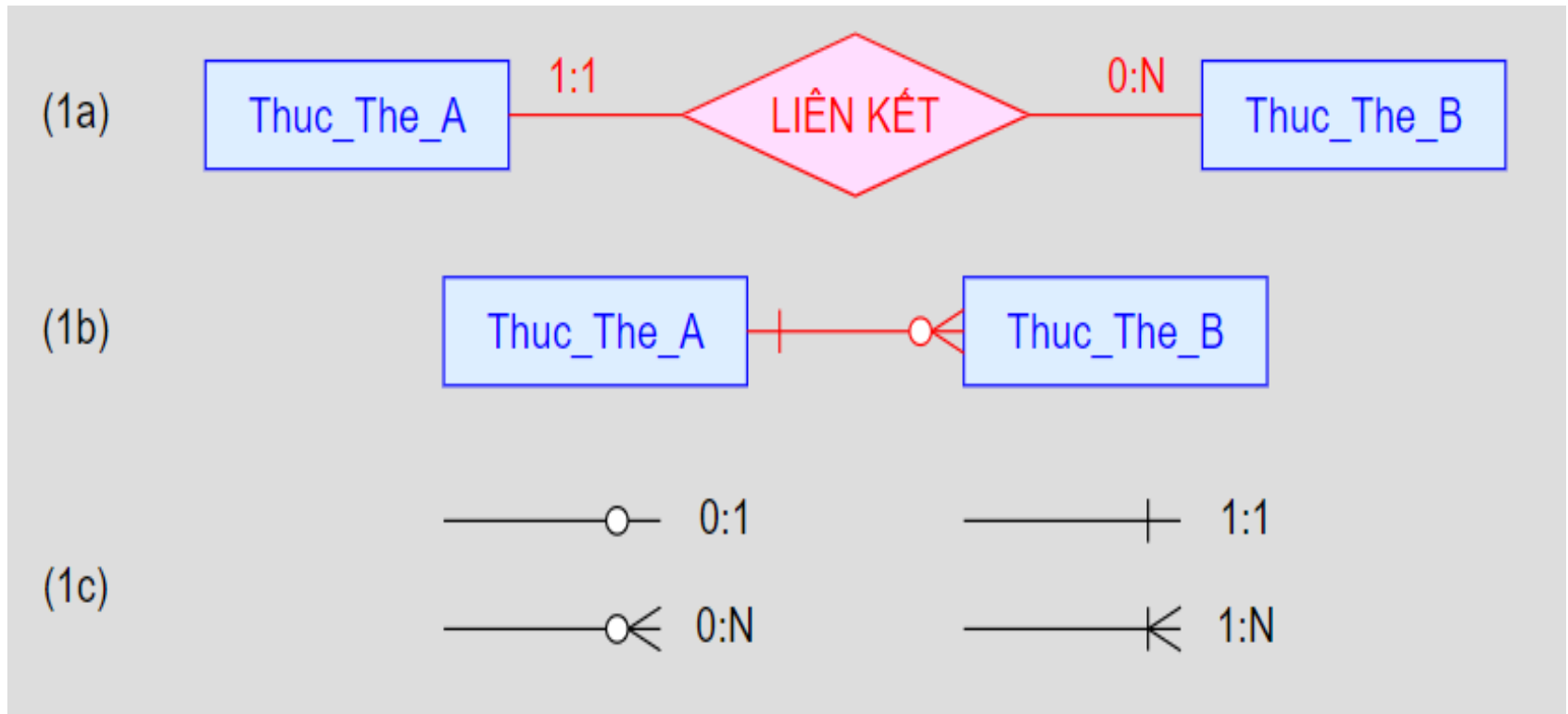
- Thuộc tính trên mối quan hệ mô tả tính chất cho mối quan hệ đó.
- Thường gặp trên mối quan hệ của hai thực thể có quan hệ N-N



- Còn được gọi là thuộc tính định danh của tập thực thể
- Dùng để phân biệt giữa các thực thể khác nhau trong tập thực thể
- Mỗi tập thực thể phải có 1 khóa
- Một khóa có thể có 1 hay kết hợp nhiều thuộc tính

- **ERD** thể hiện tất cả các **mối quan hệ** của các **Tập thực thể** và trên cùng 1 sơ đồ
- **Thực thể** được biểu diễn bằng một **hình chữ nhật** có ghi tên của thực thể (*Hình 1a và 1b, slide 10*).
- **Liên kết** được biểu diễn bằng một **hình thoi** nối đến các thực thể tham gia liên kết (*Hình 1b, slide 10*).
- **Bản số** của thực thể tham liên kết chỉ số thực thể tối đa tham gia vào liên kết
 - Được biểu diễn bằng một chữ số (0,1 hay N)
 - Hoặc cặp số như 0:N lần lượt là bản số tối thiểu và tối đa (*Hình 1a, slide 10*).

■ Quy ước các biểu diễn trong ERD:



Hình 1 Một số quy ước về lược đồ của mô hình dữ liệu thực thể - liên kết

■ Quy ước các biểu diễn trong ERD:

