

CHƯƠNG 2.

CÁC MÔ HÌNH DỮ LIỆU

(Phần 1)

CÁC MÔ HÌNH DỮ LIỆU

- ❖ Giới thiệu
- ❖ Quá trình thiết kế một CSDL
- ❖ Lược đồ thực thể liên kết E-R
- ❖ Một số vấn đề cần quan tâm khi thiết kế lược đồ E-R
- ❖ Lược đồ dữ liệu quan hệ
- ❖ Ánh xạ lược đồ thực thể liên kết sang lược đồ quan hệ

GIỚI THIỆU

- ❖ **Mô hình dữ liệu** là một tập hợp các khái niệm dùng cho việc mô tả và thao tác dữ liệu, các mối quan hệ và các ràng buộc trên dữ liệu của tổ chức.
- ❖ Mô hình là một biểu diễn đối tượng và sự kiện trong thế giới thực, và mối liên hệ của chúng. Mô hình là một khái niệm trừu tượng tập trung vào các khía cạnh bản chất của một tổ chức và bỏ qua những thuộc tính ngẫu nhiên.
- ❖ Mô hình dữ liệu phải cung cấp các khái niệm và ký hiệu cơ bản, cho phép người thiết kế CSDL và người dùng trao đổi với nhau những hiểu biết về dữ liệu của tổ chức một cách chính xác và không đa nghĩa.

GIỚI THIỆU (cont.)

- ❖ Một mô hình dữ liệu có thể bao gồm 3 thành phần:
 1. **Phần cấu trúc** gồm tập các luật mà theo đó CSDL được xây dựng.
 2. **Phần thao tác, định nghĩa các thao tác** được phép trên dữ liệu (gồm cả các thao tác cập nhật, lấy dữ liệu từ CSDL, và thay đổi cấu trúc của CSDL).
 3. Có thể có một **tập các luật về tính toàn vẹn**, nhằm đảm bảo CSDL luôn chính xác.

GIỚI THIỆU (cont.)

- ❖ Dựa vào mô hình kiến trúc 3 lớp, có thể xác định 3 loại mô hình dữ liệu khác nhau:
 1. **Mô hình dữ liệu ngoài** biểu diễn từng khung nhìn của người dùng trong tổ chức.
 2. **Mô hình dữ liệu khái niệm** biểu diễn khung nhìn mức logic, độc lập với hệ quản trị CSDL.
 3. **Mô hình dữ liệu trong** biểu diễn lược đồ khái niệm theo cách mà hệ quản trị CSDL hiểu được.

GIỚI THIỆU (cont.)

- ❖ Có nhiều hệ CSDL khác nhau đã được lý thuyết hóa, sử dụng, phát triển và cài đặt qua nhiều năm, được chia thành 3 loại chính: **hướng đối tượng**, **hướng bản ghi** và **vật lý**.
- ❖ Có 3 mô hình hướng bản ghi cơ bản: **mô hình dữ liệu quan hệ**, **mô hình dữ liệu mạng** và **mô hình dữ liệu phân cấp**.

=> Trong chương trình này tập trung vào mô hình dữ liệu quan hệ.

GIỚI THIỆU (cont.)

- ❖ **Mô hình dữ liệu ngữ nghĩa** cố gắng nắm bắt được ý nghĩa của CSDL => cung cấp một cách tiếp cận cho việc mô hình hóa dữ liệu mức khái niệm.
- ❖ **Mô hình dữ liệu quan hệ (Relational Data Model)** là mô hình dữ liệu ngữ nghĩa được sử dụng phổ biến nhất.
- ❖ RDM thường được dùng như một phương tiện trao đổi giữa những người thiết kế CSDL và người dùng trong các quá trình phát triển một CSDL.

QUÁ TRÌNH THIẾT KẾ MỘT CSDL

Gồm 6 bước cơ bản: (3 bước đầu liên quan đến mô hình dữ liệu ngữ nghĩa.)

1. *Phân tích yêu cầu:* Phải xác định được:

- Dữ liệu nào được lưu trữ trong CSDL,
- Ứng dụng nào sẽ được xây dựng trên CSDL này,
- Các thao tác nào được sử dụng thường xuyên và các yêu cầu về hiệu năng của hệ thống.

=> Quá trình này liên quan đến những trao đổi của các nhóm người dùng và nhóm nghiên cứu môi trường hiện tại. Tìm hiểu các ứng dụng đang có xem có cần thay thế hoặc bổ trợ cho hệ CSDL không.

QUÁ TRÌNH THIẾT KẾ MỘT CSDL (cont.)

2. **Thiết kế CSDL mức khái niệm:** Những thông tin có được từ bước phân tích yêu cầu sẽ được dùng để phát triển một mô tả mức tổng quát dữ liệu được lưu trong CSDL, cùng với các ràng buộc cần thiết trên dữ liệu này.
3. **Thiết kế CSDL mức logic:** Một hệ quản trị CSDL sẽ được chọn để cài đặt CSDL và chuyển thiết kế CSDL mức khái niệm thành một lược đồ CSDL với mô hình dữ liệu của hệ quản trị CSDL đã chọn.

QUÁ TRÌNH THIẾT KẾ MỘT CSDL (cont.)

