

THIẾT KẾ HỆ CSDL QUAN HỆ

Nguyễn Đình Hóa

dinhhoa@gmail.com

094-280-7711

B1. Phân tích yêu cầu

Cần xác định được một cách chi tiết về:

- Dữ liệu nào được lưu trữ trong CSDL, thông tin nào không cần lưu trữ.
- Ứng dụng nào sẽ được xây dựng trên CSDL này, quy mô ứng dụng ra sao.
- Các thao tác nào được sử dụng thường xuyên và các yêu cầu về hiệu năng của hệ thống.

B2. Thiết kế CSDL mức khái niệm

- **Lược đồ thực thể liên kết** (The Entity-Relationship Diagram) gồm 3 khái niệm cơ bản: **tập thực thể**, **tập liên kết** và **tập 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, ...).
- Mỗi thực thể được biểu diễn bởi một tập các **thuộc tính** (là các thuộc tính mô tả hoặc các đặc tính của thực thể).

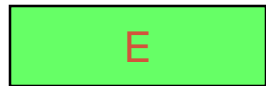
Lược đồ thực thể - liên kết

- **Tập thực thể** là một nhóm các thực thể có cùng một tập 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. Một đối tượng có thể tham gia một hoặc nhiều tập thực thể. **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 đối tượng người nào đó có thể là một *nhân viên* hoặc một *khách hàng*, hoặc cả hai, hoặc không phải cả hai.
- Một CSDL gồm nhiều tập thực thể.

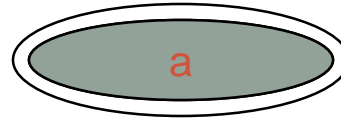
Lược đồ thực thể liên kết

- 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.

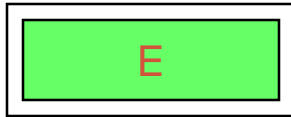
Các ký hiệu trong ERD



Tập thực thể



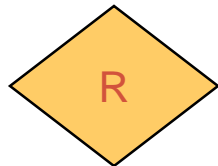
Thuộc tính đa trị



Tập thực thể yếu



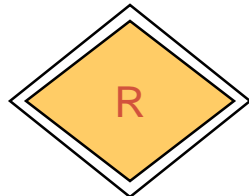
Thuộc tính dẫn xuất



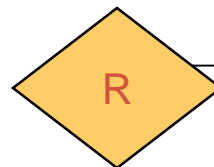
Liên kết



Thuộc tính phân biệt của một tập thực thể yếu



Liên kết xác định cho một tập thực thể yếu



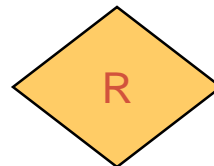
Sự tham gia đầy đủ của tập thực thể trong một liên kết