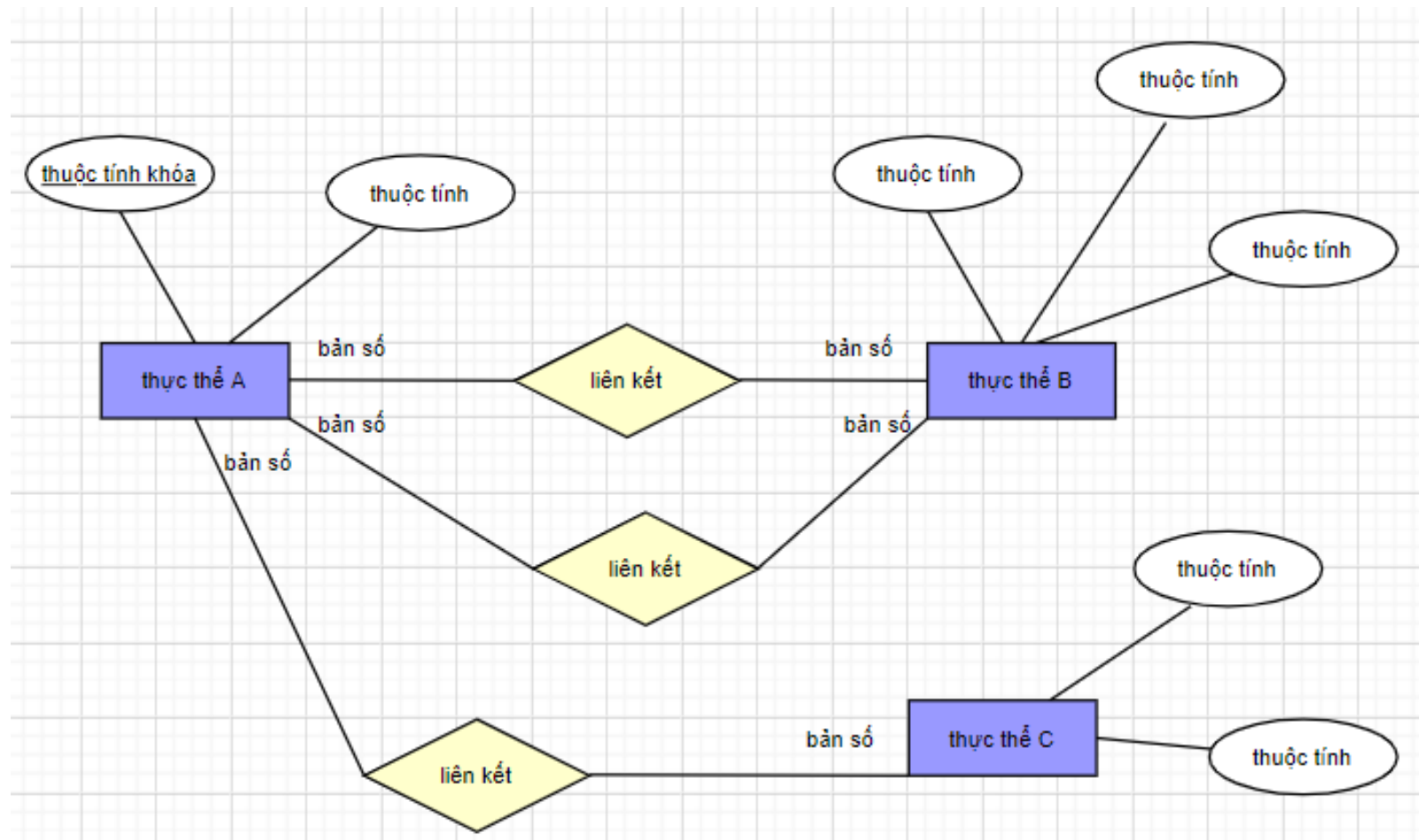
CARDINALITY

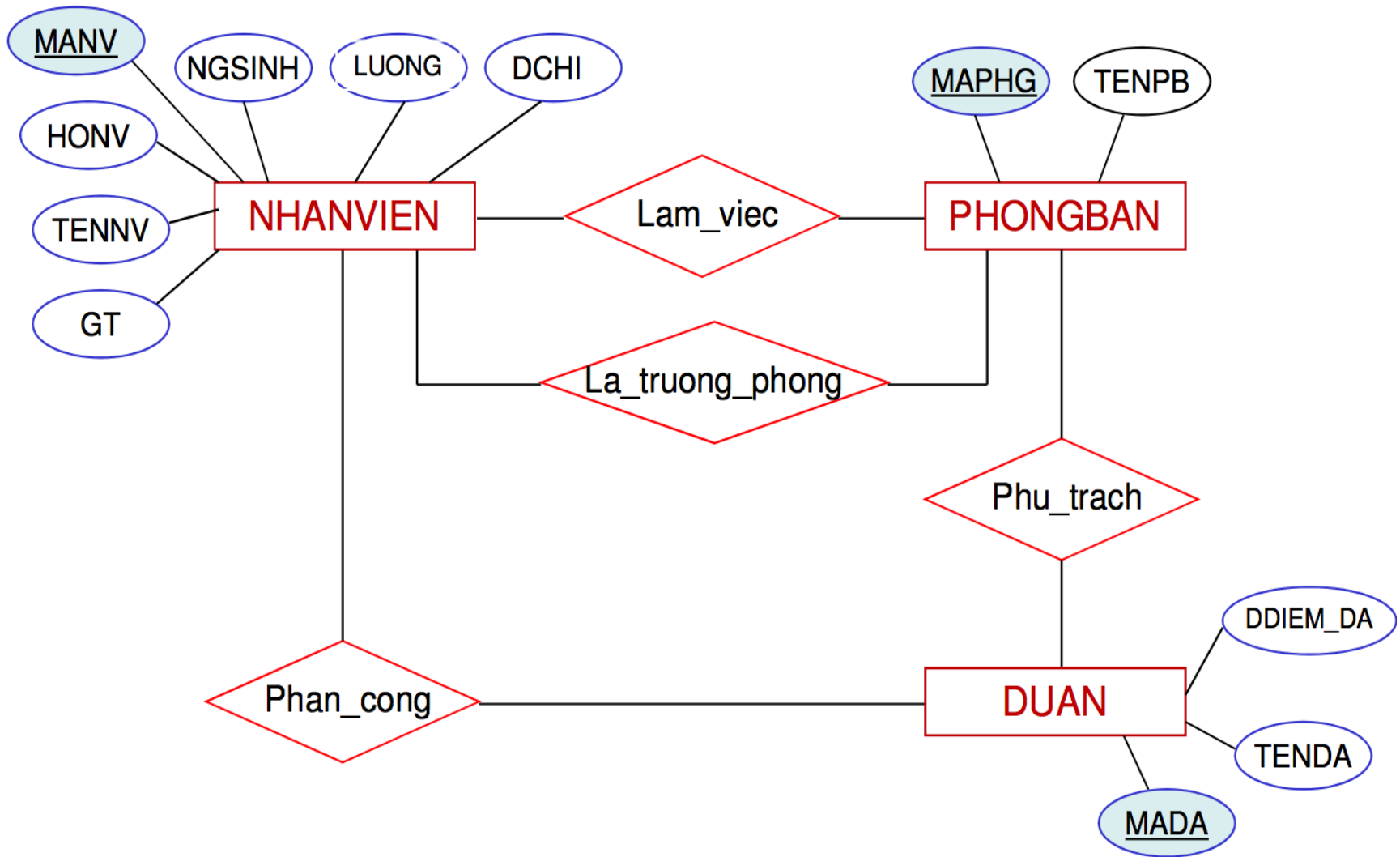
1/	—————	Một
2/	—————	Một và bắt buộc phải có một
3/	————— ○	Không hoặc một
4/	—————<	Nhiều
5/	—————<	Một hoặc nhiều
6/	————— ○<	Không hoặc nhiều

