



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 ERD và các thành phần của nó
 - ❖ Từ mô hình ERD hướng đến xây dựng CSDL mức vật lí
 - ❖ Giới thiệu về chuẩn hoá dữ liệu
 - ❖ Quy trình chuẩn hoá dữ liệu

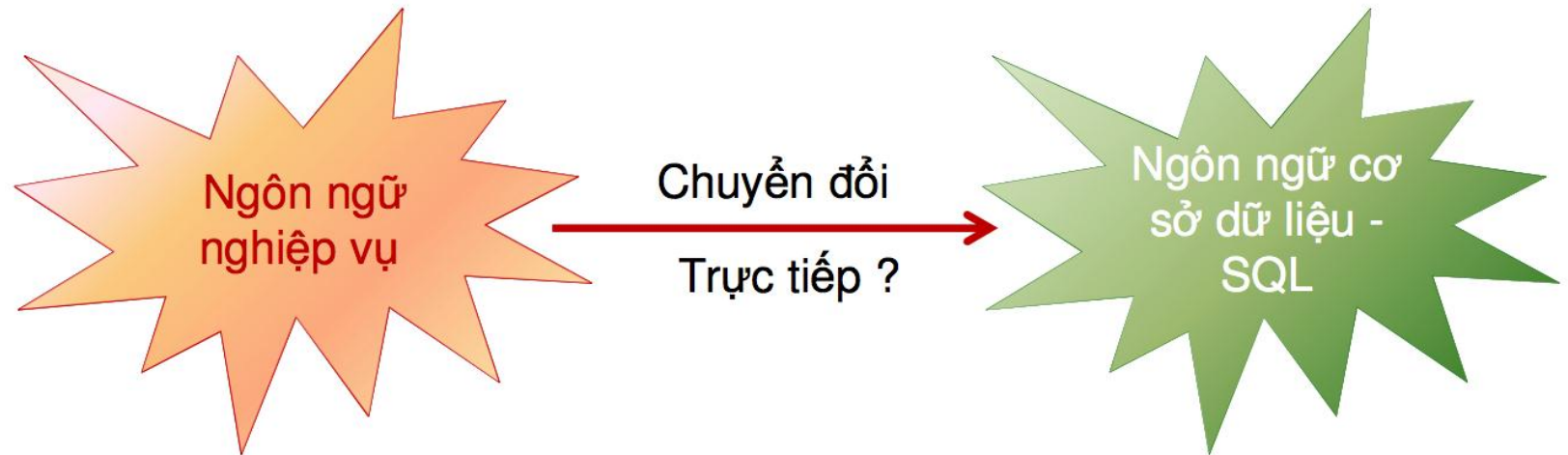




PHẦN 1

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



Người phân tích nghiệp vụ

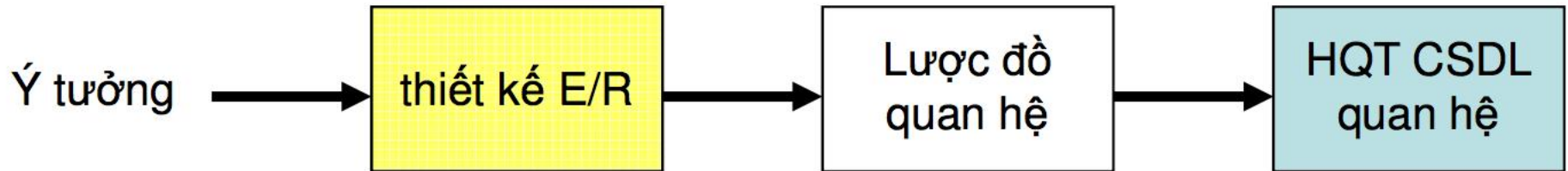


Độc lập hệ quản trị; Độc lập mô hình DL



Lập trình viên

QUÁ TRÌNH THIẾT KẾ CSDL



❑ Thu thập và phân tích yêu cầu

- ❖ Các yêu cầu về CSDL
- ❖ Các yêu cầu về chức năng (thao tác trên CSDL)

❑ Thiết kế quan niệm và phân tích chức năng

- ❖ Tạo một sơ đồ quan niệm, ví dụ: ERD
- ❖ Đặc tả giao tác cấp cao tương ứng với các thao tác trên CSDL

