CƠ SỞ DỮ LIỆU

CHƯƠNG 3. PHÂN TÍCH VÀ THIẾT KẾ CSDL

NỘI DUNG

- 1. Vòng đời phát triển của hệ thống CSDL
- 2. Phân tích CSDL trên bài toán thực tế
- 3. Mô hình thực thể liên hệ (ER)
- 4. Chuẩn hóa

NỘI DUNG

- 1. Vòng đời phát triển của hệ thống CSDL
- 2. Phân tích CSDL trên bài toán thực tế
- 3. Mô hình thực thể liên hệ (ER)
- 4. Chuẩn hóa

1. VÒNG ĐỜI PHÁT TRIỂN CỦA HỆ THỐNG CSDL

- Vòng đời của hệ thống thông tin
- Vòng đời phát triển của hệ thống CSDL
- Các công cụ hỗ trợ

VÒNG ĐỜI CỦA HỆ THỐNG THÔNG TIN

Information system

IS

Hệ thống thông tin là một cơ cấu có tổ chức của

- con người,
- · cơ sở hạ tầng,
- dữ liệu (thông tin),
- quy chế & thủ thục (quy trình nghiệp vụ),
- giải pháp phần mềm...

để lưu trữ, xử lý, chuyển đổi và phân phối thông tin nhằm phục vụ các mục tiêu hoạt động trong doanh nghiệp

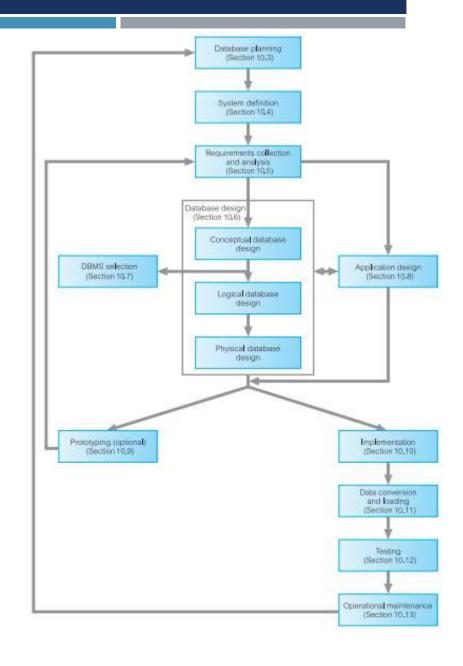
- HTTT dựa trên máy tính bao gồm:
 - Cơ sở dữ liệu
 - phần mềm cơ sở dữ liệu
 - phần mềm ứng dụng
 - phần cứng máy tính
 - và nhân sự sử dụng và phát triển hệ thống
- CSDL là một thành phần cơ bản của HTTT
- Vòng đời của một HTTT được liên kết với vòng đời của hệ thống CSDL hỗ trợ cho HTTT đó

VÒNG ĐỜI CỦA HỆ THỐNG THÔNG TIN

- Vòng đời của HTTT:
 - Lập kế hoạch planning
 - Thu thập yêu cầu và phân tích yêu cầu requirements collection and analysis
 - Thiết kế design
 - Tạo mẫu prototyping
 - Cài đặt implementation
 - Kiểm thử testing
 - Chuyển đổi conversion
 - Vận hành và bảo trì operational maintenance

VÒNG ĐỜI PHÁT TRIỂN CỦA HỆ THỐNG CSDL

- Vòng đời phát triển của hệ thống CSDL
 - Lập kế hoạch CSDL
 - Phạm vi hệ thống
 - Phân tích và tập hợp các yêu cầu
 - Thiết kế CSDL
 - Lựa chọn hệ quản trị CSDL
 - Thiết kế ứng dụng
 - Tạo mẫu (prototyping) nếu cần
 - Cài đặt
 - Nạp và chuyển đổi dữ liệu
 - Kiểm thử
 - Bảo trì vận hành



VÒNG ĐỜI PHÁT TRIỂN CỦA HỆ THỐNG CSDL

Database planning

Lập kế hoạch CSDL: xác định các trạng thái trong vòng đời phát triển hệ thống CSDL có tính hiệu quả nhất

- Lập kế hoạch CSDL phải tích hợp với chiến lược HTTT tổng thể của tổ chức.
- Các vấn đề liên quan đến xác định chiến lược của HTTT:
 - Xác định các kế hoạch và mục tiêu của doanh nghiệp cùng với việc xác định HTTT cần có;
 - Đánh giá các HTTT hiện tại để xác định các thế mạnh hiện có và những điểm yếu;
 - Đánh giá các cơ hội về CNTT có thể mang lại lợi thế cạnh tranh.

VÒNG ĐỜI PHÁT TRIỂN CỦA HỆ THỐNG CSDL

Trạng thái	Hoạt động chính
Lập kế hoạch CSDL	Lập kế hoạch làm thế nào các giai đoạn của vòng đời có thể được thực hiện một cách hiệu quả và hiệu quả nhất.
Phạm vi hệ thống	Chỉ định phạm vi và ranh giới của hệ thống cơ sở dữ liệu, bao gồm người dùng và phạm vi ứng dụng.
Phân tích và tập hợp các yêu cầu	Thu thập và phân tích các yêu cầu cho hệ thống cơ sở dữ liệu mới
Thiết kế CSDL	Thiết kế CSDL các mức khái niệm, logic và vật lý
Lựa chọn hệ quản trị CSDL	Lựa chọn hệ quản trị CSDL phù hợp với hệ thống CSDL
Thiết kế ứng dụng	Thiết kế giao diện người dùng và chương trình ứng dụng mà có sử dụng hệ thống CSDL
Tạo mẫu – nếu cần	Xây dựng mô hình làm việc của hệ thống CSDL, cho phép người thiết kế hoặc người dùng hình dung và đánh giá hệ thống cuối sẽ trông như thế nào và hoạt động ra sao
Cài đặt	Tạo các định nghĩa CSDL ở mức vật lý và các chương trình ứng dụng
Nạp và chuyển đổi dữ liệu	Nạp dữ liệu từ hệ thống cũ sang hệ thống mới và chuyển đổi bất kỳ ứng dụng hiện có để chạy trên CSDL mới (nếu có thể)
Kiểm thử	Hệ thống CSDL được kiểm tra lỗi và xác nhận theo các yêu cầu được chỉ định bởi người dùng
Bảo trì vận hành	Hệ thống được theo dõi liên tục và duy trì. Khi cần thiết, các yêu cầu mới được đưa vào hệ thống CSDL thông qua các giai đoạn trước của vòng đời.