CÓ THỂ HIỂU

<i>Đơn thuần là một</i>
<i>Required (chỉ được một, và bắt buộc có một dữ liệu)</i>
<i>Optional (không có dữ liệu cũng được)</i>
<i>Đơn thuần là nhiều</i>
<i>Required (ít nhất phải có một dữ liệu)</i>
<i>Optional (không có dữ liệu cũng được)</i>

- Dùng diagrams.net để vẽ lược đồ ERD





THIẾT KẾ CƠ SỞ DỮ LIỆU Ở MỨC LOGIC

- Khóa chính còn được gọi là thuộc tính định danh của một bảng.
- Khóa chính có thể một hoặc nhiều thuộc tính dùng để phân biệt giữa các thực thể khác nhau trong một quan hệ.
- Ví dụ:
 - Quan hệ **Nhân Viên** có thuộc tính khóa là Mã NV
 - Quan hệ **Phân Công** có khóa chính là: Mã NV, Mã CV

- Khi một quan hệ kết nối được với một quan hệ khác, luôn tồn tại một/một số thuộc tính đóng vai trò là cột dữ liệu chung kết nối hai quan hệ .
- **Khóa ngoại (Foreign Key/Reference Key):** là một/một số thuộc tính của quan hệ này có liên kết tham chiếu với quan hệ kia.
- Các thuộc tính khóa ngoại của quan hệ này là khóa chính của quan hệ kia, do đó các giá trị của cột khóa ngoại phải phù hợp với những giá trị trong cột khóa chính của quan hệ tham chiếu

- Chuyển hóa từ sơ đồ thực thể - liên kết ở mức khái niệm → lược đồ quan hệ ở mức logic.
- Quá trình chuyển hóa này gồm:
 - Chuyển hóa các thực thể.
 - Chuyển hóa các liên kết

- Chuyển hóa các thực thể thường tương đối đơn giản. Phương pháp chung là:
 - **Mỗi thực thể** từ sơ đồ thực thể - liên kết sẽ chuyển thành **một quan hệ** trong mô hình logic.
 - Với mỗi quan hệ, ta xác định các **thuộc tính** của nó.
 - Xác định **thuộc tính định danh** và chọn thuộc tính này làm **khóa chính**.

Với hai quan hệ tương ứng với hai thực thể đã chuyển hóa. Tiếp tục chuyển hóa các liên kết.

- Đối với liên kết 1-N:

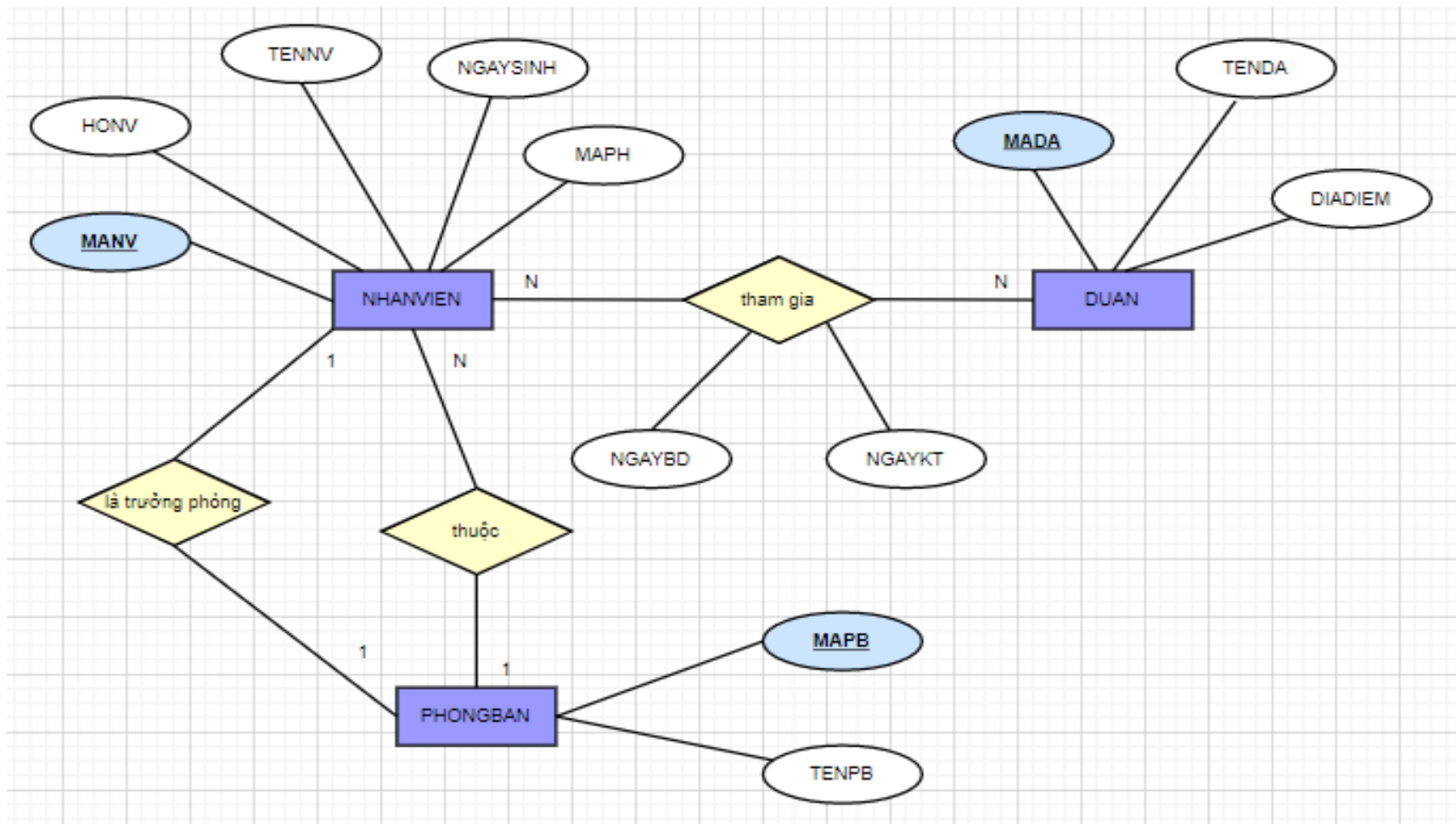
- Lấy khóa chính của quan hệ ở phía "một" để làm khóa ngoại của quan hệ phía "nhiều".

- Đối với liên kết 1-1:

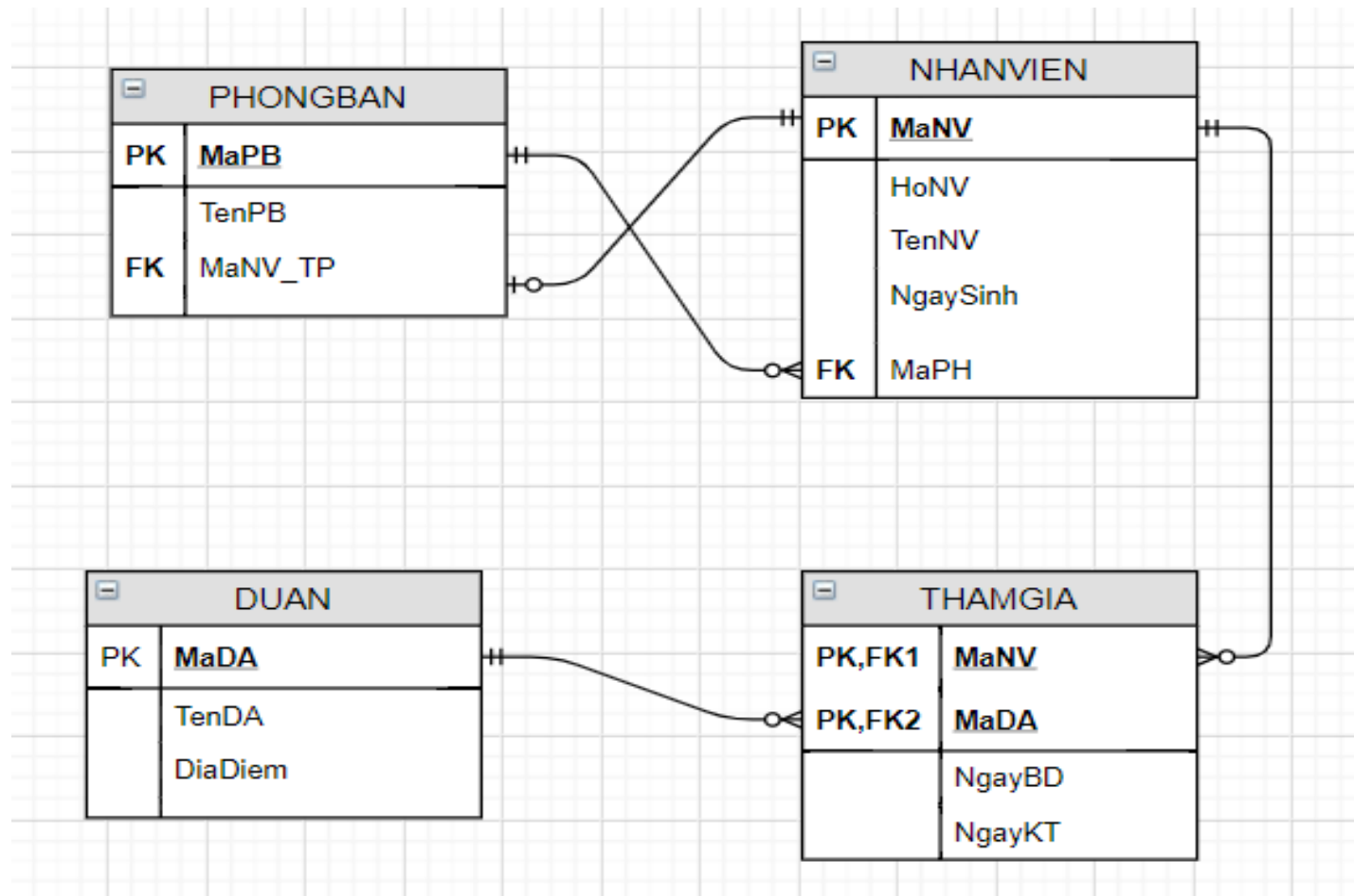
- Lấy khóa chính của quan hệ không có sự tham gia toàn bộ vào liên kết làm khóa ngoại của quan hệ kia.