❑ Thiết kế logic

- ❖ Ánh xạ lược đồ quan niệm thành lược đồ logic: vd: mô hình quan hệ

- ❑ Tiếp cận từ trên xuống – Top Down
 - ❖ Xác định các thực thể dữ liệu
 - ❖ Xác định các thuộc tính thực thể
 - ❖ Xác định các mối liên hệ giữa các thực thể
 - > **Tính trừu tượng cao - khó**

- ❑ Tiếp cận từ dưới lên – Bottom Up
 - ❖ Tập hợp các yếu tố dữ liệu từ thực tế
 - ❖ Gom thành các thực thể
 - ❖ Xác định mối quan hệ giữa các thực thể
 - > **Xuất phát từ thực tế - dễ**

- ❑ Một thực thể (entity) là một đối tượng của thế giới thực
- ❑ Tập hợp các thực thể giống nhau tạo thành 1 tập thực thể (entity set)

Mô hình ER	Mô hình OOP
Thực thể (Entity)	Đối tượng (Object)
Tập thực thể (Entity set)	Lớp đối tượng (Class of Object)
Thành phần: <ul style="list-style-type: none">- Thuộc tính/Dữ liệu (attributes)	Thành phần: <ul style="list-style-type: none">- Thuộc tính (Attribute)- Thao tác trên dữ liệu (method)

- ❑ Là những đặc tính riêng biệt của tập thực thể
- ❑ Là tính chất của thực thể cần được quản lý
- ❑ Chỉ quan tâm tới những tính chất có liên quan tới ứng dụng
- ❑ Ví dụ tập thực thể **NHAN_VIEN** có các thuộc tính
 - ❖ Họ tên
 - ❖ Ngày sinh
 - ❖ Trình độ
 - ❖ Địa chỉ
 - ❖ ...

☐ 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)