Khóa chính



Thuộc tính

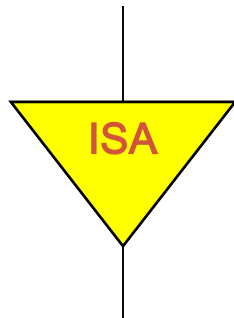


Sự tham gia một phần của tập thực thể trong một liên kết

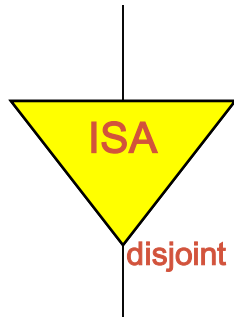
Các ký hiệu trong ERD



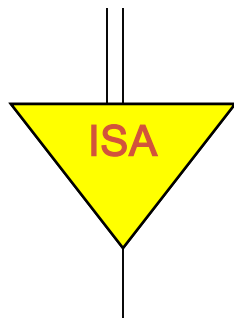
Các ký hiệu trong ERD



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

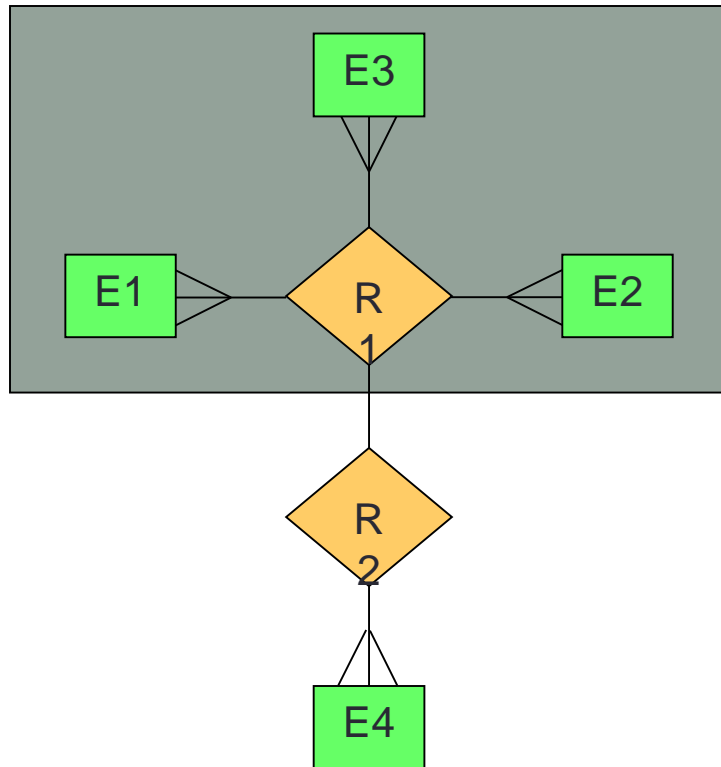


Là một phân tách (cụ thể hóa hoặc tổng quát hóa)