NỘI DUNG

- 1. Vòng đời phát triển của hệ thống CSDL
- 2. Phân tích CSDL trên bài toán thực tế
- 3. Mô hình thực thể liên hệ (ER)
- 4. Chuẩn hóa

DreamHome

Fact - finding

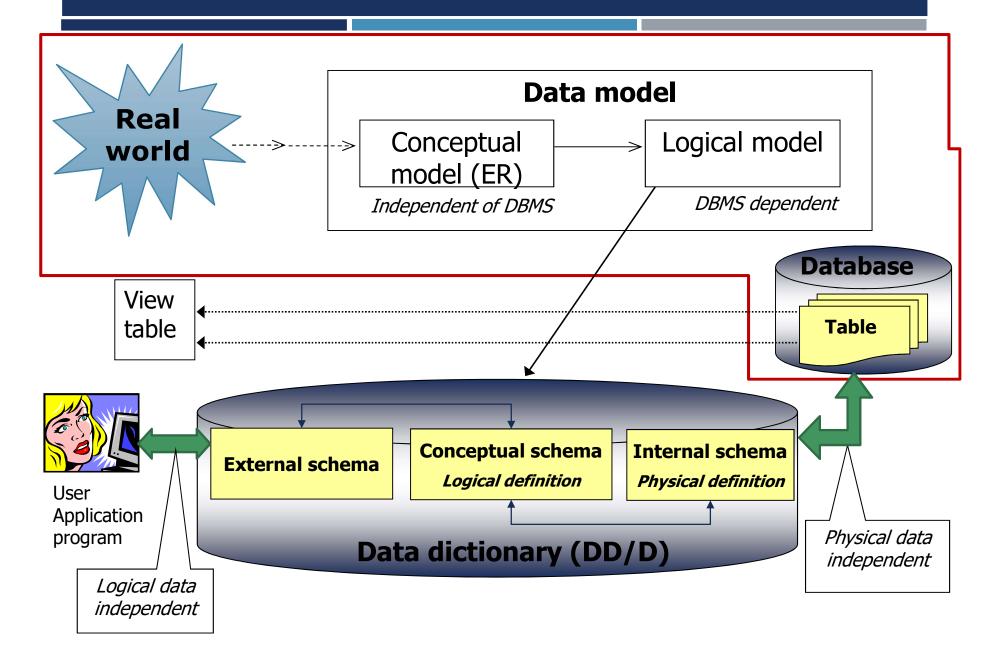
Tìm hiểu thực tế: thu thập thông tin về hệ thống, yêu cầu và mục đích của hệ thống.

Một số kỹ thuật tìm hiểu thực tế như

- Kiểm tra tài liệu;
- Phỏng vấn;
- Quan sát doanh nghiệp đang hoạt động;
- Nghiên cứu;
- Bảng hỏi

Trạng thái	Dữ liệu cần thu được Tài liệu đầu ra		
Lập kế hoạch CSDL	Mục đích và mục tiêu của CSDL	Mục tiêu và sứ mệnh của hệ thống CSDL	
Phạm vi hệ thống	Mô tả các khung nhìn của người dùng (bao gồm phạm vi công việc hoặc lĩnh vực ứng dụng)	Định nghĩa phạm vi và ranh giới của hệ thống CSDL; định nghĩa khung nhìn hỗ trợ cho người dùng	
Phân tích và tập hợp các yêu cầu	Các yêu cầu của người dùng, đặ tả hệ thống, bao gồm hiệu năng và yêu cầu bảo mật	Đặc tả yêu cầu của người dùng và hệ thống	
Thiết kế CSDL	Trách nhiệm của người dùng để kiểm tra các mức khái niệm/logic của thiết kế CSDL;	Thiết kế mức Khái niệm/logic (bao gồm môn hình ER, từ điển dữ liệu, lược đồ quan hệ), thiết kế CSDL ở mức vật lý	

Trạng thái	Dữ liệu cần thu được	Tài liệu đầu ra		
Lựa chọn hệ quản trị CSDL	Các chức năng được cung cấp bởi Hệ QTCSDL	Đánh giá Hệ QTCSDL và khuyến nghị		
Thiết kế ứng dụng	Phản hổi của người dùng về thiết kế giao diện	Thiết kế ứng dụng (bao gồm mô tả của chương rình và giao diện người dùng)		
Tạo mẫu	Phản hồi của người dùng về tạo mẫu	Chỉnh sửa đặc tả người dùng và hệ thống		
Cài đặt	Các chức năng được cung cấp bởi Hệ QTCSDL			
Nạp và chuyển đổi dữ liệu	Định dạng dữ liệu hiện hành, Nạp dữ liệu vào Hệ QTCSDL			
Kiểm thử	Kết quả kiểm thử Chiến lược kiểm thử đã sử dụng; phâ tích kết quả kiểm thử			
Bảo trì vận hành	Kết quả kiểm tra hiệu suất; mới hoặc thay đổi yêu cầu người dùng và hệ thống	Hướng dẫn sử dụng; phân tích hiệu suất; chỉnh sửa yêu cầu đặ tả của người dùng và hệ thống		



DreamHome Case study (1)

- DreamHome chuyên quản lý bất động sản, giữ vai trò trung gian giữa chủ nhà và khách hàng của DreamHome (người thuê bất động sản trong một thời gian cố định)
- DreamHome hiện có khoảng 2000 nhân viên làm việc tại 100 chi nhánh.
- Khi một thành viên của nhân viên gia nhập Công ty, công ty sẽ gửi mẫu đăng ký nhân viên gồm các thông tin như hình



- DreamHome Case study (2)
 - Mỗi chi nhánh có số lượng và loại nhân viên phù hợp bao gồm Quản lý, Giám sát viên, và Trợ lý. Người quản lý chịu trách nhiệm điều hành hoạt động hàng ngày của chi nhánh và mỗi Giám sát viên chịu trách nhiệm giám sát một nhóm nhân viên gọi là Trợ lý.



