

TRƯỜNG ĐÀO TẠO LẬP TRÌNH VIÊN VÀ QUẢN TRỊ MẠNG QUỐC TẾ BACHKHOA-APTECH

Bài 02

Mô hình ER và chuẩn hóa

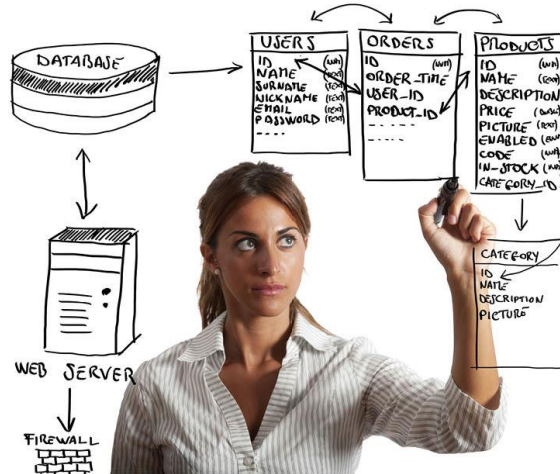
Tóm tắt

- Thiết kế CSDL quan hệ
- Thành phần trong mô hình quan hệ thực thể E-R
- Xác định mối quan hệ giữa các thực thể.
- Sơ đồ E-R
- Các dạng chuẩn hóa



Thiết kế CSDL quan hệ

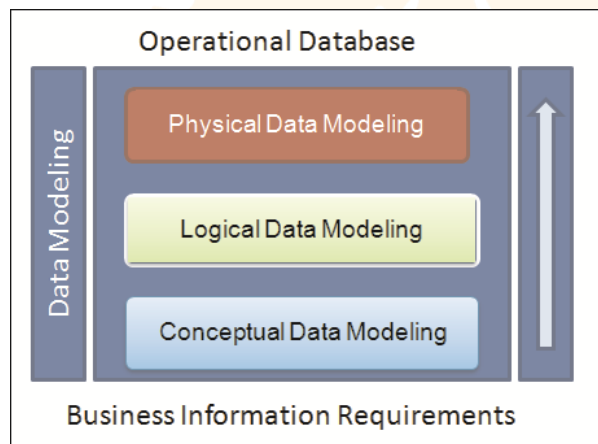
- Khi phát triển sản phẩm phần mềm, một việc rất quan trọng khi bắt đầu đó chính là thiết kế CSDL.
- Việc phân tích thiết kế CSDL giúp định hình chức năng, hoạt động phần mềm.
- Thiết kế csdl tốt giúp cho việc phát triển sản phẩm dễ dàng, an toàn, bảo mật, dễ bảo trì, nâng cấp, mở rộng.



Thiết kế CSDL quan hệ

Thiết kế CSDL được phân thành các mức khác nhau:

- Thiết kế các thành phần dữ liệu **mức khái niệm**
- Thiết kế các thành phần dữ liệu **mức logic**
- Thiết kế các thành phần dữ liệu **mức vật lý**



Thiết kế mức
khái niệm

Thiết kế
Logic

Thiết kế
mức vật lý

Thiết kế CSDL quan hệ

Thiết kế **mức khái niệm**

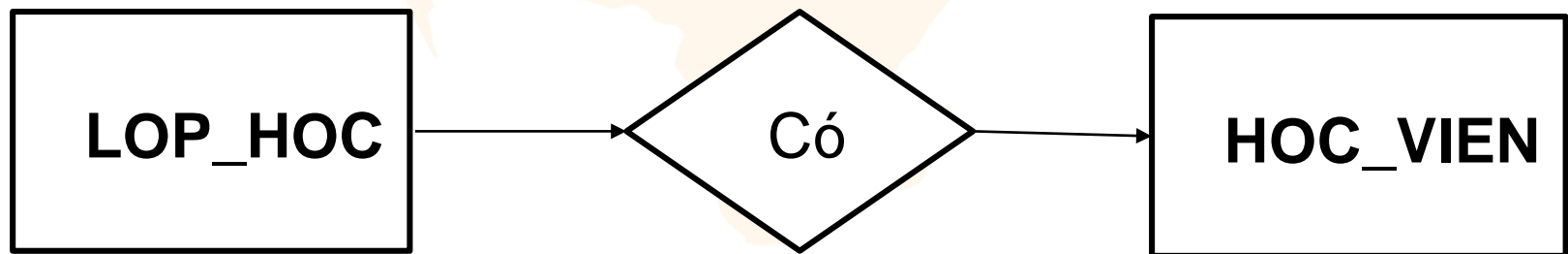
- Là sự trừu tượng hóa của thế giới thực (các nghiệp vụ có trong sản phẩm phần mềm).
- Trong DBMS, **Sơ đồ thực thể - liên kết (ERD)** dùng để mô tả lược đồ CSDL mức khái niệm.
- **Sơ đồ thực thể - liên kết** sẽ được đề cập kĩ hơn trong phần sau.