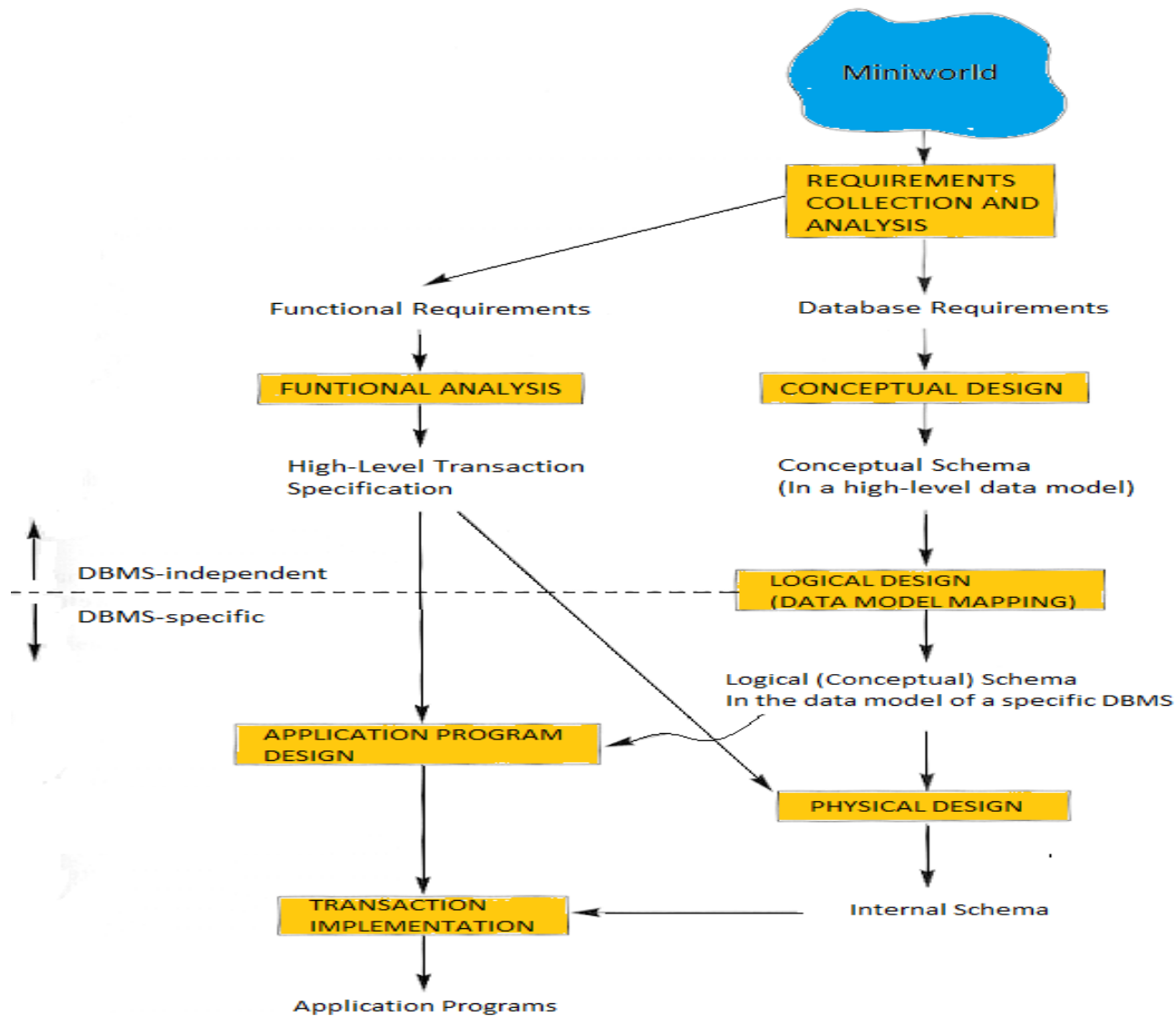
4. **Cải tiến lược đồ:** Các lược đồ được phát triển ở bước 3 sẽ được phân tích các vấn đề tiềm ẩn. Tại đây, CSDL sẽ được **chuẩn hóa**, dựa trên lý thuyết toán học.
5. **Thiết kế CSDL mức vật lý:** Khối lượng công việc tiềm ẩn và các phương pháp truy nhập được mô phỏng để xác định các điểm yếu tiềm ẩn trong CSDL mức khái niệm. Quá trình này thường là nguyên nhân tạo ra các tệp chỉ mục hoặc/và các quan hệ nhóm. Trong trường hợp đặc biệt, toàn bộ mô hình khái niệm sẽ được xây dựng lại.

QUÁ TRÌNH THIẾT KẾ MỘT CSDL (cont.)

6. **Thiết kế an toàn bảo mật:** Xác định các nhóm người dùng và phân tích vai trò của họ để định nghĩa các phương pháp truy nhập dữ liệu.

Trong quá trình phát triển, thường có bước cuối cùng (bước thứ 7), gọi là **pha điều chỉnh (tuning phase)**, trong đó CSDL sẽ được thực hiện (mặc dù nó có thể chỉ được chạy mô phỏng) và sẽ được cải tiến, chỉnh sửa để đáp ứng nhu cầu thực thi trong môi trường mong đợi.

QUÁ TRÌNH THIẾT KẾ MỘT CSDL (cont.)



Tóm tắt các bước chính trong quá trình thiết kế CSDL

LƯỢC ĐỒ THỰC THỂ LIÊN KẾT (The Entity-Relationship Diagram)

LƯỢC ĐỒ THỰC THỂ LIÊN KẾT

- ❖ Lược đồ thực thể liên kết (The Entity-Relationship Diagram) gồm 3 khái niệm cơ bản: **tập thực thể**, **tập quan hệ** và **thuộc tính**.
- ❖ **Thực thể** là một đối tượng trong thế giới thực và có thể phân biệt được với các đối tượng khác. Thực thể có thể cụ thể (một người, một quyển sách, ...) hoặc cũng có thể trừu tượng (một khoản vay ngân hàng, một khái niệm, ...).
- ❖ Thực thể được biểu diễn bởi một tập các **thuộc tính** (là các thuộc tính mô tả hoặc các đặc tính của thực thể).

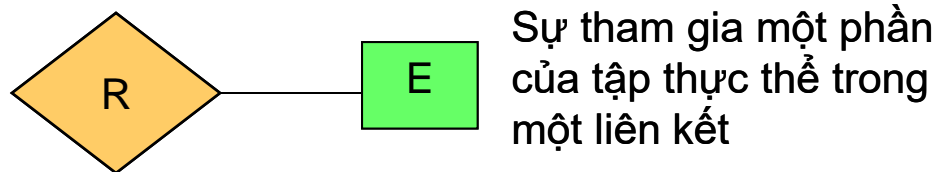
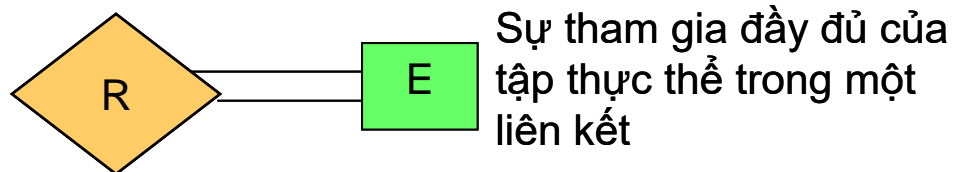
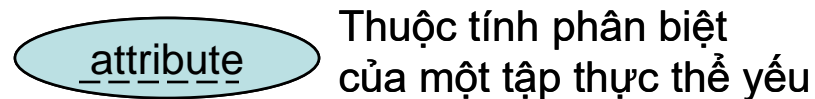
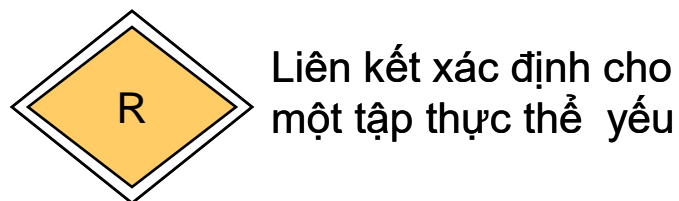
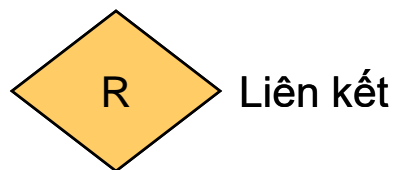
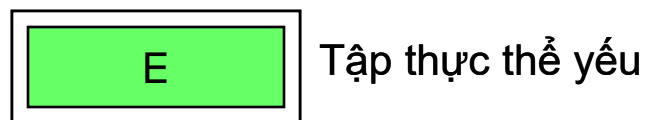
LƯỢC ĐỒ THỰC THỂ LIÊN KẾT (cont.)