- DreamHome Case study (3)
 - Mỗi chi nhánh cung cấp một số các tài sản cho thuê.
 - Chủ nhà cho thuê thường liên lạc với chi nhánh DreamHome gần nhất. Chủ nhà cung cấp các chi tiết của ngôi nhà và đồng ý một giá thuê thích hợp cho tài sản với Giám đốc chi nhánh.
 - Mẫu đăng ký nhà cho thuê:



DreamHome Case study (4)

- Khi một tài sản được đăng ký, DreamHome tổ chức cho khách hàng (người thuê nhà) xem nhà, quảng cáo nhà tại địa phương hoặc báo quốc gia (khi cần thiết), và đàm phán cho thuê.
- Sau khi cho thuê, DreamHome chịu trách nhiệm về tài sản bao gồm cả việc thu tiền thuê.
- Người muốn thuê bất động sản phải liên hệ với chi nhánh
 DreamHome gần nhất để đăng ký trở thành khách hàng.
- Mẫu đăng ký khách hàng



- DreamHome Case study (5)
 - Sau khi đăng ký hoàn tất, KH được cung cấp các báo cáo hàng tuần liệt kê tài sản hiện có sẵn cho thuê.
 - Khi KH xác định được nhà thuê phù hợp, nhân viên sẽ đưa ra hợp đồng thuê

roperty Numner PG4 ype Flat			Property Address 6 Lawrence St, Glasgow		
lent 350					
	Name	Date	Comments		
Client No	Name John Kay	Date 20/04/13	Comments		
	Name John Kay Aline Stewart	Date 20/04/13 26/05/13	Comments Too remote.		
Client No	John Kay	20/04/13			

DreamHome Lease Number 00345810					
Client Number CR74 (Enter if known) Full Name Mike Ritchie (Please print) Client Signature	Property Number PG16 Property Address 5 Novar Dr, Glasgow				
Enter payment details Monthly Rent 450 Payment Method Chaque Deposit Paid (Y or N) Yes	Rent Start				

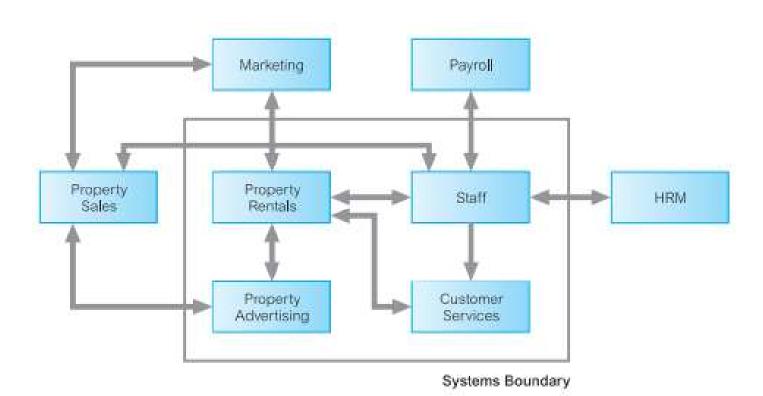
- Database Planning Lập kế hoạch CSDL
 - Phỏng vấn Giám đốc và nhân viên thích hợp theo chỉ định của Giám đốc
 - Mục đích của công ty bạn là gì?
 - Tại sao bạn cảm thấy rằng bạn cần một CSDL?
 - Làm thế nào để bạn biết rằng một CSDL sẽ giải quyết vấn đề của bạn?

...

Mục đích của hệ thống CSDL DreamHome là duy trì dữ liệu để hỗ trợ công việc cho thuê bất động sản của KH và chủ nhà và tạo điều kiện cho việc hợp tác và chia sẻ thông tin giữa các chi nhánh

Mục tiêu và sứ mệnh của hệ thống CSDL

- System Definition Phạm vi hệ thống
 - System Boundary



 Requirements Collection and Analysis - Phân tích và tập hợp các yêu cầu

Data	Access Type	Director	Manager	Supervisor	Assistant	Client
All Branches	Maintain					
	Query	×	X			
	Report	×	x			
Single Branch	Maintain		X			
	Query		X			
All Chaff	Report		X			
All Staff	Maintain					
	Query	X	х			
	Report	×	X			
Branch Staff	Maintain		X			
Branch Stan	Query		X	X		
	Report		X	X		
All Property	Maintain					
	Query	X				Х
	Report	×	X			X
Branch Property	Maintain		X	X		
	Query		x	×	×	
	Report		×	×	X	
All Owners	Maintain					
	Query	×				
	Report	×	X			-
Branch Owners	Maintain		х	х		
	Query		х	×	×	
	Report		X			

■ Database Design - Thiết kế CSDL

NỘI DUNG

- 1. Vòng đời phát triển của hệ thống CSDL
- 2. Phân tích CSDL trên bài toán thực tế
- 3. Mô hình thực thể liên hệ (ER)
- 4. Chuẩn hóa

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER)

- Thực thể (Entity)
- Quan hệ (Relationship)
- Thuộc tính (Attributes)
- Thực thể yếu, mạnh
- Thuộc tính trên quan hệ
- Ràng buộc cấu trúc
- Sự tổng quát hóa/ Chuyên môn hóa (Specialization/Generalization)
- Các vấn với mô hình thực thể liên hệ

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) THỰC THỂ

Entity

Thực thể: đối tượng cụ thể hay trừu tượng, tồn tại thực sự và khá ổn định, có thể phân biệt được với nhau

Ví dụ: sinh viên Lê Na, khách Trần Hiếu,

Entity type

Kiểu thực thể (Tập thực thể) nhóm các đối tượng có cùng tính chất Ví dụ: Sinh viên, Khách hàng, Nhân viên...