VD: Kịch bản hệ thống quản lý lớp học => có 2 thực thể cần mô hình là **LOP_HOC** và **HOC_VIEN**.

Thiết kế CSDL quan hệ

Thiết kế **mức logic**


- Thiết kế CSDL mức logic là quá trình chuyển CSDL mức khái niệm sang **mô hình sơ đồ quan hệ** và **chuẩn hóa các quan hệ**.
- Các khái niệm **sơ đồ quan hệ và chuẩn hóa** sẽ được đề cập trong phần sau.




Thiết kế CSDL quan hệ

Thiết kế **mức vật lý**

- Mức thấp nhất của kiến trúc một CSDL là cơ sở dữ liệu vật lý. CSDL vật lý là sự cài đặt cụ thể của CSDL mức khái niệm.
- CSDL vật lý bao gồm các **Bảng (Table)** và **mối quan hệ (Relationship)** giữa các bảng này.

tblLopHoc	
	id_lop
	ma_lop
	ten_lop
	he_dao_tao
	khai_giang
	ghi_chu

tblSinhVien	
	id_sv
	id_lop
	ma_sv
	ten_sv
	gioi_tinh
	ngay_sinh
	sdt
	dia_chi

Mô hình E-R

Mô hình quan hệ thực thể E-R gồm 5 thành phần cơ bản:

Thực thể

- Là đối tượng riêng biệt, vd: **Lớp học, sinh viên, môn học...**

Mối quan hệ

- Là liên kết hoặc ràng buộc giữa 2 hay nhiều thực thể. VD: Lớp Học **có** Sinh Viên, Sinh Viên **học** Môn Học

Thuộc tính

- Là đặc điểm mà thực thể có. VD: Sinh Viên có: **mã, tên, địa chỉ....**

Tập thực thể

- Là tập các thực thể giống nhau. VD: các sinh viên hợp thành **tập thực thể Sinh Viên**.

Tập mối quan hệ

- Là tập các mối liên hệ giống nhau. VD: tập quan hệ **“học tại”** giữa các sinh viên với lớp học.

Mô hình E-R

Bảng sau mô tả **thực thể** và **thuộc tính** của nó:

Tên thực thể	Thuộc tính	Mô tả
LOP_HOC	Ma_Lop	Dùng để lưu trữ mã lớp học, xác định duy nhất trong thực thể, ví dụ: 1,2,3,4...
	Ten_lop	Dùng để lưu trữ thông tin tên lớp học, ví dụ: C1707G, C1808H,...
	He_dao_tao	Lưu trữ thông tin hệ đào tạo là gì, ví dụ: ACCPi10; ACCPi13;...
	Ngay_khai_giang	Lưu trữ thông tin ngày bắt đầu khai giảng của lớp
	Tinh_trang	Tình trạng lớp đang học hay tạm ngừng hay đã hoàn thành.

Mối quan hệ (relationships)

Là mối liên kết giữa các tập thực thể (còn gọi là bảng).

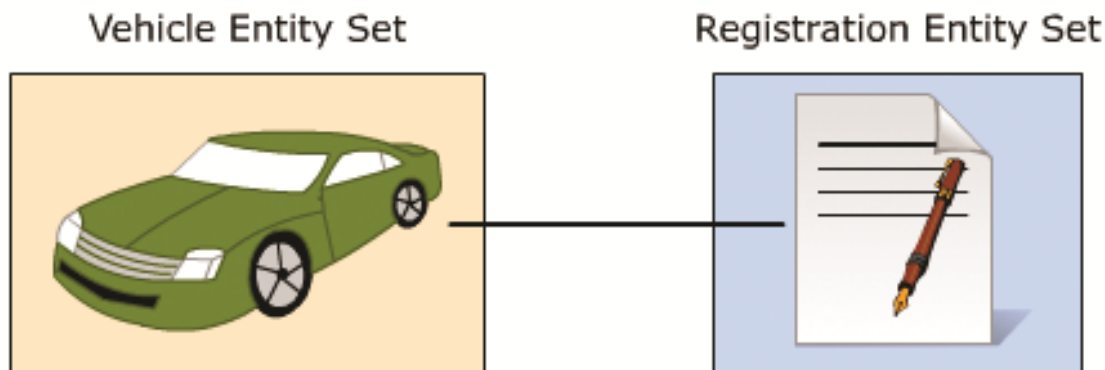
Phân loại:

- Quan hệ **1-1**
- Quan hệ **1-n (một – nhiều)**
- Quan hệ **n-n (nhiều – nhiều)**
- Quan hệ **đệ quy**

Mối quan hệ (relationships)