- ❖ *Tập thực thể* là một nhóm các thực thể có cùng thuộc tính. Ví dụ: tập tất cả khách hàng của ngân hàng có thể được định nghĩa là tập khách hàng.
- ❖ Các tập thực thể không nhất thiết phải tách rời nhau. Ví dụ: có thể định nghĩa tập tất cả những người làm việc ở ngân hàng là *nhân viên*, tập tất cả những khách hàng của ngân hàng là *khách hàng*. Một thực thể người nào đó có thể là một *nhân viên* hoặc một *khách hàng* hoặc cả hai, hoặc không phải cả hai.

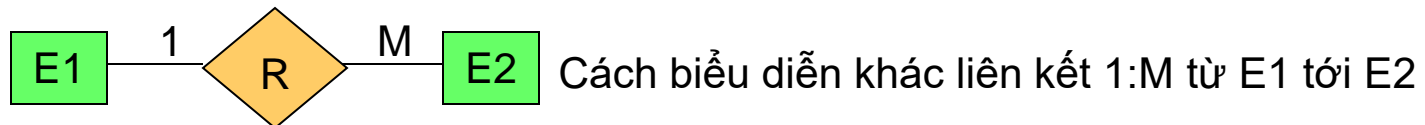
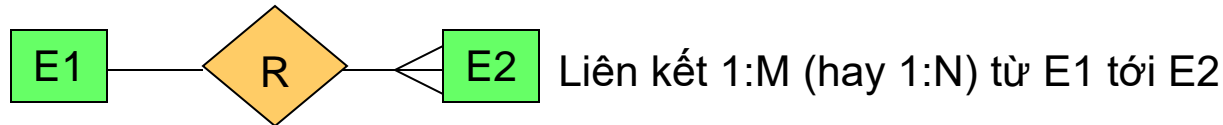
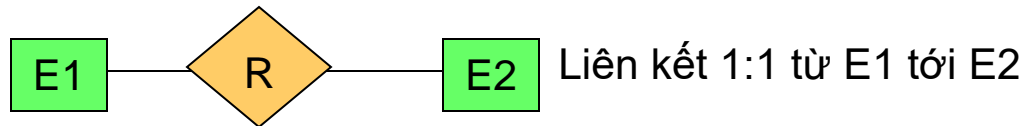
LƯỢC ĐỒ THỰC THỂ LIÊN KẾT (cont.)

- ❖ Mỗi thuộc tính có một tập giá trị cho phép, được gọi là **miền** (hay tập giá trị) của thuộc tính đó.
 - Một thuộc tính của một tập thực thể là một hàm ánh xạ từ một tập thực thể vào một miền giá trị.
 - Một tập thực thể có thể có nhiều thuộc tính.
=> Mỗi thực thể trong tập có thể được mô tả bởi một tập các cặp *<thuộc tính, giá trị>*, ứng với từng thuộc tính trong tập thực thể.
- ❖ Một CSDL bao gồm một tập các thực thể.

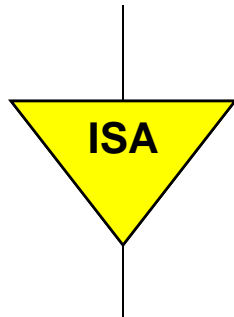
KÝ HIỆU TRONG LƯỢC ĐỒ E-R



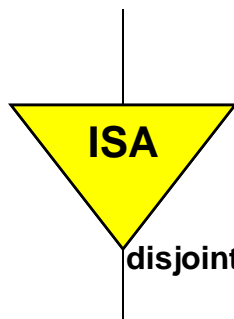
KÝ HIỆU TRONG LƯỢC ĐỒ E-R (cont.)



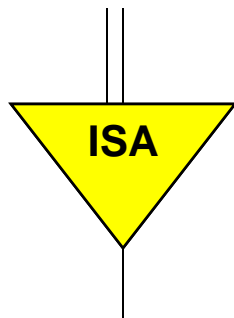
KÝ HIỆU TRONG LƯỚI ĐỒ E-R (cont.)



Là-một (cụ thể hóa hoặc tổng quát hóa) (tham gia một phần)

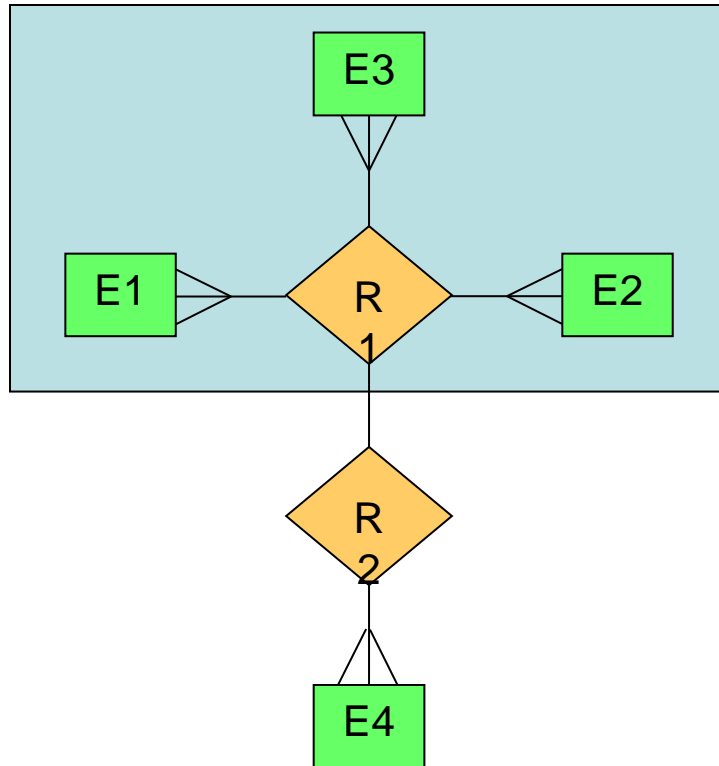


Là một phân tách (cụ thể hóa hoặc tổng quát hóa)

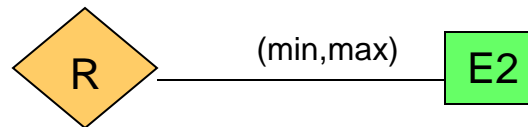


Tổng quát hóa đầy đủ

KÝ HIỆU TRONG LƯỢC ĐỒ E-R (cont.)