Mỗi kiểu thực thể gồm:

- Tên thực thể
- Thuộc tính

Tên tập thực thể là Danh từ phản ánh chung các đối tượng cần trừu tượng hóa



3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) QUAN HỆ

Relationship

Quan hệ: mối liên quan có ý nghĩa giữa của từng thực thể xác định

Ví dụ: sinh viên Lê Na học CSDL1

Tập Quan hệ (kiểu quan hệ): tập mối quan hệ Relationship type có ý nghĩa giữa các tập thực thể Ví dụ: sinh viên học môn học



- Quan hệ bao giờ cũng có chiều
- Thường là các động từ

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) QUAN HỆ

Degree of a relationship type

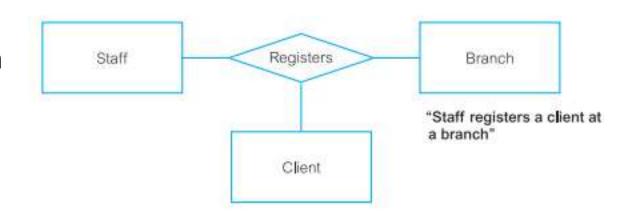
Bậc của tập quan hệ: Số lượng các tập thực thể tham gia trong một mối quan hệ

Quan hệ **nhị phân**



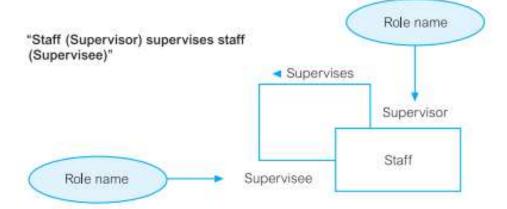
"Private owner owns property for rent"

Quan hệ tam phân

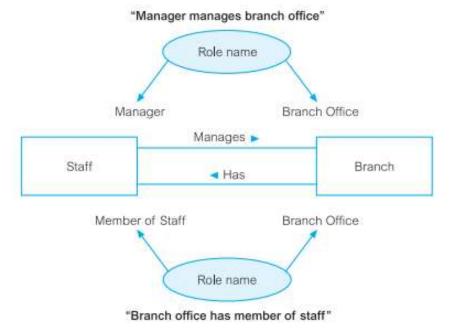


3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) QUAN HỆ

Quan hệ Đơn phân



Đa quan hệ



3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) THUỘC TÍNH

Attributes

Thuộc tính: tính chất của thực thể hoặc mối quan hệ

Ví dụ:

- thuộc tính của thực thể SV gồm: Tên SV, Ngày sinh, Giới tính
- Thuộc tính của mối quan hệ giữa SV và Môn học là Điểm thi

Attribute domain

Miền thuộc tính: tập giá trị cho phép của một hoặc nhiều thuộc tính

Ví dụ:

- Miền của Giới tính của SV là Nam, Nữ
- Miền của Điểm thi từ 1 đến 10

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) THUỘC TÍNH

Simple Attributes

Thuộc tính đơn: bao gồm một thành phần duy nhất (còn gọi là thuộc tính nguyên tố - atomic attribute), không thể chia nhỏ Ví dụ: Giới tính (Nam), Điểm thi (7)

Composite Attributes

Thuộc tính phức: bao gồm nhiều thành phần, mỗi thành phần tồn tại độc lập nhau

Ví dụ: Địa chỉ (Số 30, đường Tràng Thi, Quận Hoàn Kiểm, TP Hà Nội)

Tùy từng mục đích của mỗi bài toán mà ta cỏ thể tách/không tách Địa chỉ thành các thuộc tính Số nhà, Đường, Quận, Thành phố

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) THUỘC TÍNH

Single-valued attribute

Thuộc tính đơn trị: chứa một giá trị cho mỗi lần xuất hiện của một tập thực thể Ví dụ: Mã SV (a01234)

Multi-valued attribute

Thuộc tính đa trị: chứa nhiều giá trị cho mỗi lần xuất hiện của một tập thực thể Ví dụ: Ngoại ngữ (Anh, Pháp, Đức)
Tùy từng mục đích của mỗi bài toán khác nhau mà ta tách thuộc tính đa trị thành nhiều thuộc tính khác nhau hay không.

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) KHÓA

Candidate key

Khóa dự tuyển: Tập hợp các thuộc tính tối thiểu xác định duy nhất một thực thể Ví dụ:

- SV (Mã SV, Họ tên, Ngày sinh, Số CMT)
- Khóa dự tuyển của tập thực thể SV:
 - Mã SV
 - Hoặc Số CMT