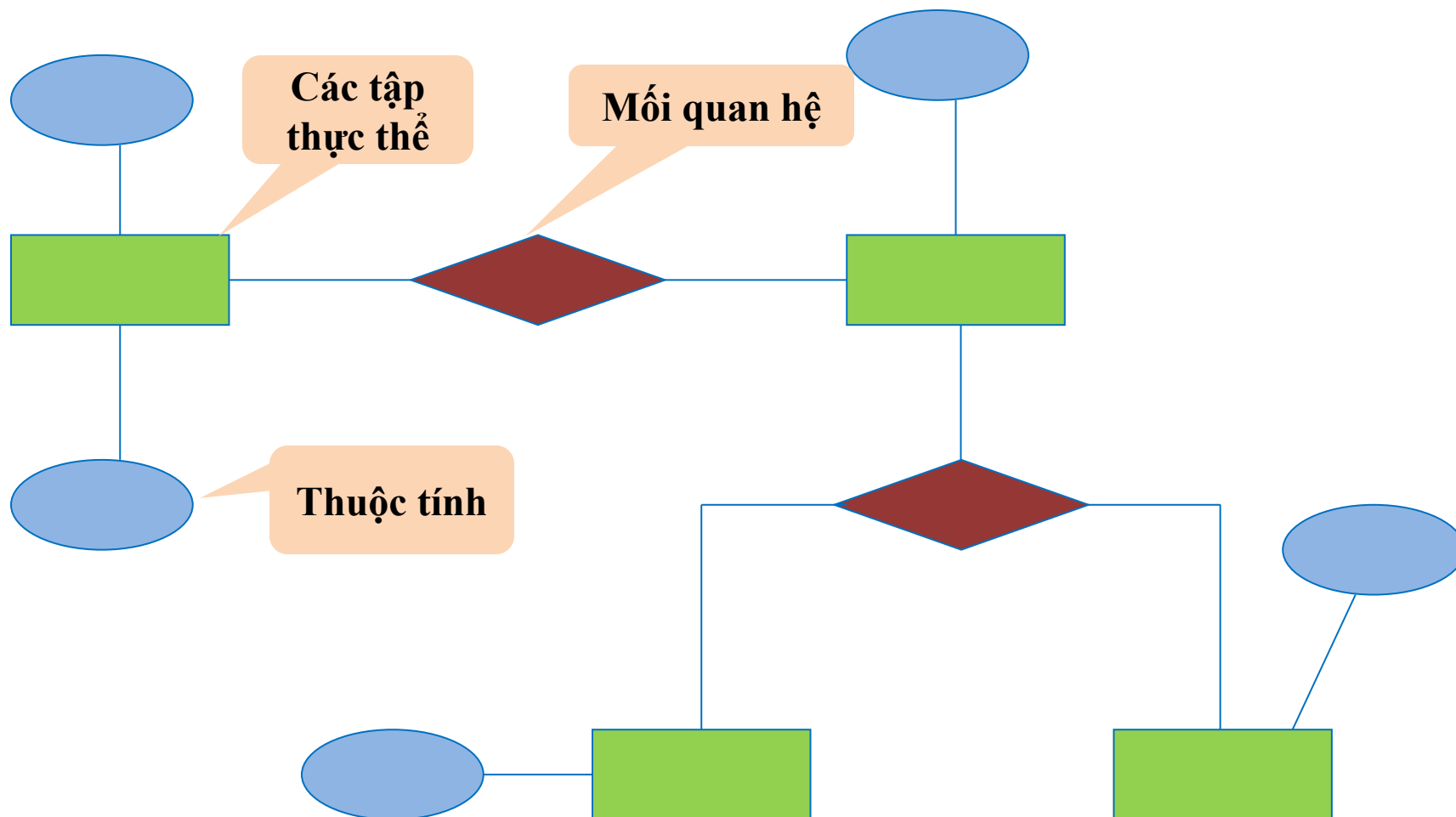
☐ Thuộc tính đa trị

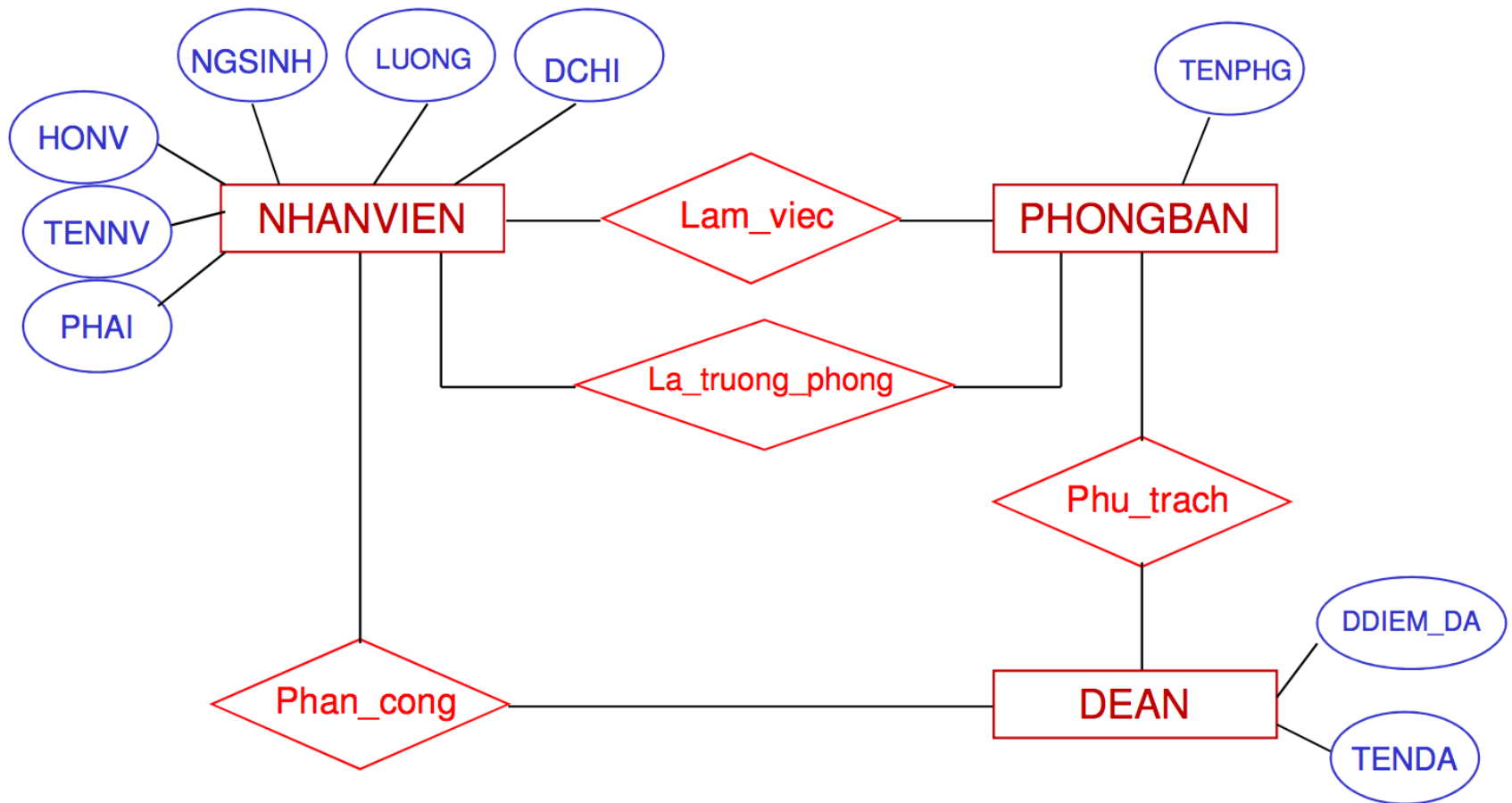
- ❖ Có thể nhiều giá trị khác nhau ở cùng 1 thực thể (vd: số đt)