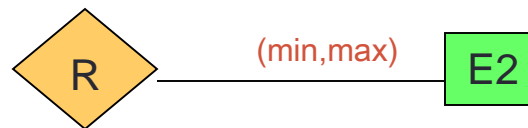


Tổng quát hóa đầy đủ

Các ký hiệu trong ERD

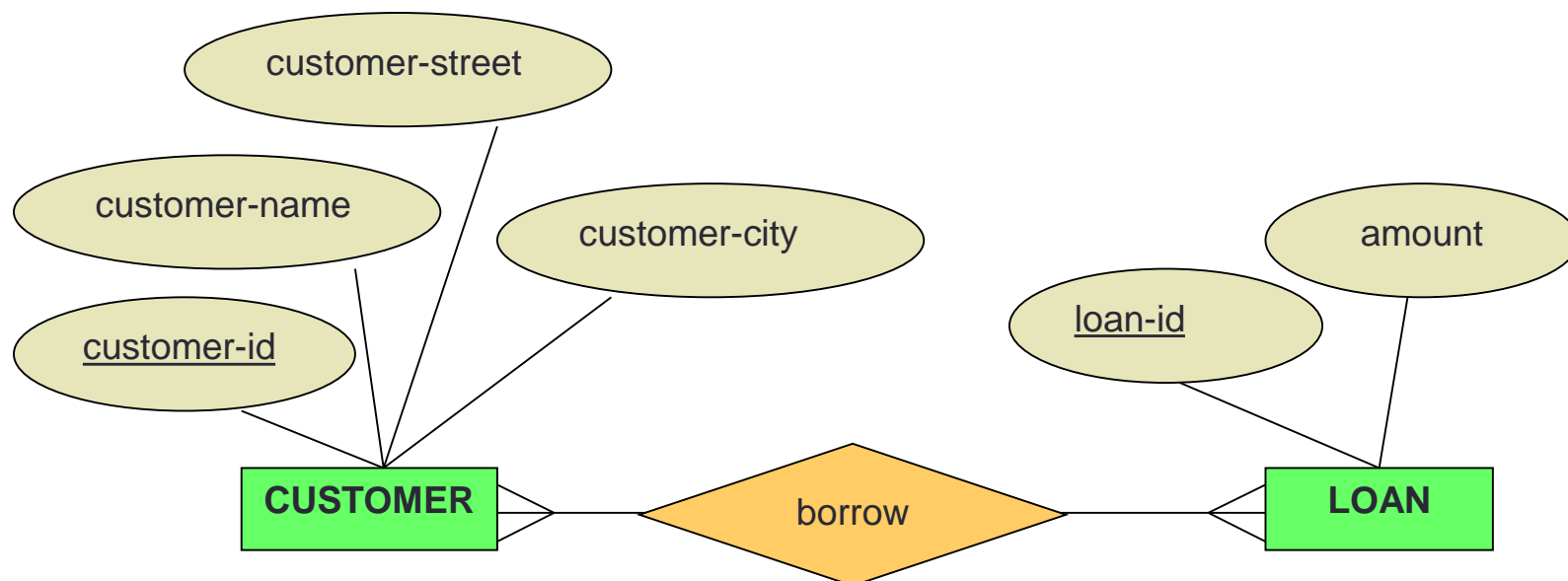


Kết hợp: hộp chữ nhật bao quanh liên kết được coi như một thực thể

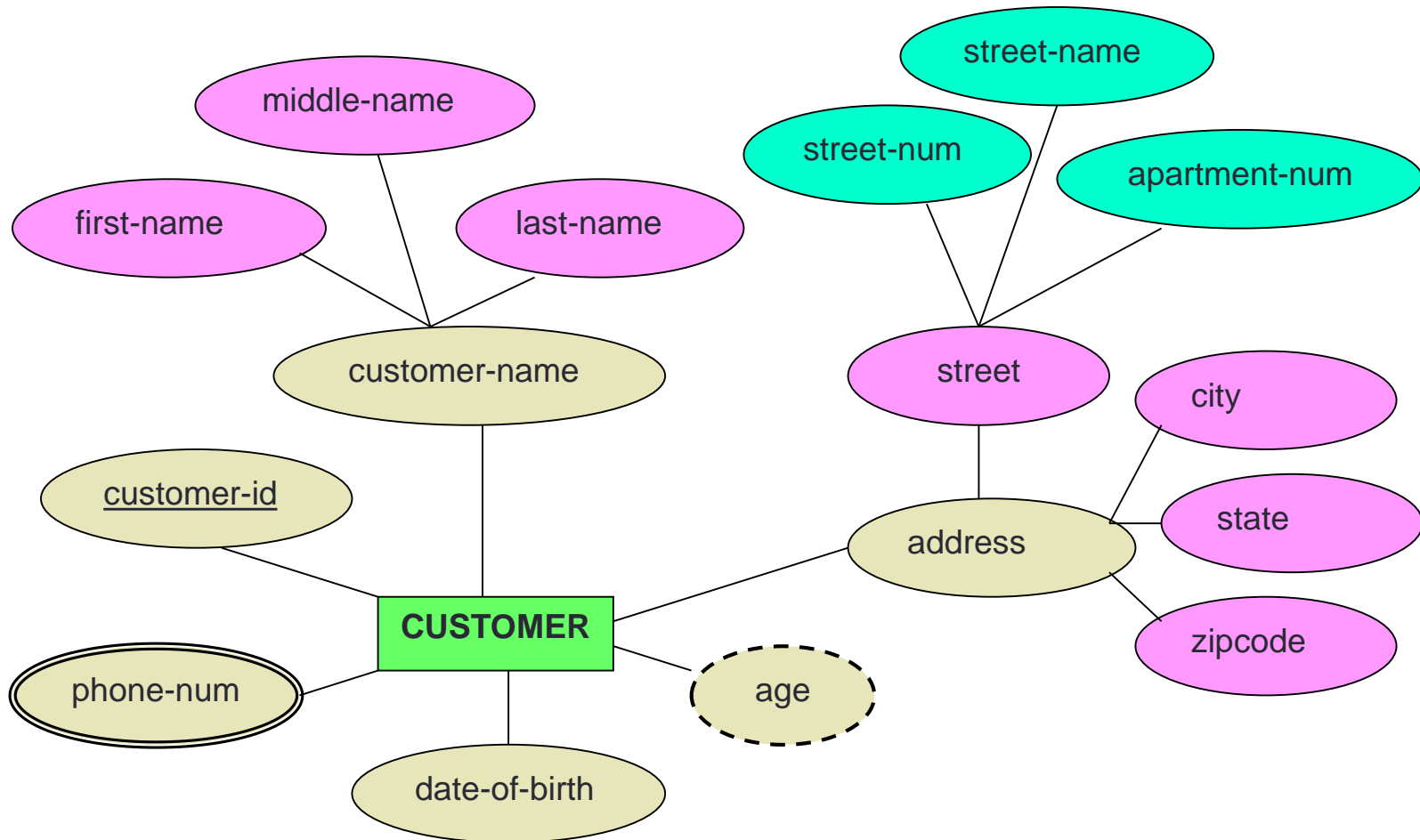


Ràng buộc có cấu trúc: (min,max) của số lượng thực thể tham gia liên kết

Ví dụ về ERD



Ví dụ về ERD



Các loại thuộc tính

- **Thuộc tính đơn** hoặc **thuộc tính kép/ghép**: Thuộc tính đơn không bao gồm các thành phần cấu thành, trong khi thuộc tính kép bao gồm các thành phần con cấu thành.
- **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.
- **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.
- **Thuộc tính rỗng (Null)**: thuộc tính không mang giá trị khi một thực thể không có giá trị cho nó.