Primary key

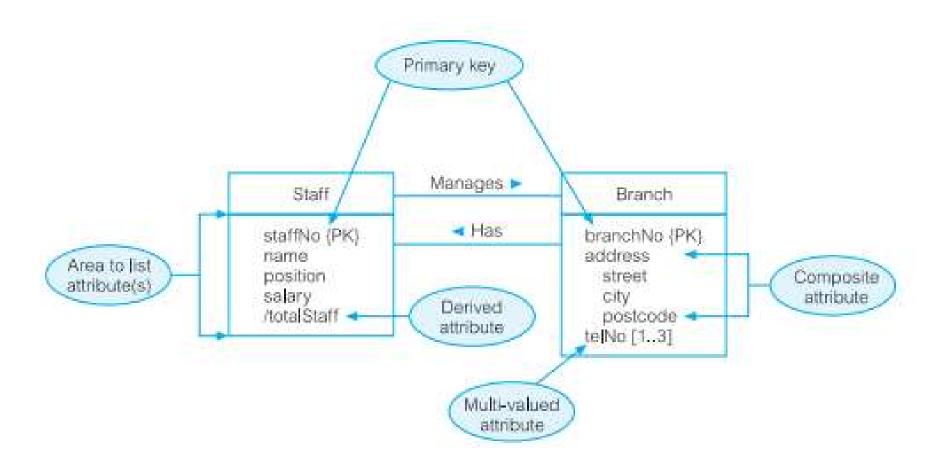
Khóa chính: Khóa dự tuyển được chọn để xác định duy nhật một thực thể Ví dụ Khóa chính của SV là Mã SV

Composite key

Khóa bao hàm: Khóa dự tuyển chứa lớn hơn hai thuộc tính

3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER)

Ký pháp trong mô hình ER



3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) THỰC THỂ MẠNH VÀ THỰC THỂ YẾU

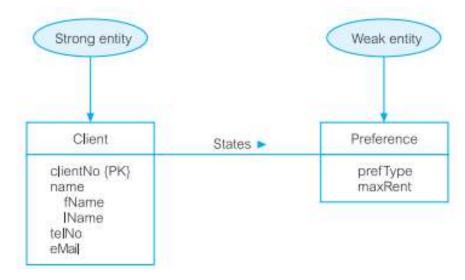
Strong entity type

Thực thể mạnh: thực thể tồn tại không phụ thuộc vào thực thể khác VD: Client

Weak entity type

Thực thể yếu: thực thể tồn tại phụ thuộc vào thực thể khác

VD: Preference



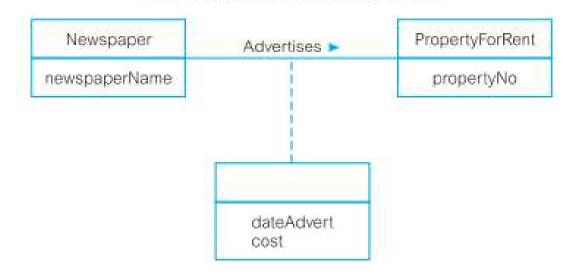
3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) THUỘC TÍNH CỦA MỐI QUAN HỆ

Attributes on Relationships

Thuộc tính của mối quan hệ: Là các đặc tính, tính chất chung của các quan hệ trong 1 tập quan hệ.

Thuộc tính của quan hệ phụ thuộc vào tất cả các tập thực thể tham gia vào quan hệ

"Newspaper advertises property for rent"



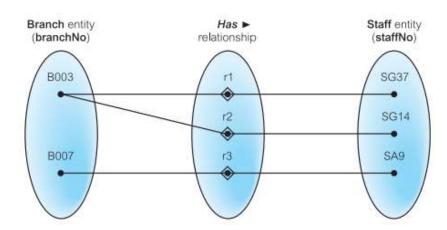
Multiplicity

Lực lượng: Số lần xuất hiện có thể có của tập thực thể

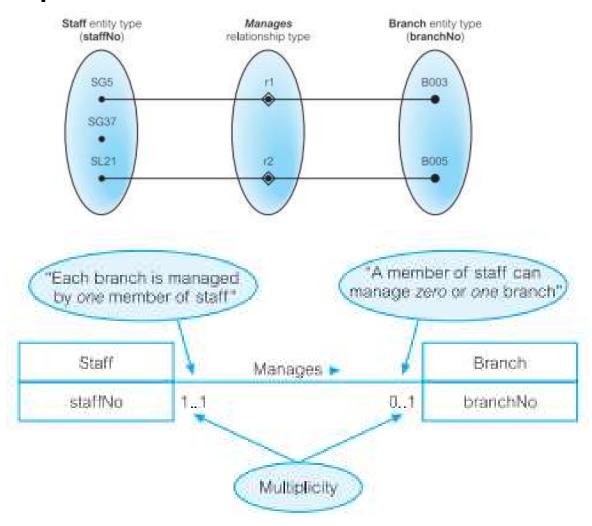
có thể liên quan đến lần xuất hiện của tập thực thể liên quan thông qua mối quan hệ

Với quan hệ nhị phân các lực lượng của mối quan hệ thường là:

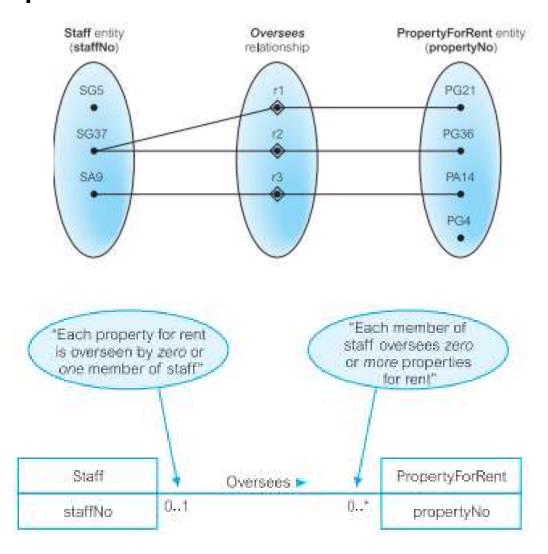
- 1 -1: một một
- 1 *: một nhiều
- * *: nhiều nhiều