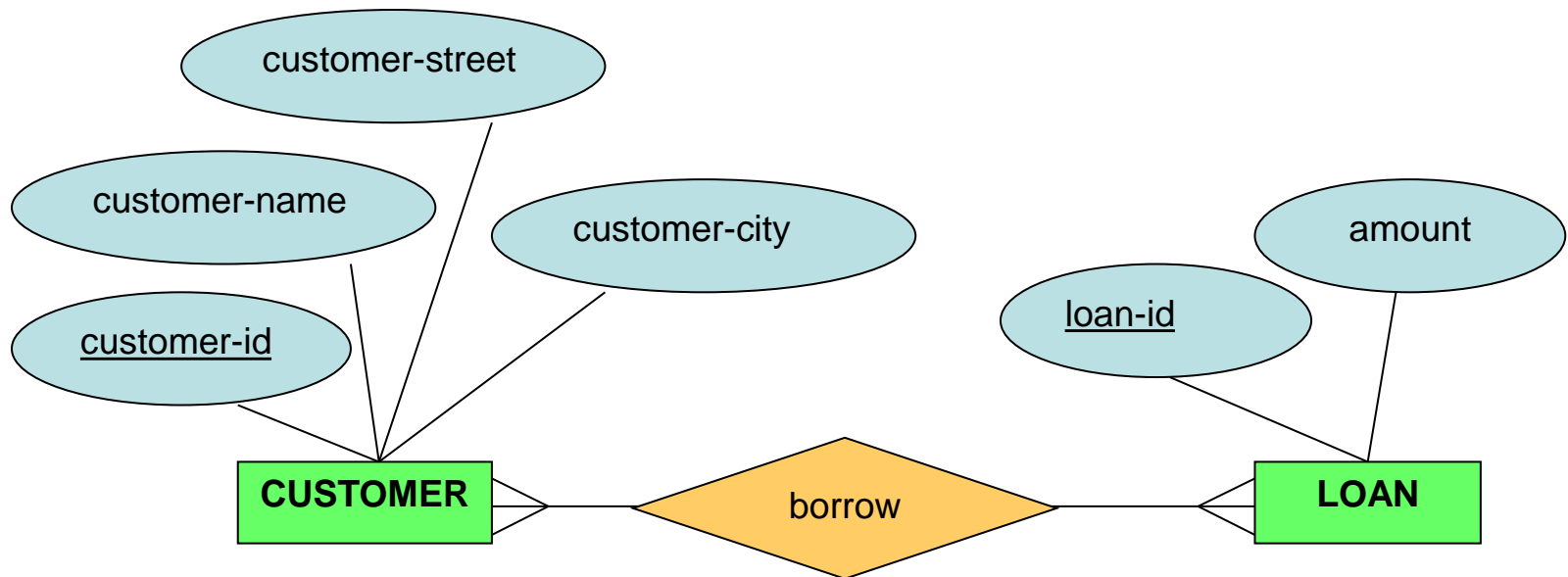


Kết hợp: hộp chữ nhật bao quanh liên kết được coi như một thực thể

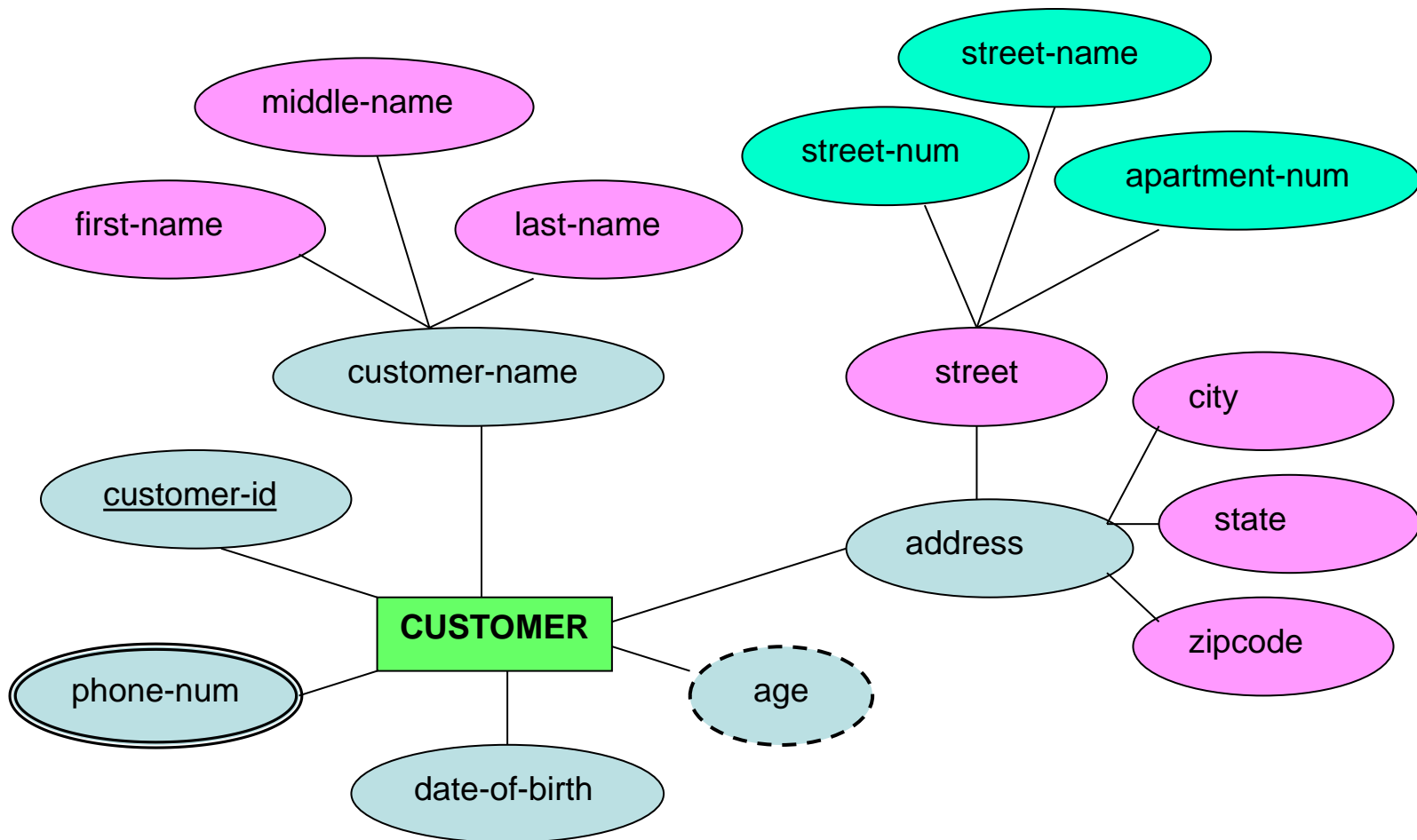


Ràng buộc có cấu trúc: (min,max) của số lượng thực thể tham gia liên kết

VÍ DỤ VỀ LƯỢC ĐỒ E-R (ERD)



VÍ DỤ KHÁC VỀ LƯỢC ĐỒ E-R



CÁC THUỘC TÍNH TRONG LƯỚI ĐỒ E-R

Thuộc tính có thể được phân chia thành các loại sau:

- ❖ **Thuộc tính đơn** hoặc **thuộc tính kép**: Thuộc tính đơn không bao gồm các thành phần cấu thành, trong khi thuộc tính kép bao gồm các thành phần con cấu thành.