Các liên kết trong ERD

- Một **liên kết** (hay một **quan hệ**) là một mối liên hệ giữa hai hay nhiều 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.

- Mối liên hệ giữa các tập thực thể được gọi là **sự tham gia** của các tập thực thể đó vào liên kết.
- Một **thể hiện liên kết** trong lược đồ E-R biểu diễn mối liên hệ giữa các thực thể xác định trong thế giới thực đang được mô hình hóa.
- Một mối liên kết cũng có thể có các thuộc tính được gọi là các **thuộc tính mô tả**.

Các ràng buộc trong ERD

- Hai loại ràng buộc quan trọng: **ánh xạ lực lượng liên kết** và **các ràng buộc tham gia**.
- **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 liên kết.
- Ràng buộc ánh xạ lực lượng liên kết có ích nhất khi mô tả các liên kết hai ngôi.
- Với một tập liên kết 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

Các ràng buộc trong ERD

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

Khoá 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.

Các khái niệm về Khoá

- **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.

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

- Nếu tập thuộc tính 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.

Khoá của các liên kết

- Cho R là một tập liên kết của 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 liên kết.

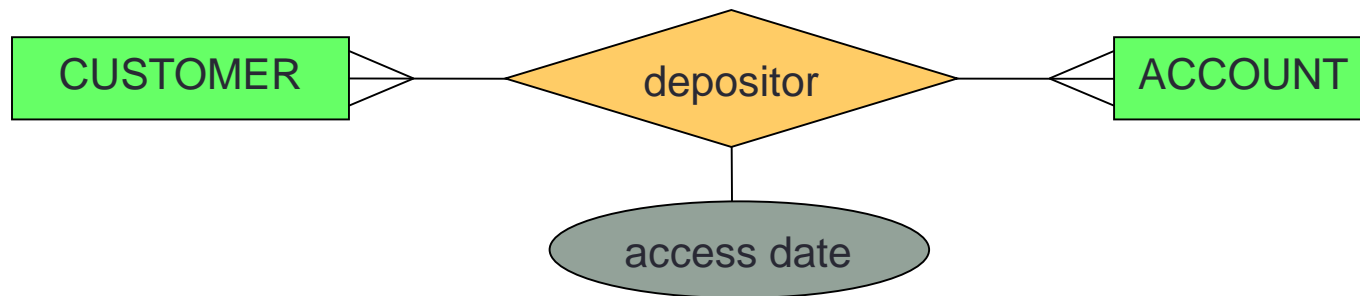
Khoá của các liên kết

- 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 liên kết.

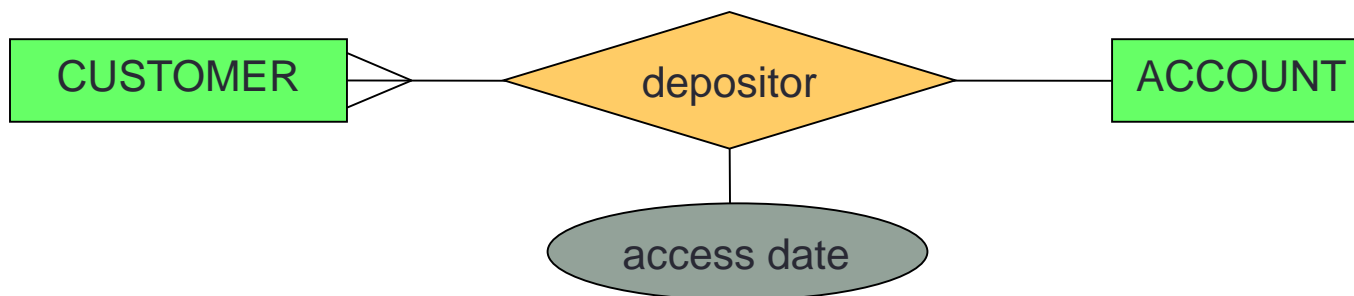
Khoá của các liên kết phụ thuộc vào ảnh xạ lực lượng tham gia liên kết

- Liên kết N-N: Khóa chính của liên kết này sẽ là tổ hợp của các khóa chính của hai tập thực thể **CUSTOMER** và **ACCOUNT**