- Đối với liên kết N-N:
 - **Phát sinh thêm một quan hệ trung gian** và chuyển một liên kết nhiều-nhiều thành hai liên kết một-nhiều.
 - Lấy khóa chính của 2 quan hệ ban đầu làm khóa ngoại của quan hệ phát sinh.
 - Khóa chính của quan hệ phát sinh là tập hợp khóa chính của các quan hệ ban đầu.
 - Chuyển thuộc tính của của mỗi quan hệ N-N (nếu có) vào quan hệ mới phát sinh.

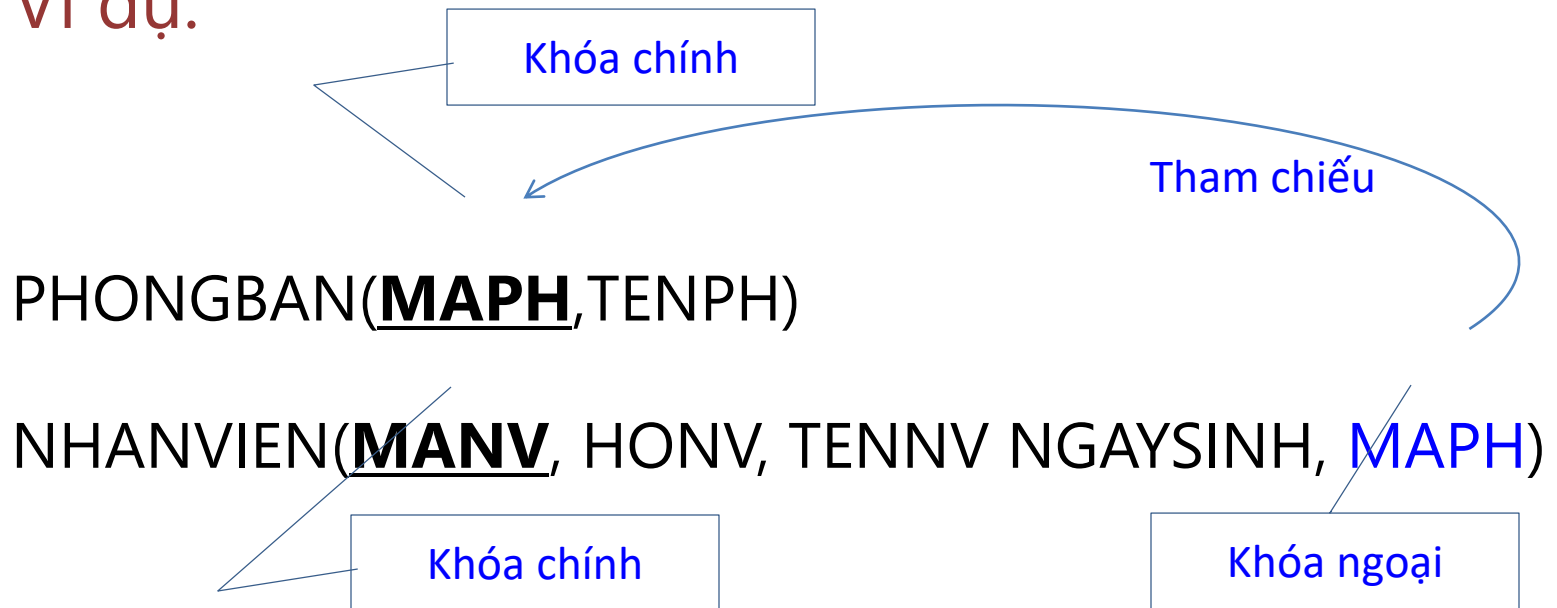
- Ví dụ chuyển hóa từ ERD → lược đồ quan hệ.



- Sau chuyển hóa ta có lược đồ quan hệ



■ Ví dụ:



- Ta có thể biểu diễn các quan hệ như sau:

PHONGBAN(**MAPB**, TENPB, MANV TP)

NHANVIEN (**MANV**, HONV, TENNV, NGAYSINH, MAPH)

DUAN (**MADA**, TENDA, DIADIEM)

THAMGIA_DA (**MANV**, **MADA**, NGAYBD, NGAYKT)

- Ghi chú:
 - Các field in đậm, gạch chân là khóa chính
 - Các field chữ nghiêng, gạch chân là khóa ngoại