Ví dụ: thuộc tính **tên**: Nếu *tên* biểu diễn một thuộc tính đơn thì có thể coi bộ ba cấu thành **tên** là **họ**, **tên đệm** và **tên gọi** là một thuộc tính nguyên tố, không phân chia được nữa. Còn nếu coi **tên** là một thuộc tính kép thì có thể lựa chọn thao tác với thuộc tính này là một tên đầy đủ hoặc có thể thao tác với từng thành phần cấu thành tên.

CÁC THUỘC TÍNH TRONG LƯỢC ĐỒ E-R (cont.)

- ❖ **Thuộc tính đơn trị** hoặc **thuộc tính đa trị**: Thuộc tính đơn trị có nhiều nhất một giá trị tại một thời điểm cụ thể. Thuộc tính đa trị có thể có nhiều giá trị khác nhau tại một thời điểm.

Ví dụ: Tại một trường học sinh viên được đăng ký học theo tín chỉ. Tại một kỳ học nào đó, số tín chỉ một sinh viên đăng ký là đơn trị, ví dụ là 7 (tín chỉ) => số tín chỉ không thể nhận giá trị đa trị.

Thuộc tính số điện thoại của sinh viên có thể chứa nhiều giá trị cùng lúc do tại một thời điểm, một sinh viên có thể có một vài số điện thoại khác nhau. => thuộc tính số điện thoại là đa trị.

CÁC THUỘC TÍNH TRONG LƯỢC ĐỒ E-R (cont.)

- ❖ **Thuộc tính dẫn xuất**: là thuộc tính mà giá trị của nó được dẫn xuất (hoặc được tính toán) từ những giá trị của các thuộc tính hoặc các thực thể có liên quan.

Ví dụ: Giả sử thực thể **KHÁCH HÀNG** của một ngân hàng có một thuộc tính tên là ***loans-held***, chứa số lượng các khoản vay của một khách hàng tại ngân hàng. Giá trị của thuộc tính này có thể được tính bằng cách đếm số lượng thực thể các khoản vay liên quan tới khách hàng.

- ❖ **Thuộc tính rỗng (Null)**: thuộc tính nhận giá trị rỗng khi một thực thể không có giá trị cho nó.

CÁC LIÊN KẾT TRONG LƯỢC ĐỒ E-R

- ❖ Một **liên kết** (hay một **quan hệ**) là một mối liên hệ giữa một vài thực thể.

Ví dụ, có thể định nghĩa một mối liên kết thể hiện sinh viên của một lớp học. Mối quan hệ này xác nhận rằng sinh viên đã đăng ký học lớp đó.

Một cách hình thức, một *tập quan hệ* là một tập các liên hệ cùng loại. Nó là một quan hệ toán học của n tập thực thể (có thể giao nhau, $n \geq 2$).

Nếu E_1, E_2, \dots, E_n là các tập thực thể, thì một tập quan hệ R là một tập con của:

$$\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$$

Với (e_1, e_2, \dots, e_n) là một quan hệ.

CÁC LIÊN KẾT TRONG LƯỢC ĐỒ E-R (cont.)

- ❖ Mỗi liên hệ giữa các tập thực thể được gọi là **sự tham gia**; nghĩa là các tập thực thể E_1, E_2, \dots, E_n tham gia vào quan hệ R .
- ❖ Một **thể hiện quan hệ** trong lược đồ E-R biểu diễn mỗi liên hệ giữa các thực thể xác định trong thế giới thực đang được mô hình hóa.
- ❖ Một mối quan hệ cũng có thể có các thuộc tính được gọi là các **thuộc tính mô tả**.

Ví dụ: Xét ngữ cảnh ngân hàng, giả sử có tập các mối quan hệ **depositor** và các tập thực thể **CUSTOMER** và **ACCOUNT** => có thể có thuộc tính mô tả **access-date** cho mối quan hệ **depositor** để mô tả ngày gần nhất mà khách hàng truy nhập vào tài khoản của họ.