Quan hệ 1-1

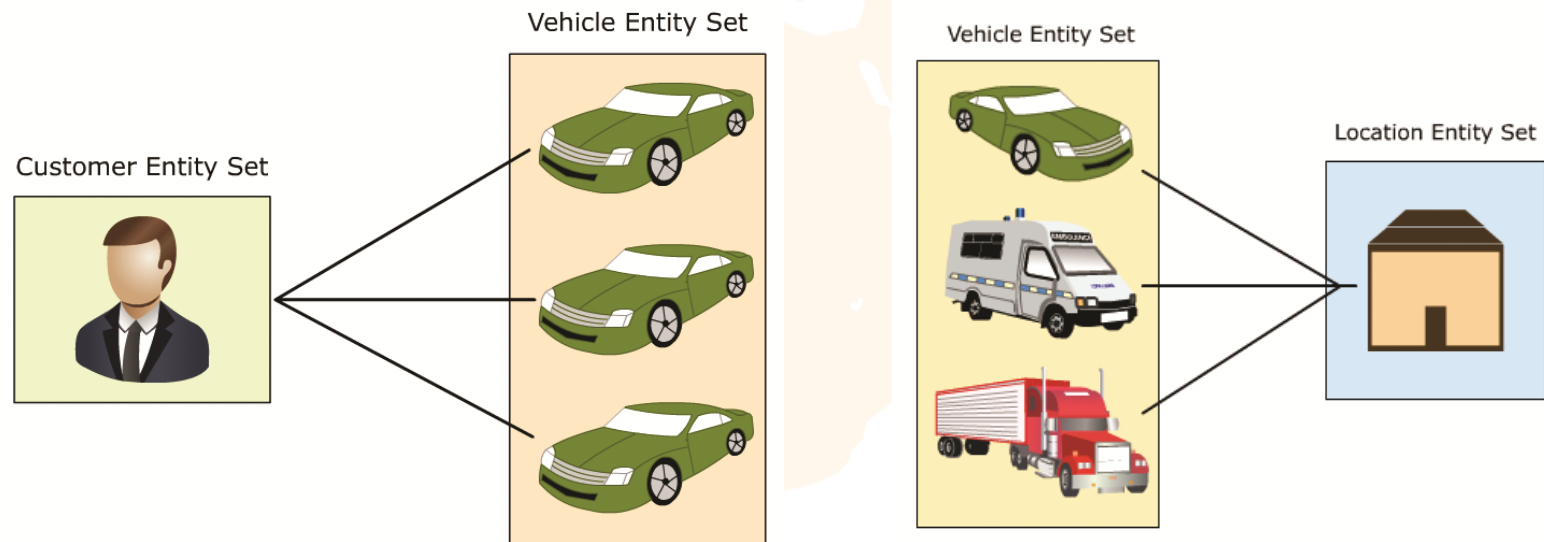
- Là quan hệ giữa hai tập thực thể trong đó mỗi thực thể của tập cha chỉ có thể liên kết **với nhiều nhất một** thực thể của tập con, và ngược lại.
- Ví dụ: quan hệ giữa thực thể **Oto** và **DangKy** là 1-1 (tức một ô tô có duy nhất một đăng ký xe).



Mối quan hệ (relationships)

Quan hệ 1-n (một – nhiều hay nhiều – một)

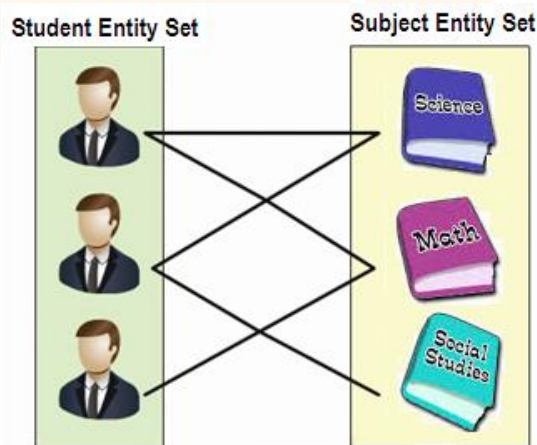
- Là quan hệ giữa hai tập thực thể trong đó mỗi thực thể của tập này có thể liên kết với nhiều thực thể của tập còn lại.
- Ví dụ: một người có thể sở hữu nhiều xe ô tô.



Mối quan hệ (relationships)

Quan hệ n-n (nhiều – nhiều)

- Là quan hệ giữa hai tập thực thể trong đó một thực thể của tập này có thể liên kết với 0, 1 hoặc nhiều thực thể của tập kia, và ngược lại.
- Thường quan hệ N-N có thêm phần dữ liệu giao nhau để thêm thông tin cụ thể cho **mỗi quan hệ**.