CƠ SỞ DỮ LIỆU

BÀI 3: MÔ HÌNH ERD VÀ CHUẨN HOÁ CƠ SỞ DỮ LIỆU

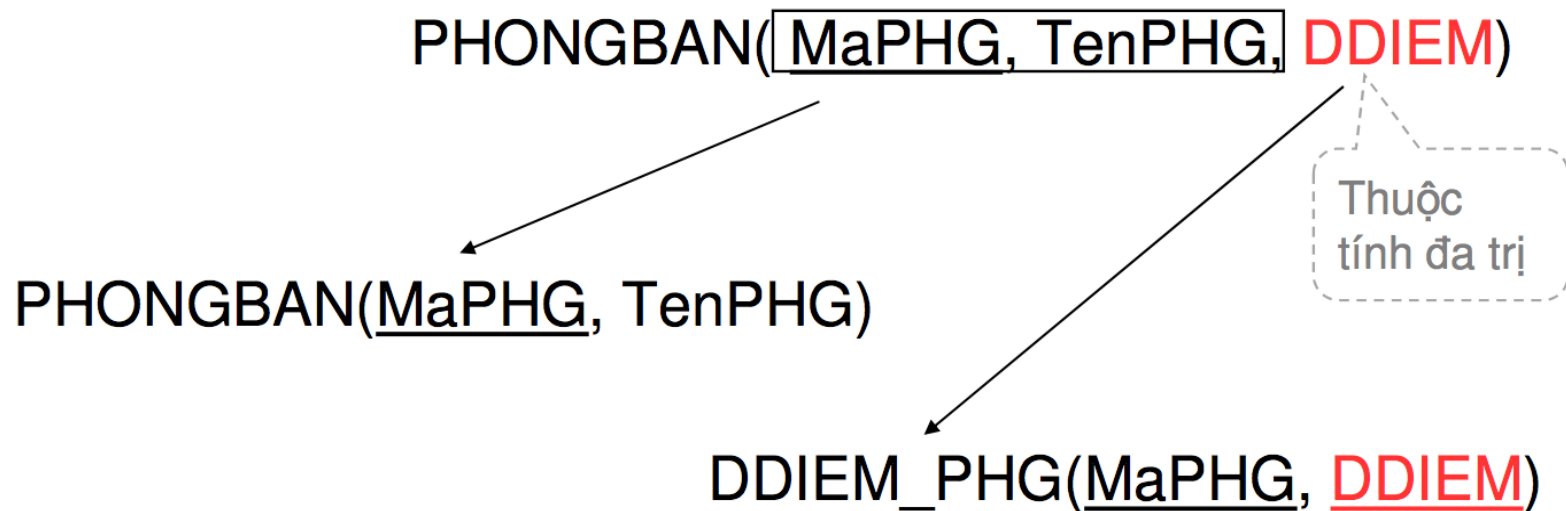
PHẦN 2 - CHUẨN HOÁ DỮ LIỆU

- Chuẩn hóa là 1 cách tiếp cận từ dưới lên (bottom-up approach) để thiết kế CSDL, bắt đầu từ các mối liên hệ giữa các thuộc tính
- Mục đích của chuẩn hóa là loại bỏ các bất thường của 1 quan hệ để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn
- Giảm thiểu sự dư thừa dữ liệu ở mức thấp nhất và cho phép người dùng thêm, sửa, xóa mà không gây ra mâu thuẫn dữ liệu

- Dạng chuẩn 1 (1NF – first normal form)
- Dạng chuẩn 2 (2NF – second normal form)
- Dạng chuẩn 3 (3NF – third normal form)
- Dạng chuẩn BCNF (Boyce-Codd normal form)
- Dạng chuẩn 4NF

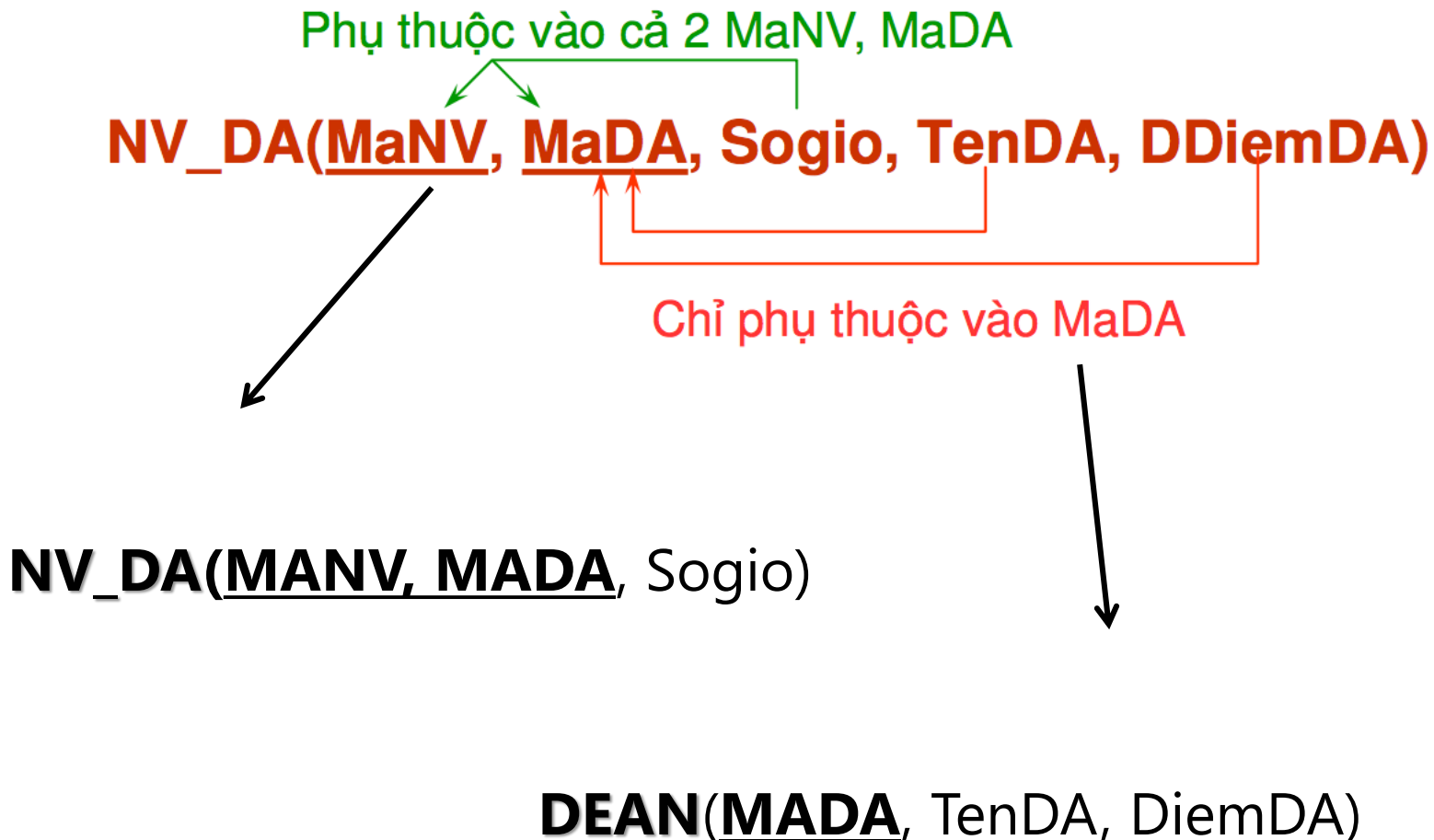
- Một bảng được gọi là ở dạng 1NF nếu:
 - Tất cả các thuộc tính đều là thuộc tính đơn và đơn trị.
 - Có khóa chính.
 - Không có quan hệ lặp.
- Chuẩn hóa:
 - Tách các thuộc tính phức hợp hoặc đa trị thành bảng mới.