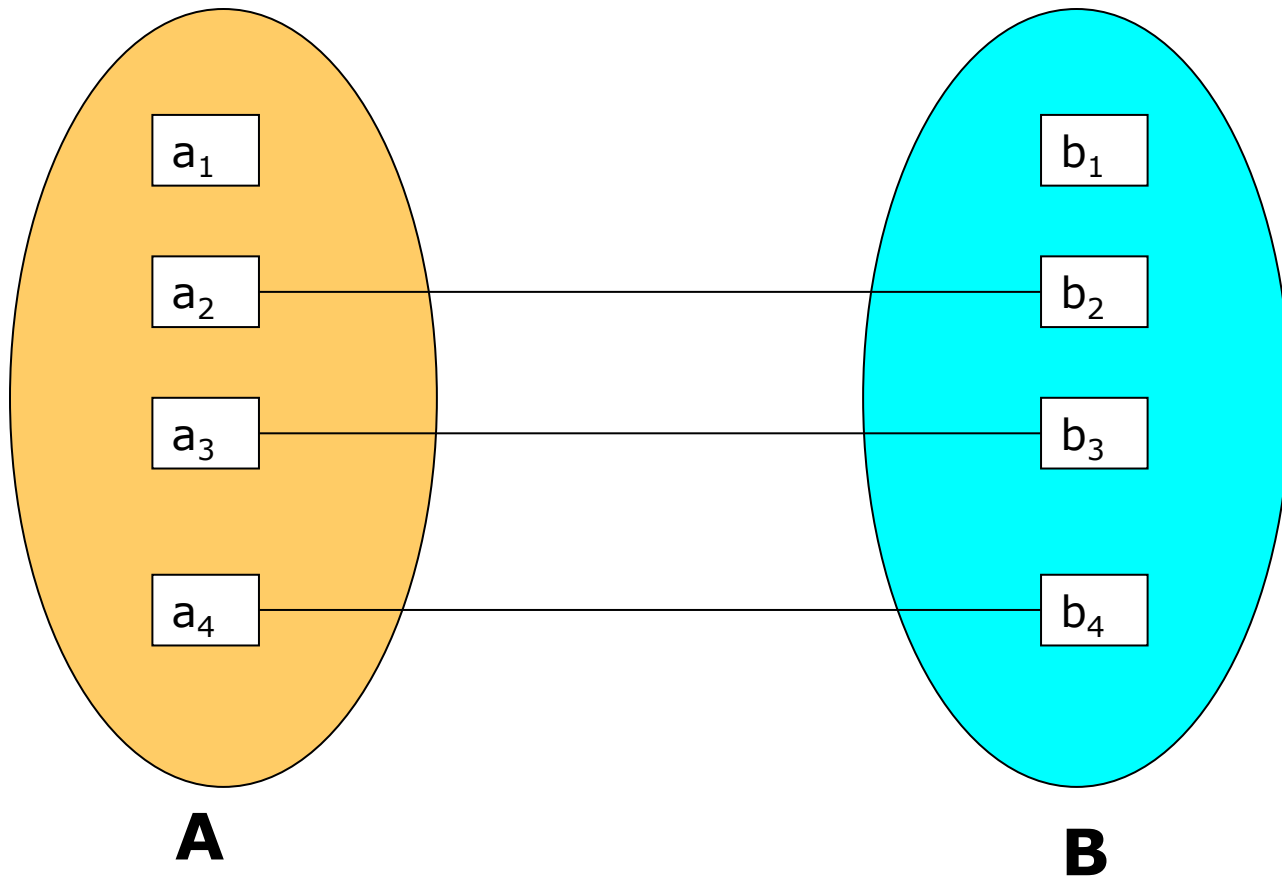
CÁC RÀNG BUỘC TRONG LỰỢC ĐỒ E-R

- ❖ Các giá trị được lưu trữ trong CSDL thường có những ràng buộc để đảm bảo rằng chúng mô hình hóa một cách chính xác toàn bộ thể giới thực của tổ chức đang được thể hiện trong CSDL.
- ❖ Lược đồ E-R có khả năng mô hình hóa các loại ràng buộc này.
- ❖ Tập trung vào hai loại ràng buộc quan trọng: **ánh xạ lực lượng liên kết** và **các ràng buộc tham gia**.

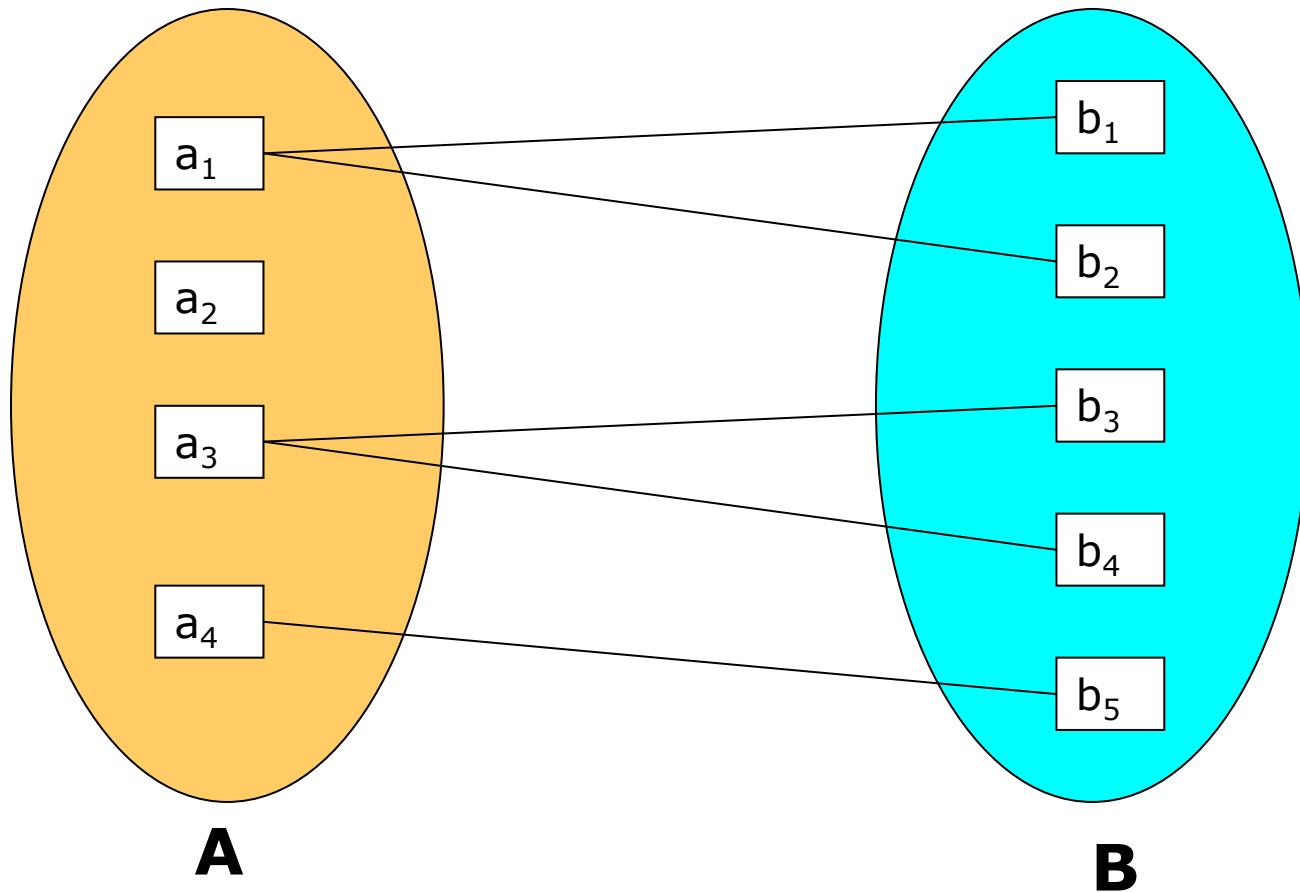
ÁNH XẠ LỰC LƯỢNG LIÊN KẾT

- ❖ **Ánh xạ lực lượng liên kết** thể hiện số lượng các thực thể mà một thực thể khác có thể liên hệ thông qua một tập quan hệ.
- ❖ Ràng buộc ánh xạ lực lượng liên kết có ích nhất khi mô tả các quan hệ hai ngôi.
- ❖ Với một tập quan hệ hai ngôi R giữa tập thực thể A và B , ánh xạ lực lượng liên kết gồm các loại sau:
 - $(1:1)$ một tới một từ A đến B
 - $(1:M)$ một tới nhiều từ A đến B
 - $(M:1)$ nhiều tới một từ A đến B
 - $(M:M)$ nhiều tới nhiều từ A đến B