☐ Thuộc tính suy diễn

- ❖ Giá trị của nó được suy ra từ thuộc tính khác (vd: Năm sinh -> tuổi)

- ❑ **Quan hệ:** Là sự liên kết giữa 2 hay nhiều tập thực thể
- ❑ Ví dụ giữa tập thực thể **NHANVIEN** và **PHONGBAN** có các liên kết
 - ❖ Một nhân viên thuộc một phòng ban nào đó
 - ❖ Một phòng ban có một nhân viên làm trưởng phòng





– Một-Nhiều



– Một-Một



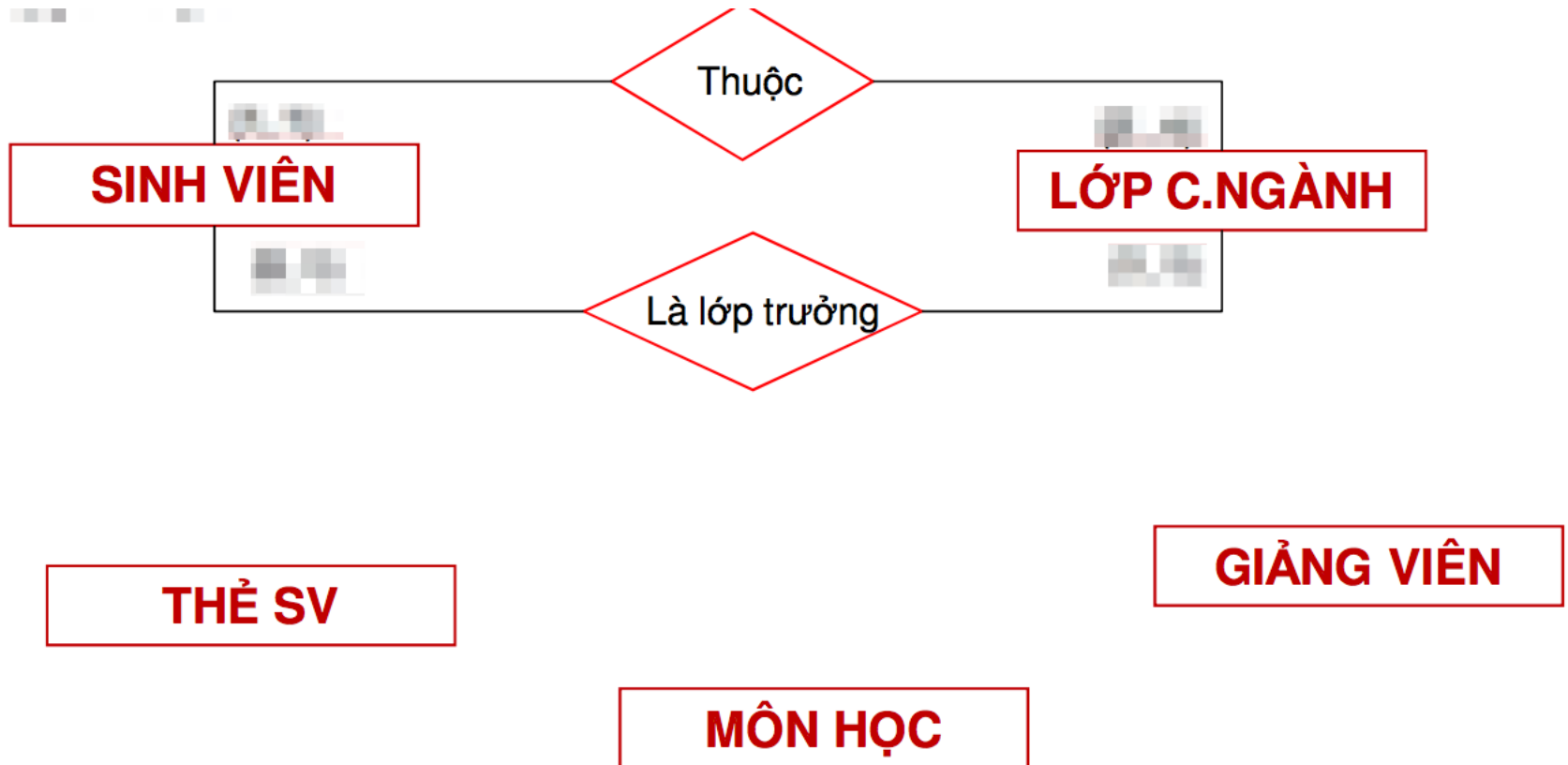
– Nhiều-Một



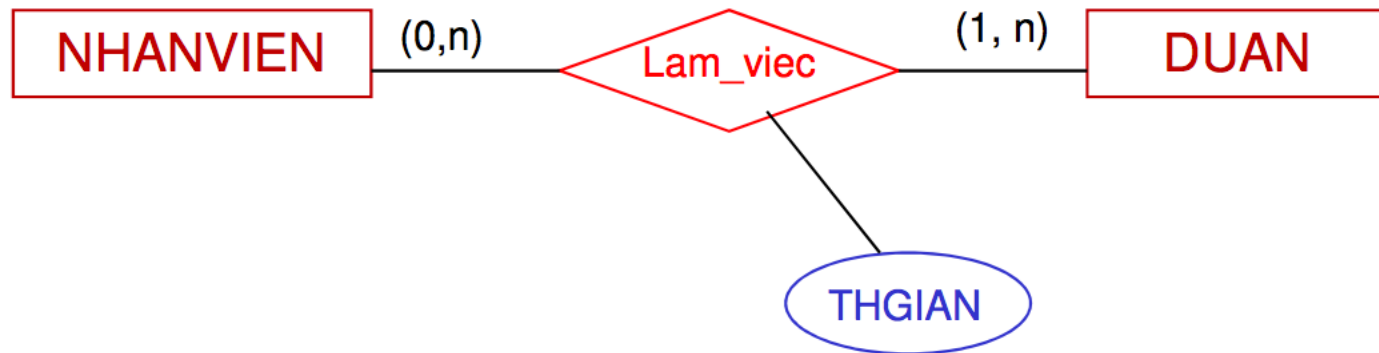
– Nhiều-Nhiều



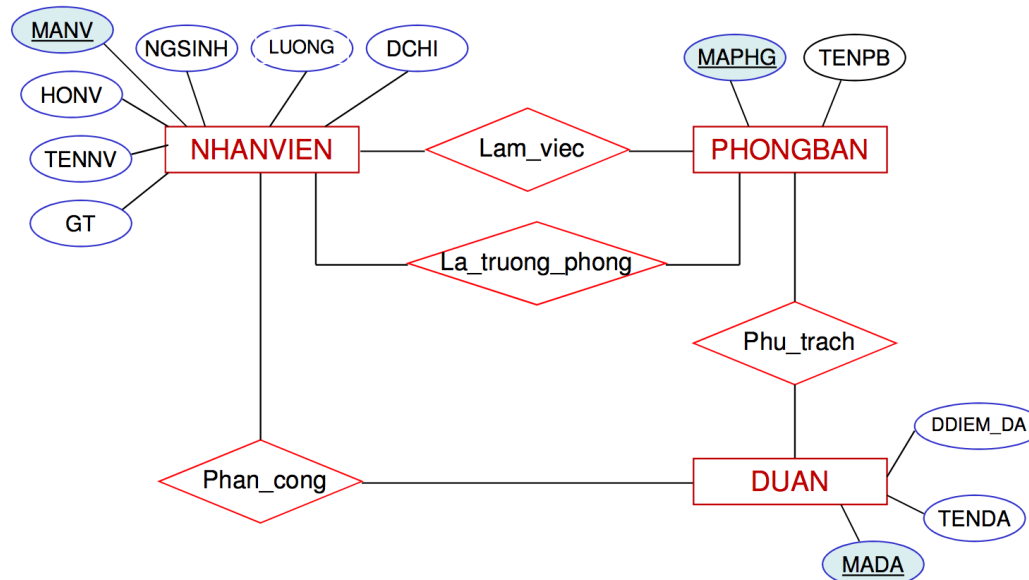
- ❑ Xác định mối quan hệ và bản số cho các tập thực thể sau:



- ❑ Thuộc tính trên mối quan hệ mô tả tính chất cho mỗi quan hệ đó

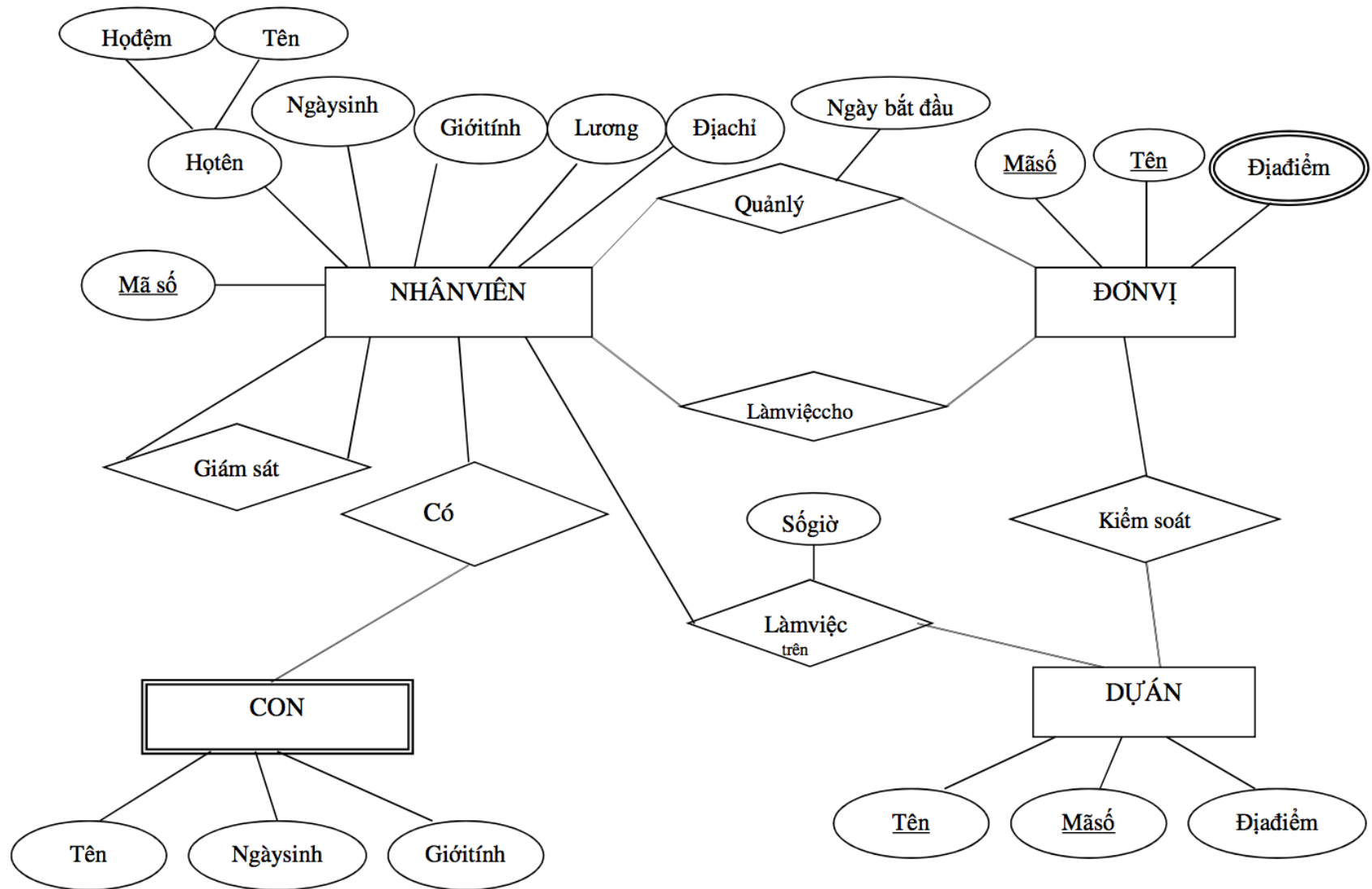


- ❑ Còn được gọi là thuộc tính định danh của tập thực thể
- ❑ Dùng để phân biệt giữa các thực thể khác nhau trong tập thực thể
- ❑ Mỗi tập thực thể phải có 1 khóa
- ❑ Một khóa có thể có 1 hay nhiều thuộc tính