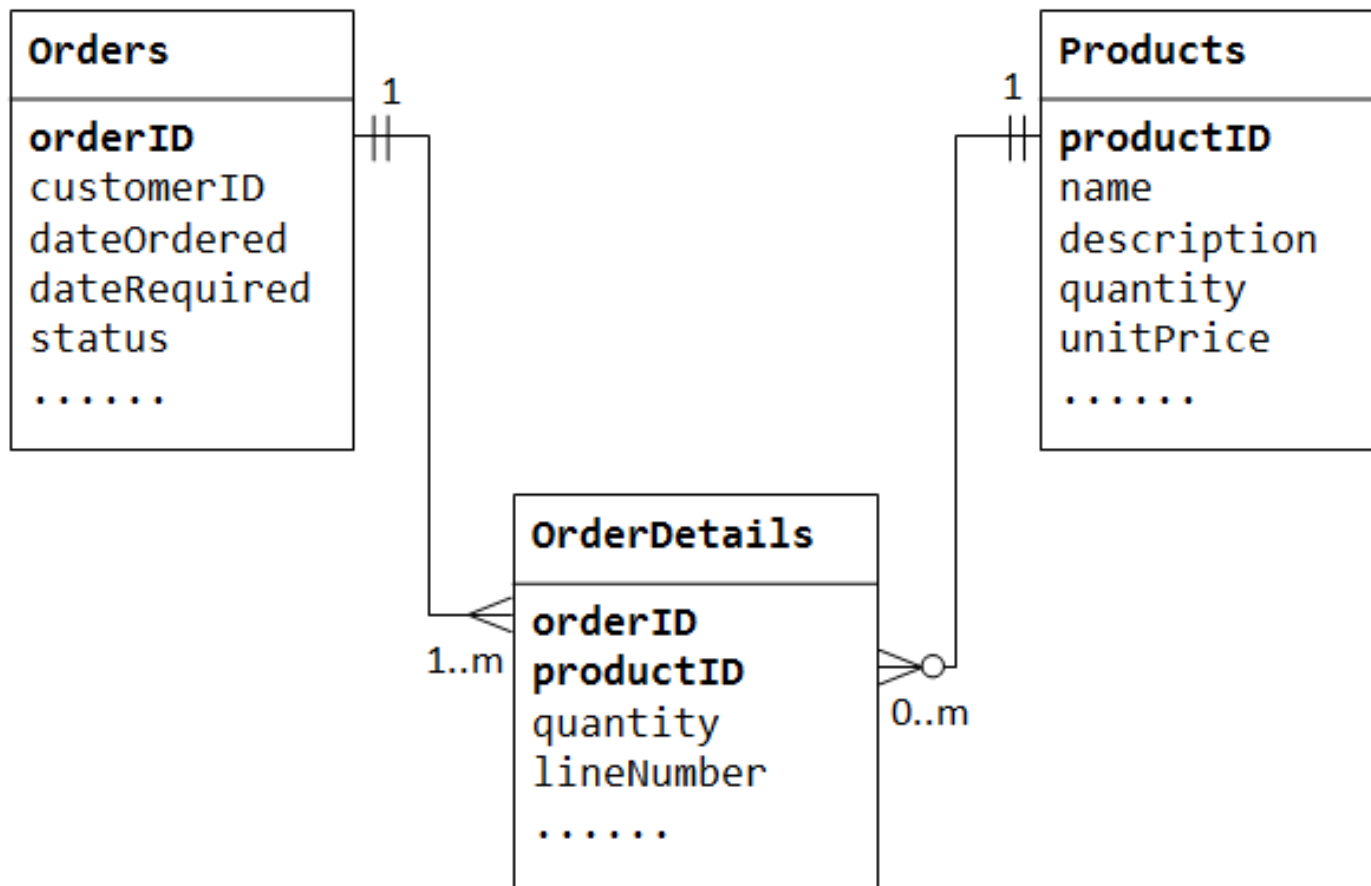
Mối quan hệ (relationships)

Quan hệ n-n (nhiều – nhiều)

- Ví dụ: quan hệ giữa hai thực thể **Order** và **Product** là **N-N** vì mỗi đơn đặt hàng có thể gồm nhiều sản phẩm, và ngược lại mỗi sản phẩm có thể xuất hiện ở nhiều đơn đặt hàng
- Phần dữ liệu giao nhau cho biết cụ thể **Số lượng đặt hàng**, **giá đặt** và **chiết khấu** bao nhiêu.