■ Ví dụ:



- Một bảng được gọi là ở dạng 2NF
 - Thỏa mãn 1NF
 - Phụ thuộc hàm đầy đủ vào khóa chính (Mỗi thuộc tính không khóa phụ thuộc hàm vào toàn bộ khóa chính.)
- Chuẩn hóa: Nếu quan hệ đã đạt dạng chuẩn 1 và
 - Khóa chính đơn thì nó đạt dạng chuẩn 2.
 - Nếu quan hệ chỉ chứa khóa chính thì nó cũng đạt dạng chuẩn 2.
 - Với các bảng có khóa chính kết hợp, kiểm tra nếu có chứa phụ thuộc hàm bộ phận thì tách phụ thuộc hàm bộ phận thành bảng mới.

- Ví dụ phụ thuộc hàm bộ phận:



- 3NF dựa trên khái niệm phụ thuộc bắc cầu.
- Một lược đồ quan hệ là ở 3NF nếu nó:
 - Thỏa mãn 2NF
 - Không có thuộc tính không khóa nào của R là phụ thuộc bắc cầu vào khóa chính.
- Chuẩn hóa:
 - Tất cả các thuộc tính không khóa chỉ phụ thuộc vào thuộc tính khóa → đạt dạng chuẩn 3.
 - Các thuộc tính phụ thuộc vào thuộc tính không phải là khóa thì tách nhóm các thuộc tính đó thành bảng mới.

■ Ví dụ phụ thuộc bắc cầu:

Phụ thuộc vào MaNV

NV_DV(MaNV, TenNV, NS, DCHI, MaDV, TenDV, TruongPHG)

Phụ thuộc vào MaDV

NHANVIEN(MaNV, TenNV, NS, DCHI, MaDV)

DONVI(MaDV, TenDV, TruongPHG)

TÓM TẮT 3 DẠNG CHUẨN 1-3

NF	Nhận biết	Cách chuẩn hoá
1	Quan hệ không có thuộc tính phức hợp, đa trị và quan hệ lặp.	Chuyển tất cả quan hệ lặp hoặc đa trị, phức hợp thành một quan hệ mới
2	Phụ thuộc 1 phần vào thuộc tính khóa	Tách thuộc tính phụ thuộc một phần thành lược đồ mới, đảm bảo quan hệ với lược đồ liên quan
3	Phụ thuộc ẩn, tồn tại phụ thuộc hàm giữa các thuộc tính không phải là khóa	Tách các thuộc tính đó thành lược đồ mới

THIẾT KẾ CƠ SỞ DỮ LIỆU Ở MỨC VẬT LÝ

- Phần thiết kế vật lý CSDL sẽ phụ thuộc vào một HQTCSDL mà bạn sẽ sử dụng để cài đặt CSDL.
- Chúng ta cần chuyển từ các bản ghi logic với các thuộc tính thành các quan hệ được biểu diễn dưới dạng bảng với các trường hay các cột được cài đặt trong một HQTCSDL cụ thể.
- Quá trình chuyển hóa này sẽ được thuận tiện hơn nếu các giai đoạn trước đã được thực hiện kỹ lưỡng.

- Dữ liệu được biểu diễn dưới dạng các **bảng (table)**.
- Mỗi **quan hệ** trong lược đồ CSDL là một **bảng**
- **Bảng** bao gồm các **cột (field)**, các **hàng/bộ (record)**
 - Mỗi cột biểu diễn một thuộc tính và có **kiểu dữ liệu (Data type)** nhất định.
 - Mỗi hàng/bộ thể hiện một bộ thông tin
 - Mỗi bảng có một **Khóa chính (Primary key)** – xác định tính duy nhất của bộ dữ liệu trong tập dữ liệu - khóa gồm **một** hoặc **vài** cột của bảng.