BÀI TẬP – XÂY DỰNG SƠ ĐỒ ER

- ☐ CSDL dự án công ty theo dõi các thông tin liên quan đến nhân viên, phòng ban và các dự án
- ☐ Cty có nhiều phòng ban, mỗi phòng ban có tên duy nhất, mã phòng ban duy nhất, một trưởng phòng và ngày nhận chức. Mỗi phòng ban có thể ở nhiều địa điểm khác nhau.
- ☐ Dự án có tên duy nhất, mã duy nhất, do 1 một phòng ban chủ trì và được triển khai ở 1 địa điểm.
- ☐ Nhân viên có mã số, tên, địa chỉ, ngày sinh, giới tính và lương. Mỗi nhân viên làm việc ở 1 phòng ban, tham gia vào các dự án với số giờ làm việc khác nhau. Mỗi nhân viên đều có một người quản lý trực tiếp.
- ☐ Một nhân viên có thể có những người con được hưởng bảo hiểm theo nhân viên. Mỗi người con của nhân viên có tên, giới tính, ngày sinh





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 miền giá trị của một thuộc tính chỉ chứa giá trị nguyên tố đơn (không phân chia được) và giá trị của mỗi thuộc tính cũng là một giá trị đơn lấy từ miền giá trị của nó

❑ Ví dụ:

PHONGBAN(MaPHG, TenPHG, **DDIEM**)

PHONGBAN(MaPHG, TenPHG)

DDIEM_PHG(MaPHG, DDIEM)

Thuộc
tính đa trị

❑ Xét lược đồ DDIEM_PHG(MaPHG, DDIEM)

- Vẫn bị lặp lại
- Ẩn chứa các phụ thuộc hàm bộ phận

MAPHG	DIADIEM
1	TP HCM
4	HA NOI
5	VUNGTAU
5	NHATRANG
5	TP HCM

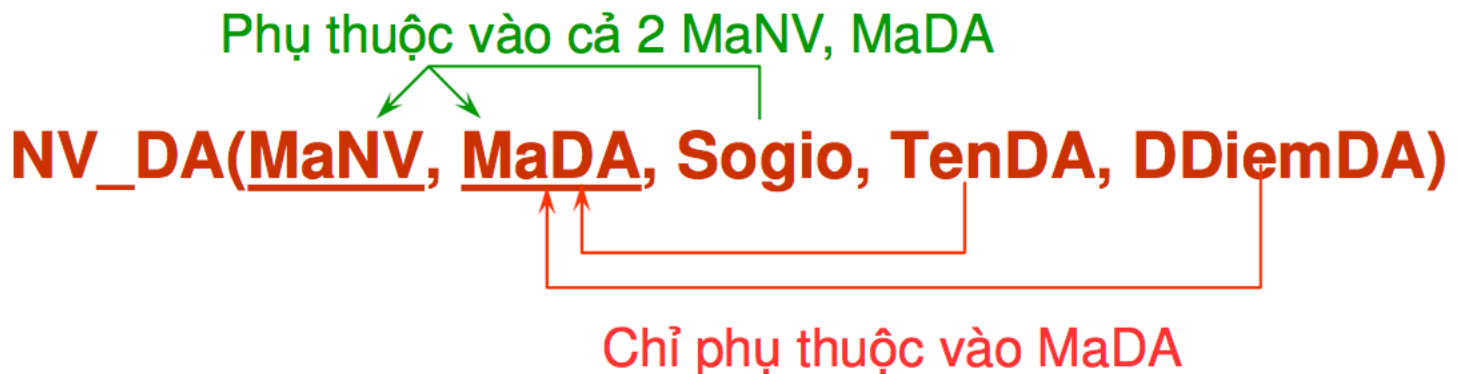
- ❑ Một phụ thuộc hàm $X \rightarrow Y$ là một **phụ thuộc hàm đầy đủ** nếu loại bỏ bất kỳ thuộc tính A nào ra khỏi X thì phụ thuộc hàm không còn đúng nữa.