Mối quan hệ (relationships)

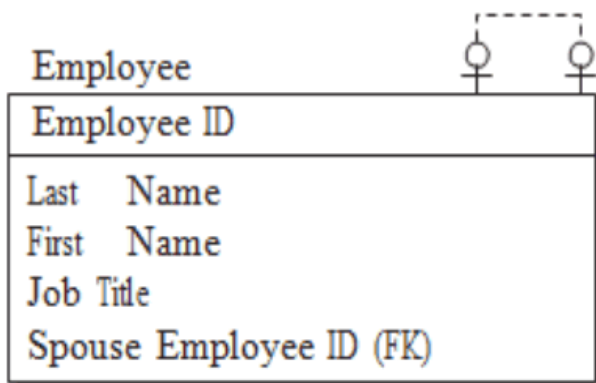
Quan hệ n-n (nhiều – nhiều)



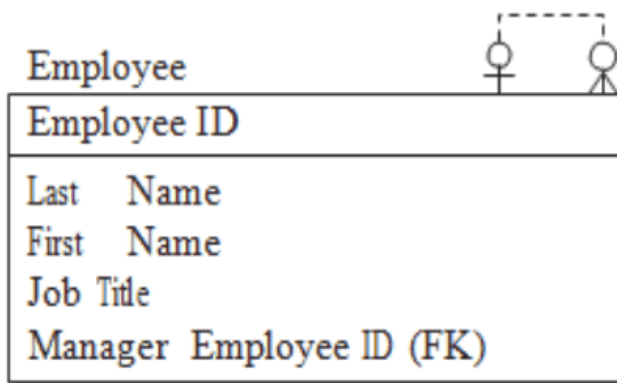
Mối quan hệ (relationships)

Quan hệ đệ quy

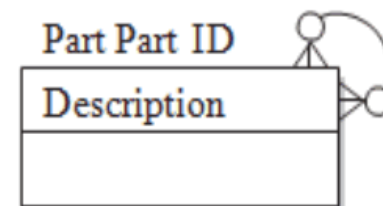
- Là quan hệ tồn tại giữa hai thực thể thuộc cùng một tập thực thể.
- Phân loại: 1-1, 1-N, N-N.



Quan hệ 1-1



Quan hệ 1-N






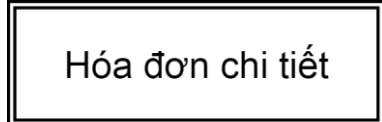
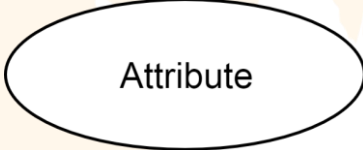
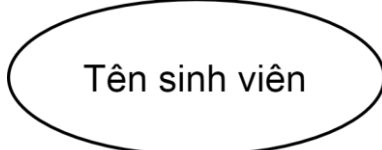
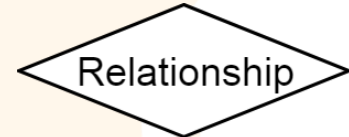
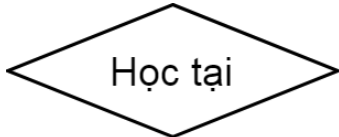
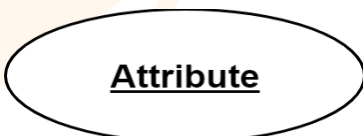
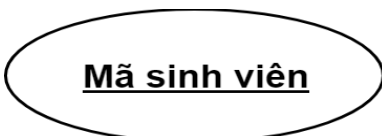
Quan hệ N-N

Sơ đồ E-R

- Là sự biểu diễn đồ họa của **mô hình E-R**.
- Sơ đồ sử dụng nhiều **ký hiệu, hình vẽ** để giúp biểu diễn các thành phần khác nhau trong mô hình quan hệ thực thể - liên kết.
- Sơ đồ E-R là cần thiết trong quá trình xây dựng CSDL cho sản phẩm phần mềm.

Sơ đồ E-R

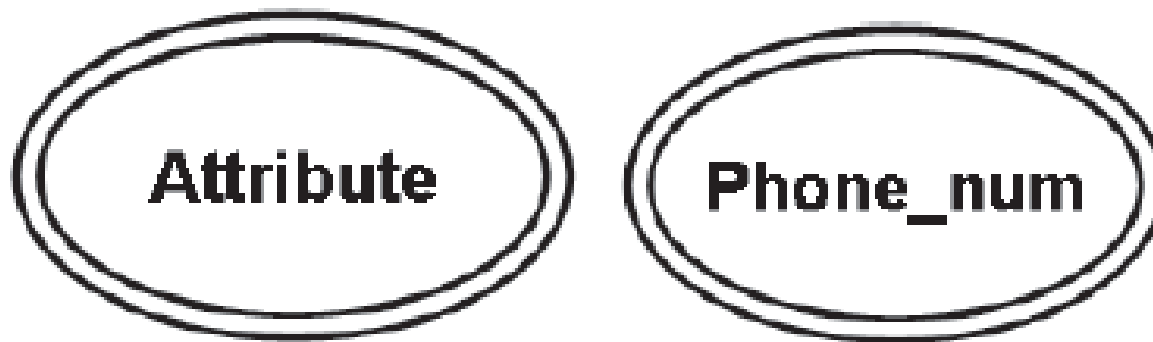
Ký hiệu biểu diễn các thành phần trong sơ đồ:

Thành phần	Ký hiệu	Ví dụ
Entity – thực thể		
Weak Entity – thực thể yếu		
Attribute – thuộc tính		
Relationship – quan hệ		
Key Attribute – khóa chính		

Sơ đồ E-R

Đa trị (Multi-valued)

- Thuộc tính đa trị là: một thuộc tính có **nhều hơn một giá trị** cho một thực thể, ký hiệu là một hình **Elip nét đôi**.
- VD: thuộc tính số điện thoại, sở thích của Sinh Viên có thể nhiều giá trị.