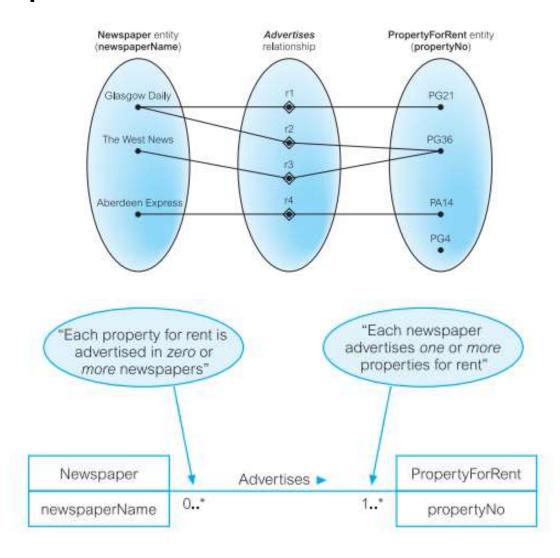
■ Quan hệ 1 – 1



■ Quan hệ 1 – *

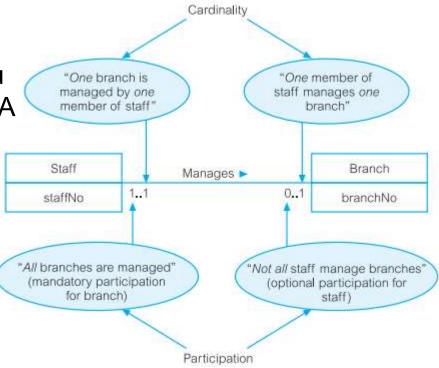


■ Quan hệ * – *



3. MÔ HÌNH THỰC THỂ LIÊN HỆ (ER) RÀNG BUỘC CẤU TRÚC

- Ràng buộc tham gia của quan hệ: Thể hiện cách các tập thực thể tham gia quan hệ.
 - Có hai loại ràng buộc tham gia:
 - Quan hệ tùy chọn: thực thể trong A tồn tại không phải kết hợp với thực thể khác trong B
 - Quan hệ bắt buộc: thực thể trong B được đưa ra khi có yêu cầu kết hợp với thực thể trong A



- Là 2 tiến trình ngược nhau
 - Specialization: thiết kế Top Down nhằm chia 1 lớp cha (Superclass) thành nhiều thực thể lớp con (Subclass)
 - Generalization: thiết kế Bottom Up nhằm gộp vài thực thể có chung một số đặc tính thành một thực thể mức cao hơn.
- Hai cách thiết kế trên kết quả thể hiện trên lược đồ ER giống nhau.
- Thực thể lớp con có thể tham gia
 - trực tiếp vào một quan hệ
 - gián tiếp vào một quan hệ thông qua lớp cha

Superclass

Chứa các thuộc tính chung của các thực thể con

Subclass

 Chứa các thuộc tính của riêng và khóa của tập thực thể cha

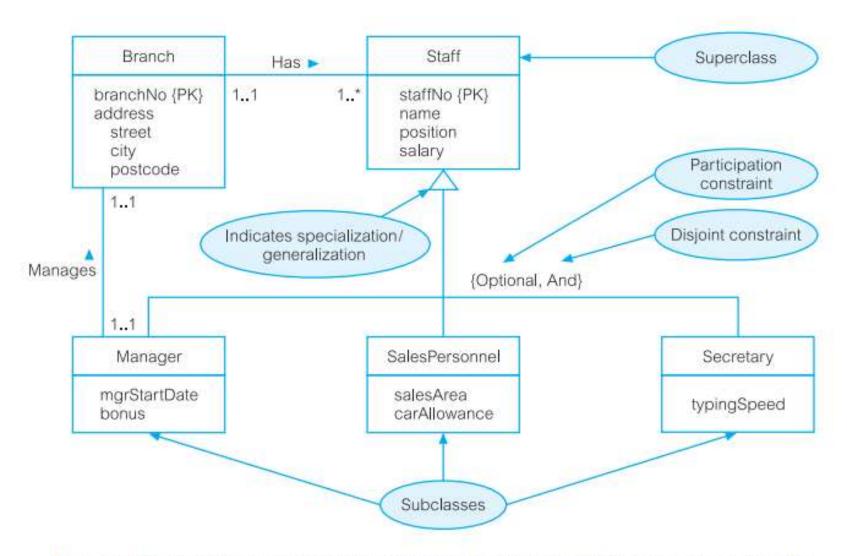
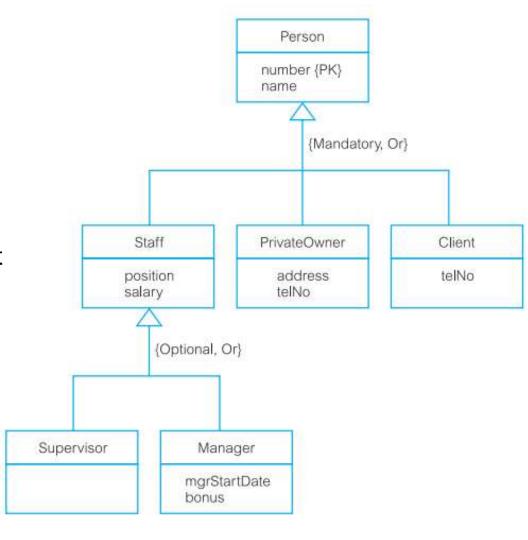


Figure 13.2 Specialization/generalization of the Staff entity into subclasses representing job roles.