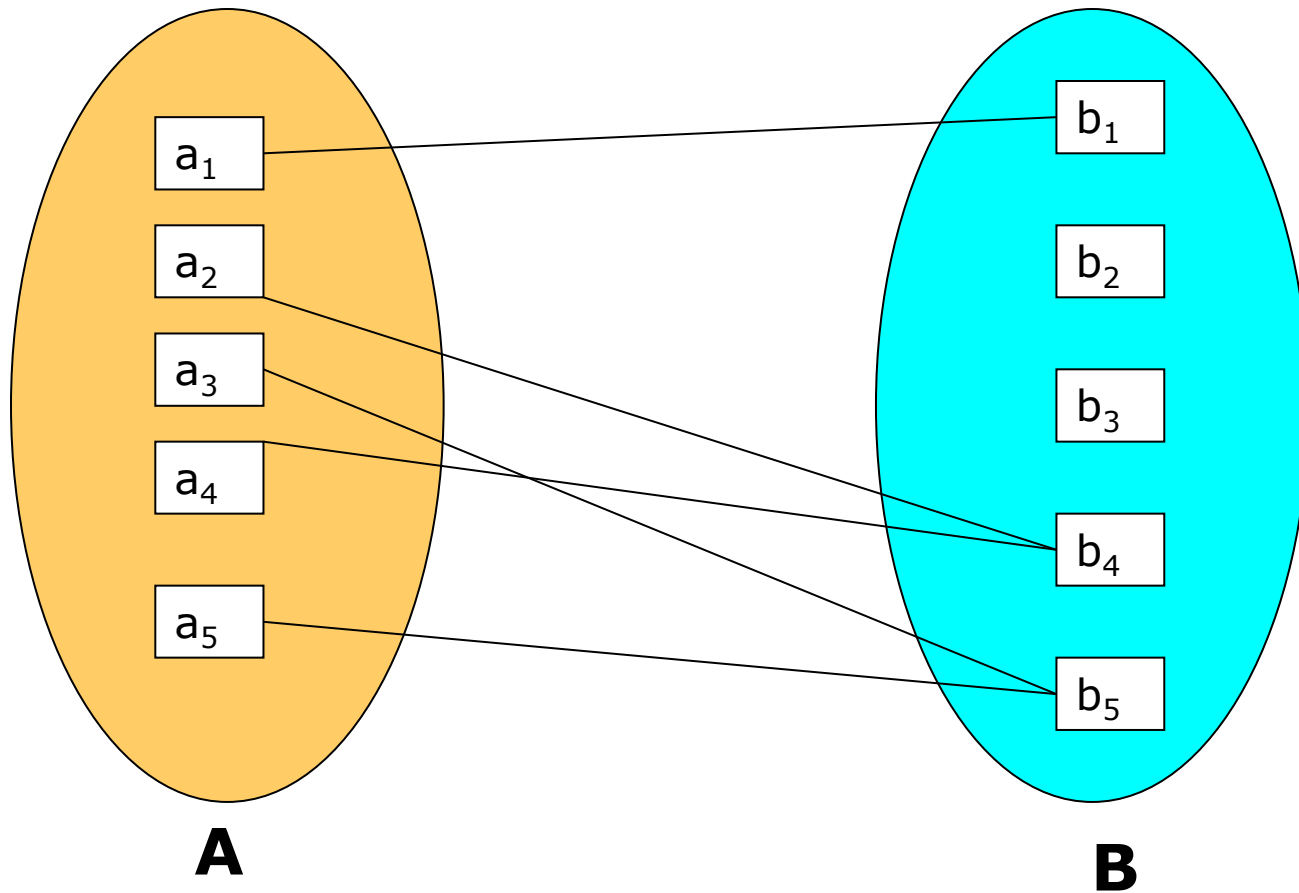
ẢNH XẠ LỰC LƯỢNG LIÊN KẾT: 1:1 TỪ A ĐẾN B



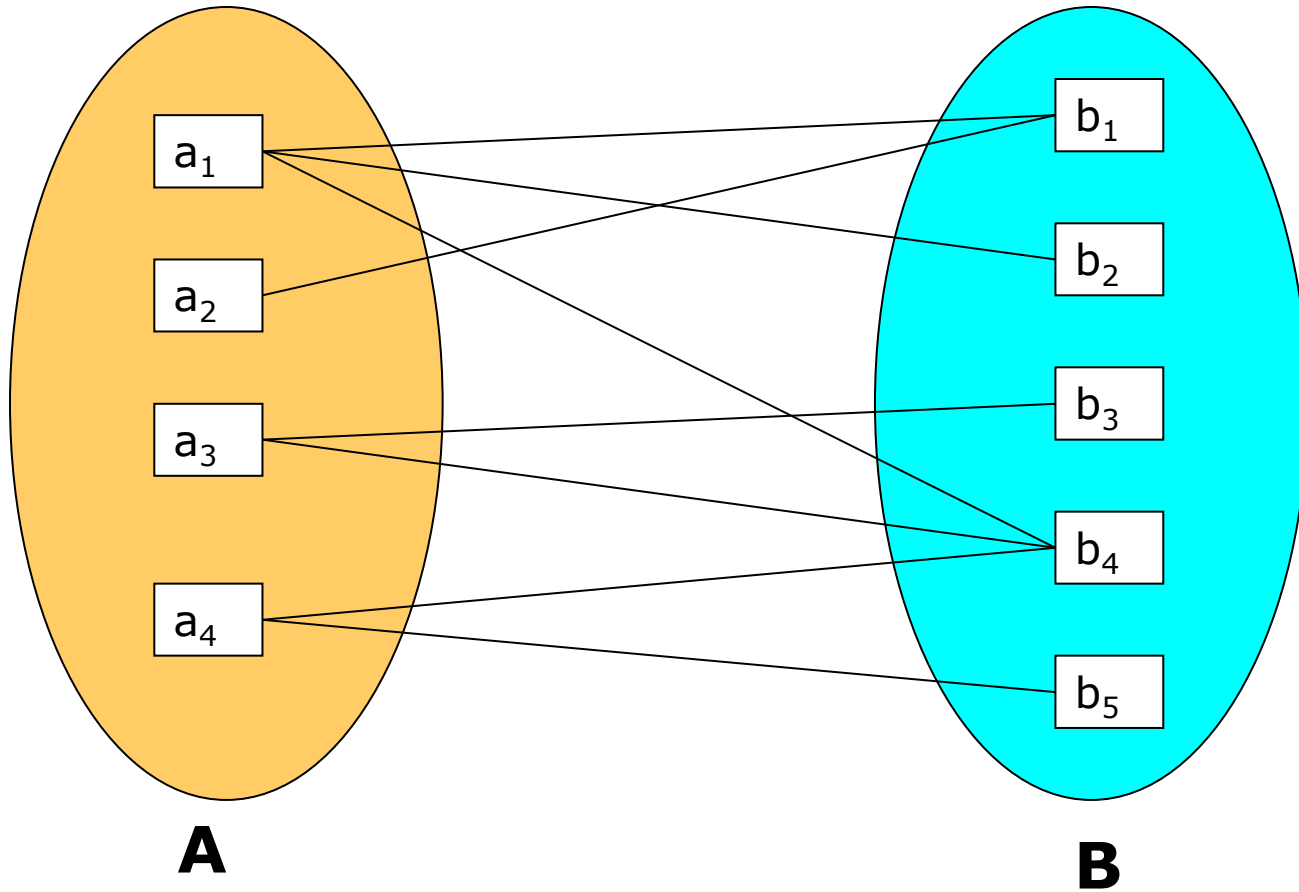
ÁNH XẠ LỰC LƯỢNG LIÊN KẾT: 1:M TỪ A ĐẾN B