Sơ đồ E-R

Đa trị (Multi-valued)

- Trong bảng sau: để thể hiện sở thích có thể đặt giá trị trong cột **So_thích**. Tuy nhiên điều này dễ dẫn tới **khó khăn trong truy vấn** dữ liệu.

Ma_sinhvien	Ho_sinhvien	Ten_sinhvien	So_thích
1639	George	Barnes	Đọc sách
5629	Susan	Noble	Đi bộ, xem phim
3388	Erwin	Star	Bóng đá, lướt ván
5772	Alice	Buck	
1911	Frank	Borders	Chụp ảnh, du lịch, vẽ
4848	Hanna	Diedrich	Xem phim

Sơ đồ E-R

Đa trị (Multi-valued)

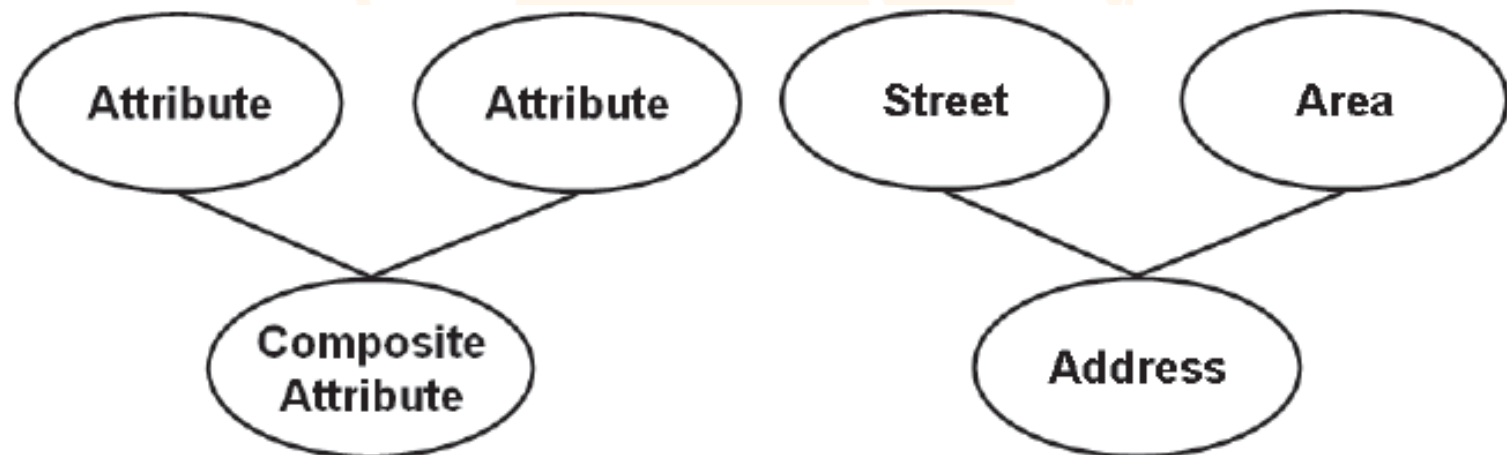
- Với thể hiện thuộc tính **"Sở thích"** là đa trị cho thực thể **"Sinh Viên"**, một **bảng** dữ liệu sẽ đảm nhiệm riêng biệt để lưu trữ sở thích.

Ma_sinhvien	So_thich
1639	Đọc sách
5629	Đi bộ
5629	Xem phim
3388	Bóng bầu dục
3388	Lướt ván
1911	Chụp ảnh
1911	Du lịch
1911	Vẽ
4848	Nấu ăn

Sơ đồ E-R

Kết hợp (Composite)