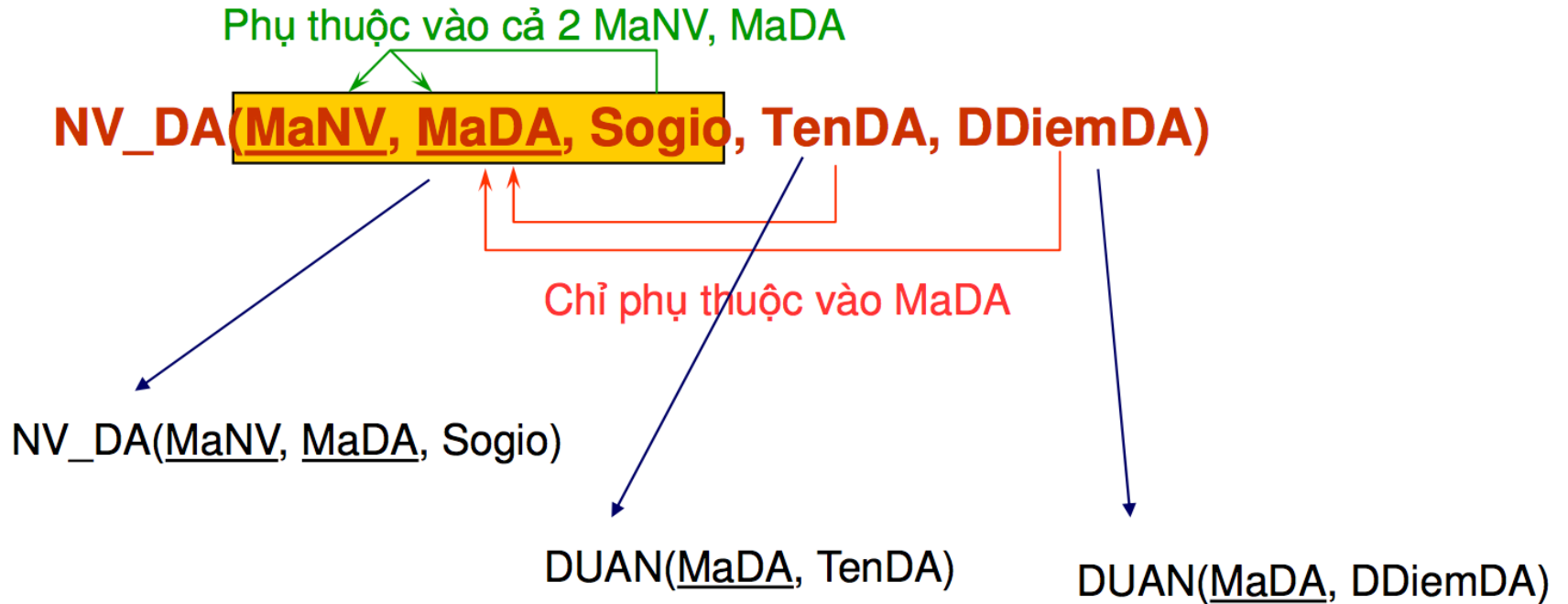
$\forall A, A \in X, (X - \{A\}) \rightarrow Y$: là sai.

- ❑ Một phụ thuộc hàm $X \rightarrow Y$ là phụ thuộc bộ phận nếu có thể bỏ một thuộc tính $A \in X$, ra khỏi X phụ thuộc hàm vẫn đúng, điều đó có nghĩa là với

$\exists A \in X, (X - \{A\}) \rightarrow Y$

- ❑ Một bảng được gọi là ở dạng 2NF
 - ❖ Thỏa mãn 1NF
 - ❖ Phụ thuộc hàm đầy đủ vào khóa chính
- ❑ Với các quan hệ có thuộc tính khóa đơn thì ko phải xét
- ❑ Chỉ kiểm tra các lược đồ có chứa phụ thuộc hàm bộ phận
- ❑ Ví dụ:





- ❑ 3NF dựa trên khái niệm phụ thuộc bắc cầu.
- ❑ Một lược đồ quan hệ R là ở 3NF nếu nó:
 - ❖ Thỏa mãn 2NF
 - ❖ Không có thuộc tính không khoá nào của R là phụ thuộc bắc cầu vào khoá chính.

Phụ thuộc vào MaNV



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



Phụ thuộc vào MaDV

- ❑ Tất cả các thuộc tính phải phụ thuộc vào thuộc tính khóa
 - ❖ Một vài thuộc tính phụ thuộc vào thuộc tính ko phải là khóa
 - ❖ Chuẩn hóaTách nhóm các thuộc tính đó thành quan hệ mới

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 ko có thuộc tính đa trị và quan hệ lặp	Chuyển tất cả quan hệ lặp hoặc đa trị thành 1 quan hệ mới
2	Phụ thuộc 1 phần vào thuộc tính khóa	Tách thuộc tính phụ thuộc 1 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 ko phải là khóa	Tách các thuộc tính đó thành lược đồ mới

- ❑ 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 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)
 - ❖ Dạng chuẩn BCNF (Boyce-Codd normal form)
 - ❖ Dạng chuẩn 4NF
 - ❖ ...



Cảm ơn