Participant constraint

- Mandatory: bắt buộc (mọi thành viên trong lớp cha phải là thành viên của một lớp con)
- Optional: tùy chọn (một thành viên của lớp cha không cần phải thuộc về bất kỳ lớp con nào)
- Disjoint constraint: chỉ áp dụng khi lớp cha có nhiều hơn một lớp con
 - And: lớp cha là thành viên đồng thời của các lớp con
 - Or: lớp cha là một thành viên của 1 trong những lớp con



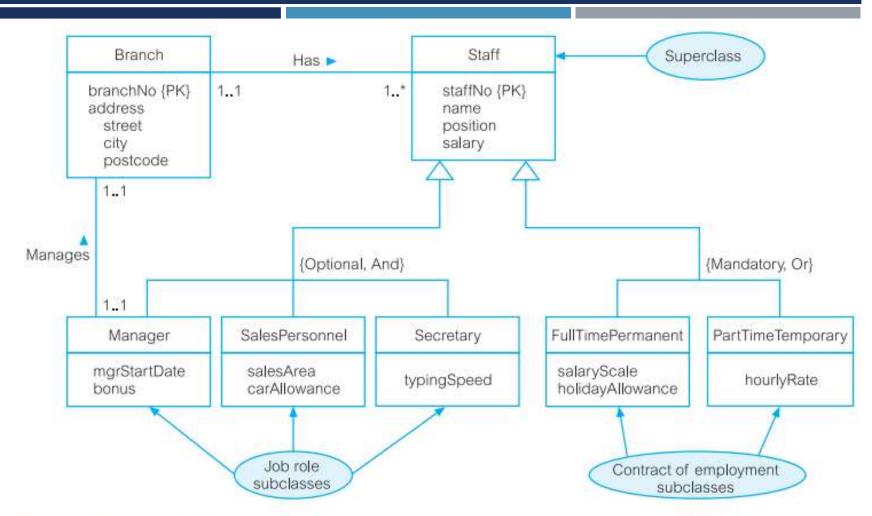


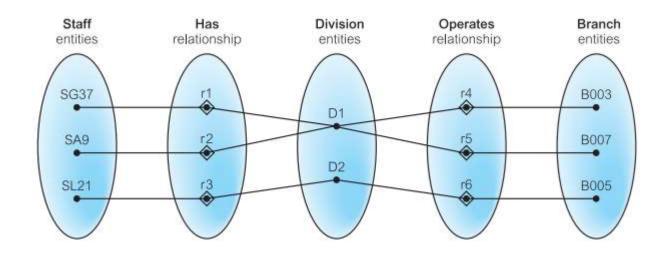
Figure 13.3 Specialization/generalization of the Staff entity into subclasses representing job roles and contracts of employment.

Fan trap

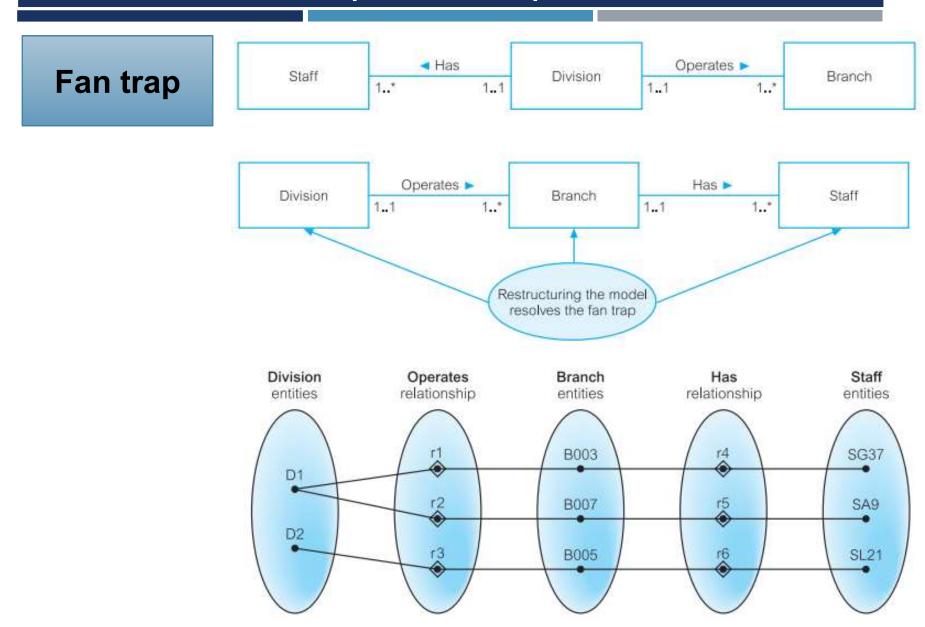
Trong mô hình quan hệ giữa các thực thể, đường nối giữa các thực thể không rõ ràng



Nhân viên SG37 làm ở phòng nào?