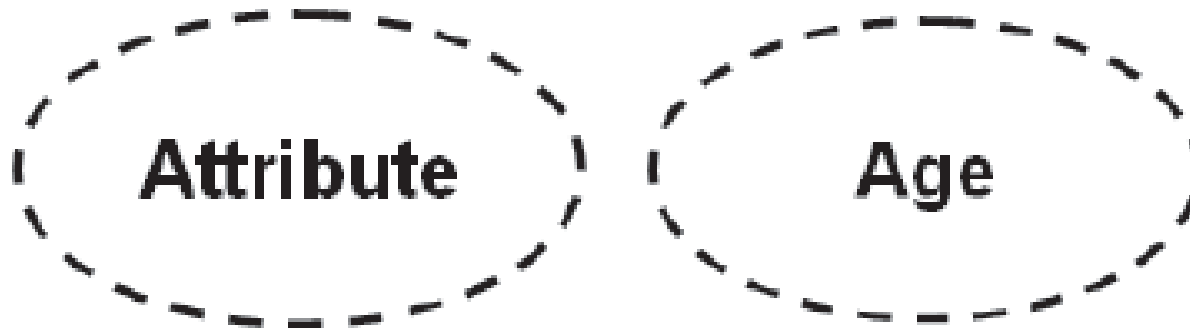
- Thuộc tính kết hợp là tự nó có thể chứa **hai hay nhiều thuộc tính**, đại diện cho thuộc tính cơ bản có ý nghĩa độc lập.
- VD: thuộc tính địa chỉ là **kết hợp thuộc tính**: số nhà, phường, xã, quận, thành phố....



Sơ đồ E-R

Dẫn xuất (Derived)

- Thuộc tính dẫn xuất là giá trị của nó được **suy ra** từ thuộc tính khác, ký hiệu là **Elip nét đứt**.
- VD: thuộc tính **tuổi** được suy ra từ **thuộc tính ngày sinh**, thuộc tính tổng tiền hóa đơn suy ra từ: **giá sản phẩm * số lượng**.



Sơ đồ E-R

Các bước xây dựng sơ đồ E-R

1

- Thu thập tất cả các dữ liệu mà cần được mô hình.

2

- Nhận diện dữ liệu là các thực thể trong thế giới thực có thể được mô hình.

3

- Xác định các thuộc tính cho mỗi thực thể.

4

- Sắp xếp các tập thực thể yếu, các tập thực thể mạnh.

5

- Sắp xếp các thuộc tính thực thể như là các thuộc tính khoá, các thuộc tính đa trị, các thuộc tính kết hợp, các thuộc tính dẫn xuất.

6

- Xác định các mối quan hệ giữa các thực thể.

Chuẩn hóa

- Ban đầu, các csdl được đặc trưng bởi **số lượng lớn** các cột và bản ghi. Phương pháp này có nhiều hạn chế.
- Bảng sau thể hiện thông tin chi tiết của nhân viên và dự án họ tham gia – **đây là bảng chưa được chuẩn hóa.**

Ma_nhanvien	Ma_project	Ten_project	Ten_nhanvien	Cap_bac	Luong
142	113, 124	STAR, MAG	Hùng	A	20.000
168	113	STAR	Đức	B	15.000
263	113	STAR	Nam	C	10.000
109	124	MAG	Tuấn	C	10.000

Chuẩn hóa

CSDL khi chưa chuẩn hóa sẽ gặp một số vấn đề sau:

Dị thường trùng lặp

- Dữ liệu **Ma_project, Ten_project, Cap_bac, Luong** bị lặp đi lặp lại nhiều lần (**cột bôi vàng**). Việc này làm ảnh hưởng tới hiệu suất lưu trữ và lấy dữ liệu.

Ma_nhanvien	Ma_project	Ten_project	Ten_nhanvien	Cap_bac	Luong
142	113, 124	STAR, MAG	Hùng	A	20.000
168	113	STAR	Đức	B	15.000
263	113	STAR	Nam	C	10.000
109	124	MAG	Tuấn	C	10.000

Chuẩn hóa

Dị thường khi chèn

- Giả sử công ty có nhân viên mới là Ngân, cô chưa được phân dự án nào nên dữ liệu tại cột liên quan **bị trống (null)**. Điều đó sẽ dẫn tới nhiều vấn đề khi truy vấn sau này.

Ma_nhanvien	Ma_project	Ten_project	Ten_nhanvien	Cap_bac	Luong
142	113, 124	STAR, MAG	Hùng	A	20.000
168	113	STAR	Đức	B	15.000
263	113	STAR	Nam	C	10.000
109	124	MAG	Tuấn	C	10.000
195			Ngân	C	10.000

Chuẩn hóa

Dị thường khi xóa

- Giả sử **Tuấn rời khỏi dự án MAG** (do nghỉ việc).
- Việc xóa bản ghi đồng nghĩa **mất toàn bộ thông tin** của Tuấn.
- Gây ra khó khăn khi tra cứu sau này.

Ma_nhanvien	Ma_project	Ten_project	Ten_nhanvien	Cap_bac	Luong
142	113, 124	STAR, MAG	Hùng	A	20.000
168	113	STAR	Đức	B	15.000
263	113	STAR	Nam	C	10.000
109	124	MAG	Tuấn	C	10.000

Chuẩn hóa

Dị thường khi cập nhật

- Giả sử công ty có chính sách nâng **bậc C lên \$11.000**. Việc cập nhật này buộc phải thực hiện trên **tất cả** bản ghi của nhân viên có bậc C.