- Tập hợp các cột
 - Cố định số lượng cột
 - Mỗi cột được đặt tên, có kiểu dữ liệu
- Mỗi một giá trị trong một hàng (tuple – bộ)
 - Hoặc chứa giá trị nguyên tố
 - Hoặc chứa giá trị rỗng (null)
- Không có 2 hàng trùng nhau hoàn toàn trong một bảng
- Thứ tự của các hàng trong một bảng là không quan trọng
- Tập hợp các dòng thay đổi theo thời gian

- Với mỗi bảng cần chỉ rõ:
 - Khoá chính (có thể gồm một hay nhiều thuộc tính).
 - Mô tả của tất cả các cột (trường).
- Với mỗi cột (trường) cần phải có:
 - Một tên duy nhất (trong bảng lưu giữ nó).
 - Một mô tả ngắn gọn.
 - Một kiểu dữ liệu (ví dụ: integer, char, date, logical,... phụ thuộc vào HQTCSDL cụ thể cài đặt CSDL)
 - Một kích thước (mặc định hay chỉ rõ tùy từng kiểu dữ liệu)
- Chú ý: tên của bảng, cột không nên quá dài, cần đủ nghĩa và thường không có dấu.

- Các cột (trường) tùy chọn và các cột (trường) bắt buộc:
 - Nếu cột là bắt buộc thì người sử dụng cần phải cung cấp một giá trị cho cột này trong mỗi dòng (bản ghi) thêm vào bảng khi cập nhật CSDL. Cột sẽ cần phải được chỉ rõ là NOT NULL
 - Nếu cột là tùy chọn thì nó có thể nhận giá trị null.
- Giá trị mặc định có thể được sử dụng để giảm bớt việc gõ máy cho người sử dụng. Nó giúp cho việc nhập dữ liệu nhanh hơn và giảm bớt lỗi.

- Tên các cột của một bảng là duy nhất
- Mô tả ý nghĩa cho các giá trị tại cột đó
- Tất cả các dữ liệu trong cùng một cột đều có cùng kiểu dữ liệu

Thuộc tính

TENNV	HONV	NS	DIACHI	GT	LUONG	PHG
Tung	Nguyen	12/08/1955	638 NVC Q5	Nam	40000	5
Hang	Bui	07/19/1968	332 NTH Q1	Nu	25000	4
Nhu	Le	06/20/1951	291 HVH QPN	Nu	43000	4
Hung	Nguyen	09/15/1962	Ba Ria VT	Nam	38000	5

- Mỗi cột trong bảng được quy định bởi một kiểu dữ liệu
- Kiểu dữ liệu cho phép xác định:
 - Loại dữ liệu của cột như dạng số, dạng kí tự, ngày tháng...
 - Giới hạn miền giá trị cho cột

- **Ràng buộc (Constraint):** là những quy tắc cần tuân theo khi nhập liệu vào CSDL để hạn chế miền giá trị các thuộc tính.
- **Ràng buộc toàn vẹn (Integrity Constraint):** là ràng buộc nhằm nhấn mạnh sự chính xác của dữ liệu nhập vào.
- Bao gồm 3 kiểu ràng buộc toàn vẹn:
 - Ràng buộc **NOT NULL**
 - Ràng buộc **CHECK**
 - Ràng buộc **Trigger** (Trigger là chương trình/macro tự động thực hiện khi có một sự kiện (bất thường) xảy ra trong CSDL)

- Ví dụ: cột HOSV của bảng SINHVIEN có ràng buộc NOT NULL, tức là đòi hỏi phải nhập dữ liệu

sinhvien - Table

Table Name: Schema: **mytest**

Charset/Collation: Engine:

Comments:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
MASV	INT	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
HOSV	VARCHAR(50)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
TENSV	VARCHAR(20)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
NGAYSINH	DATE	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
MANGANH	VARCHAR(10)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

Column Name: Data Type:

Charset/Collation:

Comments:

Storage: ☐ Virtual ☒ Stored

☐ Primary Key ☒ Not Null ☐ Unique

☐ Binary ☐ Unsigned ☐ Zero Fill

☐ Auto Increment ☐ Generated

Columns Indexes Foreign Keys Triggers Partitioning Options

- ❑ Cần phải đánh giá việc thực thi CSDL, đó là tính hợp lý của thời gian đáp ứng các câu hỏi truy vấn.
- ❑ Chạy thử CSDL với một tập đủ lớn các bản ghi trong môi trường thực tế.
- ❑ Mô phỏng một số người dùng cùng thực hiện một truy vấn trong CSDL đa người sử dụng.
- ❑ Đôi khi chúng ta cần cân nhắc đưa một số dư thừa dữ liệu vào cơ sở dữ liệu để tăng tốc độ truy vấn. Khi đó có thể chúng ta đã phá vỡ dạng chuẩn của các quan hệ đã được chuẩn hoá.

- Thực hành thiết kế CSDL bài toán Quản lý bán hàng tiếp theo slide3 phần 1
- Yêu cầu:
 1. Chuẩn hóa CSDL đã thiết kế để đạt chuẩn 3NF
 2. Thiết kế vật lý của CSDL dùng với HQT MySQL
 3. Cài đặt CSDL dùng HQT MySQL.

- ❑ Mô hình hoá dữ liệu là sự biểu diễn thế giới thực thông qua các kí hiệu
- ❑ Cố gắng rút gọn tổ chức dữ liệu thành sự mô tả của các thực thể và các mối liên hệ giữa chúng
- ❑ Một thực thể (entity) là một đối tượng của thế giới thực
- ❑ Sơ đồ quan hệ ERD biểu diễn các tập thực thể và mối quan hệ giữa chúng



- ❑ Chuẩn hoá dữ liệu là quá trình loại bỏ các bất thường của 1 quan hệ để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn
- ❑ Giảm thiểu sự dư thừa dữ liệu ở mức thấp nhất và cho phép người dùng thêm, sửa, xóa mà không gây ra mâu thuẫn dữ liệu
- ❑ Các dạng chuẩn hoá:
 - ❖ Dạng chuẩn 1 (1NF – first normal form)
 - ❖ Dạng chuẩn 2 (2NF – second normal form)
 - ❖ Dạng chuẩn 3 (3NF – third normal form)



Cảm ơn