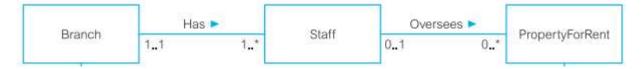


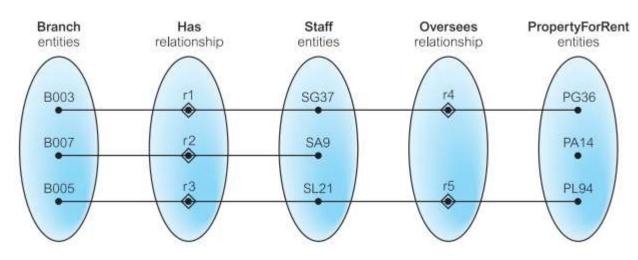
SG37 làm ở chi nhánh B003 *hoặc* B007



Chasm trap

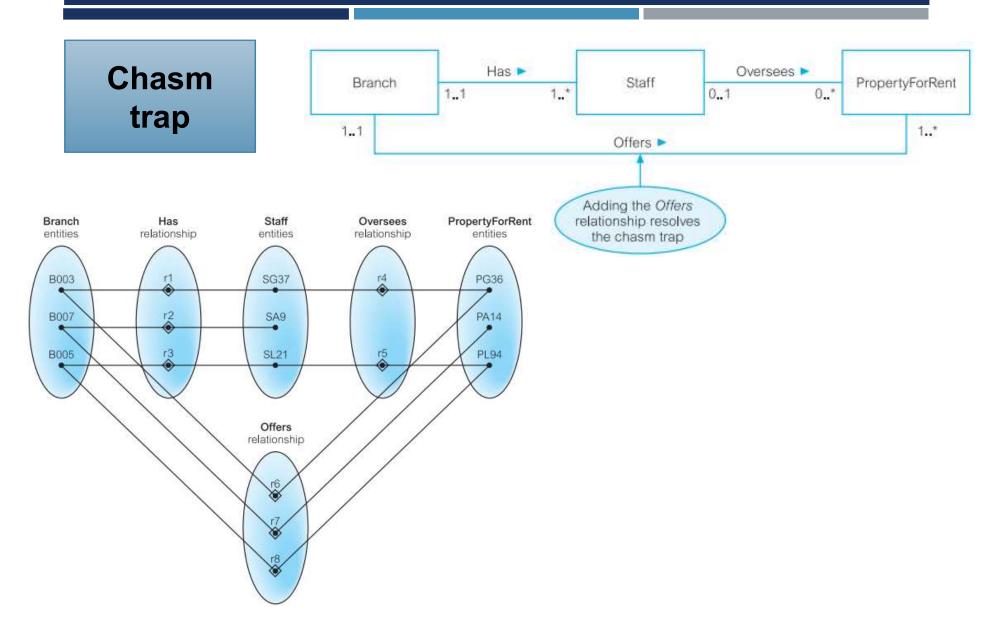
Trong mô hình, có tồn tại mối quan hệ giữa các tập thực thể, nhưng giữa các thực thể nhất định lại không có mối quan hệ



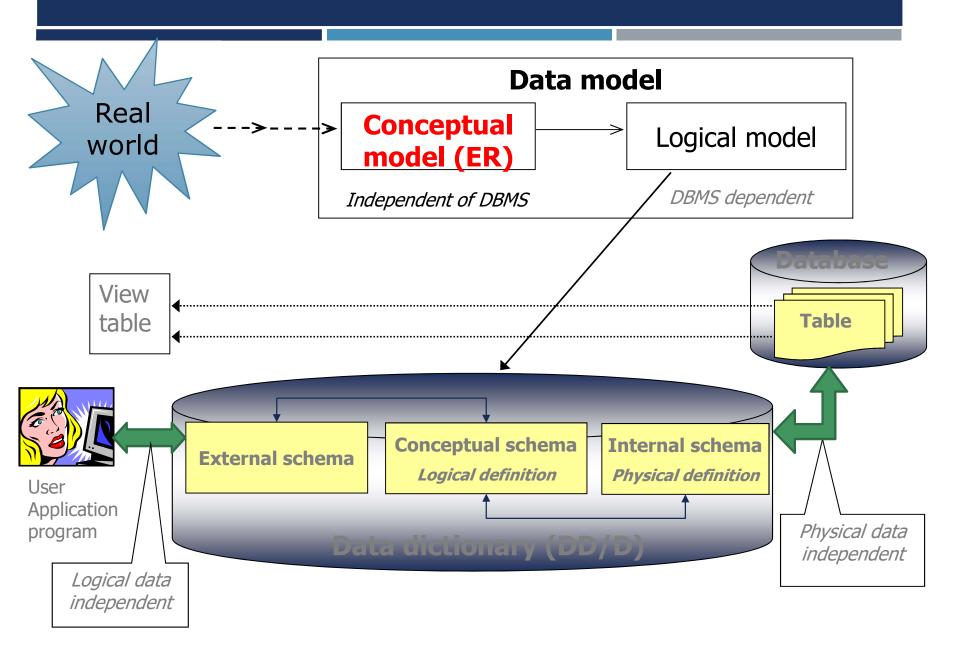


Vì PA14 chưa phân bổ cho bất kỳ nhân viên nào -> Chasm trap

Chi nhánh nào có bất động sản PA14?



CÁC BƯỚC THIẾT KẾ CSDL



- Step 1.1 Identify entity types
- Step 1.2 Identify relationship types
- Step 1.3 Identify and associate attributes with entity or relationship types
- Step 1.4 Determine attribute domains
- Step 1.5 Determine candidate, primary, and alternate key attributes

- Get requirements from user
- Create ER diagram (Conceptual modeling)
 - Nouns -> entities or attributes
 - Verbs -> relationships
- Map from ER diagram to SQL tables (Logical modeling)
 - Entities -> tables
 - Attributes -> tables or attributes
 - Relationship attributes -> tables or attributes

- Draw the ER Diagram for the following online registration system:
 - The university offers courses uniquely identified by department and course number.
 - There may be multiple sections of a given course in a particular semester, including no offerings at all.
 - Sections are taught by professors in rooms (identified by building number and room number).
 - Rooms have specific capacities, but sections may have lower capacity (unused seats).
 - Students enroll in sections.

- Draw the ER Diagram for the following online registration system:
 - The university offers courses uniquely identified by department and course number.
 - There may be multiple sections of a given course in a particular semester, including no offerings at all.
 - Sections are taught by **professors** in rooms (identified by building and room number).
 - Rooms have specific capacities, but sections may have lower capacity (unused seats).
 - Students enroll in sections.

EXAMPLE1 - SELECT ENTITIES (IN RED) (REST ARE ATTRIBUTES OR DISCARDED)

- Draw the ER Diagram for the following online registration system:
 - The university offers courses uniquely identified by department and course number.
 - There may be multiple sections of a given course in a particular semester, including no offerings at all.
 - Sections are taught by professors in rooms (identified by building and room number).
 - Rooms have specific capacities, but sections may have lower capacity (unused seats).
 - Students enroll in sections.

- Course: dept. name, course number, title, credit hours, description, ...
- Section: sectionId, semester, professor name,
 Student: studentId, name, major, ...
- EnrolledIn: relationship between Student-Section
- Offered: relationship between Section-Course
 - Relationship attributes: building name, room number, room capacity

Student - Section

```
*:*
```

Course - Section

```
1:*
```

- Add
 - Primary keys
 - Foreign keys
 - Constraints
 - Domains
 - etc