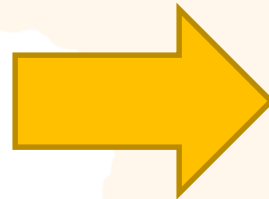
Ma_nhanvien	Ma_project	Ten_project	Ten_nhanvien	Cap_bac	Luong
142	113, 124	STAR, MAG	Hùng	A	20.000
168	113	STAR	Đức	B	15.000
263	113	STAR	Nam	C	10.000
109	124	MAG	Tuấn	C	10.000

Chuẩn hóa

Việc chuẩn hóa là:

- Loại bỏ **sự dư thừa và phụ thuộc** không mong muốn.
- Có 3 dạng chuẩn cơ bản là: **1NF, 2NF và 3NF**.
- Các dạng chuẩn nâng cao thứ tư, thứ 5 dựa trên đa giá trị và ghép các sự phụ thuộc.

Course	Content
Programming	Java, c++
Web	HTML, PHP, ASP



Chuẩn hóa

Course	Content
Programming	Java
Programming	c++
Web	HTML
Web	PHP
Web	ASP

Chuẩn hóa

Chuẩn 1NF:



- Tạo ra các bảng riêng biệt cho mỗi nhóm các dữ liệu liên quan.
- Các cột của bảng phải có giá trị nguyên tố (đơn trị).
- Tất cả các thuộc tính khóa phải được nhận diện.

=> Bảng **EmployeeProjectDetail** cần tách thành 2 bảng: **ProjectDetail** và **EmployeeDetail**

Ma_project	Ten_project
113	STAR
124	MAG

Ma_nhanvien	Ten_nhanvien	Cap_bac	Luong
142	Hùng	A	20.000
168	Đức	B	15.000
263	Nam	C	10.000
109	Tuấn	C	10.000

Chuẩn hóa

Chuẩn 2NF:

Chúng thỏa mãn các yêu cầu của dạng chuẩn một.

Không có phụ thuộc hàm bộ phận. (Tất cả các thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào khóa)

Các bảng được thiết lập quan hệ thông qua các khóa ngoại.

- Ở chuẩn 2, một bảng **EmployeeProjectDetail** cần được tạo để lưu trữ thông tin nhân viên với dự án liên quan.

Chuẩn hóa

Chuẩn 2NF:

Ma_project	Ten_project
113	STAR
124	MAG

Bảng ProjectDetail

Ma_nhanvien	Ma_project
142	113
142	124
168	113
263	113
109	124

Bảng EmployeeProjectDetail

Ma_nhanvien	Ten_nhanvien	Cap_bac	Luong
142	Hùng	A	20.000
168	Đức	B	15.000
263	Nam	C	10.000
109	Tuấn	C	10.000

Bảng EmployeeDetail

Chuẩn hóa

Chuẩn 3NF:

Các bảng thỏa mãn yêu cầu dạng chuẩn hai

Không tồn tại phụ thuộc bắc cầu trong các bảng.

Sau khi chuẩn 2NF, bảng **EmployeeDetail** vẫn tồn tại phụ thuộc bắc cầu ở cột **Cấp Bậc** và **Lương**.

Ma_nhanvien	Ten_nhanvien	Cap_bac
142	Hùng	A
168	Đức	B
263	Nam	C
109	Tuấn	C

Bảng EmployeeDetail

Cap_bac	Luong
A	20.000
B	15.000
C	10.000

Bảng SalaryDetail

Chuẩn hóa

Kết quả sau chuẩn hóa:

Ma_project	Ten_project
113	STAR
124	MAG

Bảng ProjectDetail

Ma_nhanvien	Ma_project
142	113
142	124
168	113
263	113
109	124

Bảng EmployeeProjectDetail

Ma_nhanvien	Ten_nhanvien	Cap_bac
142	Hùng	A
168	Đức	B
263	Nam	C
109	Tuấn	C

Bảng EmployeeDetail

Cap_bac	Luong
A	20.000
B	15.000
C	10.000

Bảng SalaryDetail

Tóm tắt bài học

- **Mô hình E-R** xem thế giới thực là một tập hợp các đối tượng cơ bản và mối quan hệ giữa chúng.
- **Thực thể, các thuộc tính**, tập hợp thực thể, các mối quan hệ và các bộ mối quan hệ hình thành nên thành phần cơ bản của mô hình E-R.
- Quá trình loại bỏ dữ liệu dư thừa từ các bảng của một cơ sở dữ liệu quan hệ được gọi là **chuẩn hóa**.

TRƯỜNG ĐÀO TẠO LẬP TRÌNH VIÊN VÀ QUẢN TRỊ MẠNG QUỐC TẾ BACHKHOA-APTECH

Thank for watching!