ÁNH XẠ LỰC LƯỢNG LIÊN KẾT: M:1 TỪ A ĐẾN B



ÁNH XẠ LỰC LƯỢNG LIÊN KẾT: M:M TỪ A ĐẾN B



CÁC RÀNG BUỘC THAM GIA TRONG LỰỢC ĐỒ E-R

- ❖ Sự tham gia của một tập thực thể E trong một tập các quan hệ R được gọi là **đầy đủ (total)** nếu mọi thực thể của E tham gia vào ít nhất một quan hệ trong R .
- ❖ Nếu chỉ có một vài thực thể của E tham gia vào một quan hệ trong R , thì sự tham gia của tập thực thể E trong tập các quan hệ R được gọi là **một phần (partial)**.

CÁC RÀNG BUỘC THAM GIA TRONG LỰỚI ĐỒ E-R (Cont.)

❖ Ví dụ, xem xét hệ thống ngân hàng.

- Mỗi thực thể **LOAN** (khoản vay) có liên hệ tới ít nhất một **CUSTOMER** (khách hàng) thông qua mối quan hệ *borrower* (vay mượn) => sự tham gia của thực thể LOAN trong tập quan hệ *borrower* là **đầy đủ**.
- Ngược lại, có những khách hàng không liên quan tới các khoản vay mượn ngân hàng. Như vậy, có thể nói rằng chỉ một số các thực thể **CUSTOMER** có liên hệ tới một thực thể **LOAN** thông qua quan hệ *borrower* => sự tham gia của thực thể **CUSTOMER** trong tập quan hệ *borrower* là **một phần**.

KHÓA CỦA MỘT TẬP THỰC THỂ

- ❖ Cần có một cơ chế để phân biệt các thực thể trong một tập thực thể.
- ❖ Theo khía cạnh CSDL, sự khác nhau giữa các thực thể phải được thể hiện thông qua các thuộc tính.

=> Các giá trị của thuộc tính của một thực thể phải được xác định sao cho chúng có thể **xác định duy nhất** thực thể đó. Nghĩa là, không có hai thực thể nào trong một tập thực thể được phép có giá trị trùng nhau trên tất cả các thuộc tính.
- ❖ **Khóa (key)** cho phép xác định một tập các thuộc tính đủ để phân biệt các thực thể với nhau. Khóa cũng giúp cho việc xác định duy nhất các mối quan hệ, và vì vậy, phân biệt các mối quan hệ với nhau.

KHÓA CHÍNH, SIÊU KHÓA VÀ KHÓA DỰ BỊ

- ❖ **Siêu khóa (superkey)** là một tập gồm một hoặc nhiều thuộc tính được lựa chọn cho phép xác định duy nhất một thực thể trong tập thực thể.

Giả sử có một tập thực thể sinh viên trong một lớp học với lược đồ sau:

STUDENTS(SS#, name, address, age, major, minor, gpa, spring-sch)

- ❖ Siêu khóa cho tập thực thể STUDENTS có thể là:
(SS#, name, major, minor) hoặc (SS#, name)

=> *Khái niệm siêu khóa là một định nghĩa không đầy đủ của khóa vì siêu khóa còn chứa nhiều thuộc tính dư thừa.*

KHÓA CHÍNH, SIÊU KHÓA VÀ KHÓA DỰ BỊ (Cont.)

- ❖ Nếu tập K là một siêu khóa của tập thực thể E thì mọi tập cha của K cũng là siêu khóa.
- ❖ **Khóa dự bị (candidate key)** là những siêu khóa mà không có tập con nào của nó là siêu khóa.

=> Với mỗi một tập thực thể E cho trước, tồn tại một hoặc nhiều khóa dự bị.
- ❖ Người thiết kế CSDL chỉ chọn một khóa dự bị làm **khóa chính**, hay còn gọi là **khóa** của tập thực thể.

KHÓA CHÍNH, SIÊU KHÓA VÀ KHÓA DỰ BỊ (Cont.)

- ❖ Hai thực thể khác nhau sẽ không có cùng giá trị trên tất cả các thuộc tính khóa (khóa chính, khóa dự bị hay siêu khóa) tại cùng một thời điểm.

=> **ràng buộc khóa**

- ❖ Người thiết kế CSDL phải cẩn thận khi lựa chọn tập các thuộc tính cấu thành khóa của một tập thực thể để đảm bảo:
 1. Chấn chấn rằng tập các thuộc tính xác định duy nhất thực thể.
 2. Tập thuộc tính khóa sẽ không bao giờ hoặc rất hiếm khi bị thay đổi.

TẬP CÁC QUAN HỆ

- ❖ Khóa chính của một tập thực thể cho phép phân biệt các thực thể khác nhau trong tập => phải có một cơ chế tương tự cho phép phân biệt các mối quan hệ trong một tập các quan hệ.
- ❖ Cho R là một tập quan hệ liên quan tới các tập thực thể E_1, E_2, \dots, E_n . Gọi K_i là tập các thuộc tính cấu thành khóa chính của tập thực thể E_i .

Giả thiết rằng:

1. Tên của các thuộc tính trong tất cả các khóa chính là duy nhất.
2. Mỗi tập thực thể sẽ chỉ tham dự một lần duy nhất trong quan hệ.

TẬP CÁC QUAN HỆ (Cont.)

- ❖ Việc cấu thành khóa chính cho một tập các quan hệ R phụ thuộc vào tập các thuộc tính liên quan theo các cách sau:
 1. Nếu tập quan hệ R không có thuộc tính nào liên quan thì tập các thuộc tính: $K_1 \cup K_2 \cup \dots \cup K_n$ mô tả một quan hệ riêng biệt trong tập R .
 2. Nếu tập quan hệ R có các thuộc tính liên quan a_1, a_2, \dots, a_m thì tập các thuộc tính $K_1 \cup K_2 \cup \dots \cup K_n \cup \{a_1, a_2, \dots, a_m\}$ sẽ mô tả một quan hệ riêng trong tập R .
- => Trong cả hai trường hợp, tập các thuộc tính $K_1 \cup K_2 \cup \dots \cup K_n$ đều hình thành một siêu khóa cho tập các quan hệ.