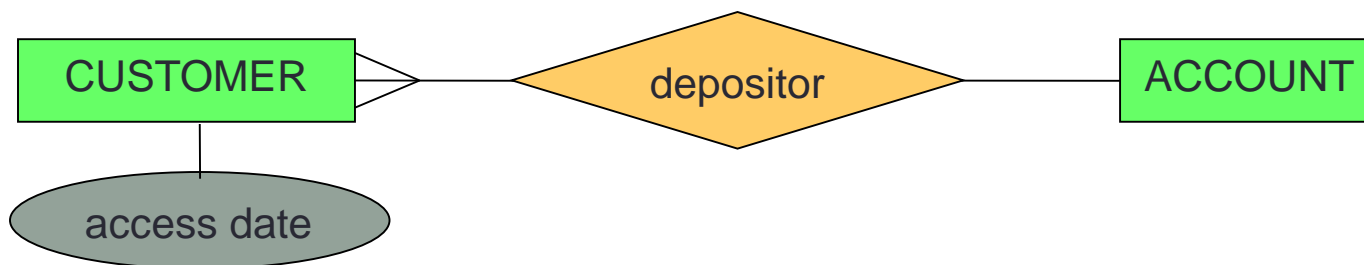


Khoá của các liên kết phụ thuộc vào ảnh xạ lực lượng tham gia liên kết

- Liên kết N-1: Khóa chính của liên kết **depositor** sẽ chỉ bao gồm khóa chính của tập thực thể **CUSTOMER**.

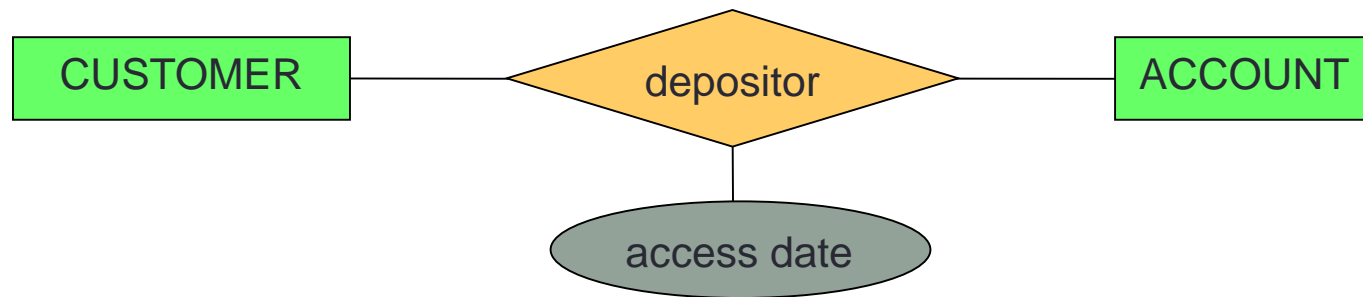


- Tương đương với:



Khoá của các liên kết phụ thuộc vào ảnh xạ lực lượng tham gia liên kết

- Liên kết 1-1: Khóa chính của liên kết **depositor** có thể là khóa chính của một trong hai tập thực thể CUSTOMER và ACCOUNT.



=> Thuộc tính **access-date** có thể gắn vào hoặc thực thể **CUSTOMER** hoặc tập thực thể **ACCOUNT** mà không làm tổn thất thông tin.

Một số vấn đề cần chú ý

- Sự phân biệt 2 vai trò thực thể và thuộc tính phụ thuộc vào cấu trúc ngữ cảnh trong thế giới thực cần mô hình hóa dữ liệu và dựa vào ngữ nghĩa liên quan đến các thuộc tính trong bài toán cụ thể.
- Việc xác định đâu là tập thực thể, đâu là tập các liên kết cũng hoàn toàn phụ thuộc vào từng tình huống và mô tả dữ liệu cụ thể.

Bài tập

Hãy xây dựng lược đồ ER để biểu diễn CSDL quản lý ngân hàng với các thông tin về khách hàng và tài khoản.

- Mỗi khách hàng có thể giao dịch trên nhiều tài khoản, mỗi tài khoản có thể có giao dịch từ nhiều khách hàng.
- Mỗi khách hàng có thể thực hiện nhiều giao dịch trên một tài khoản, thông tin về mỗi lần giao dịch đó cần được lưu trữ lại.

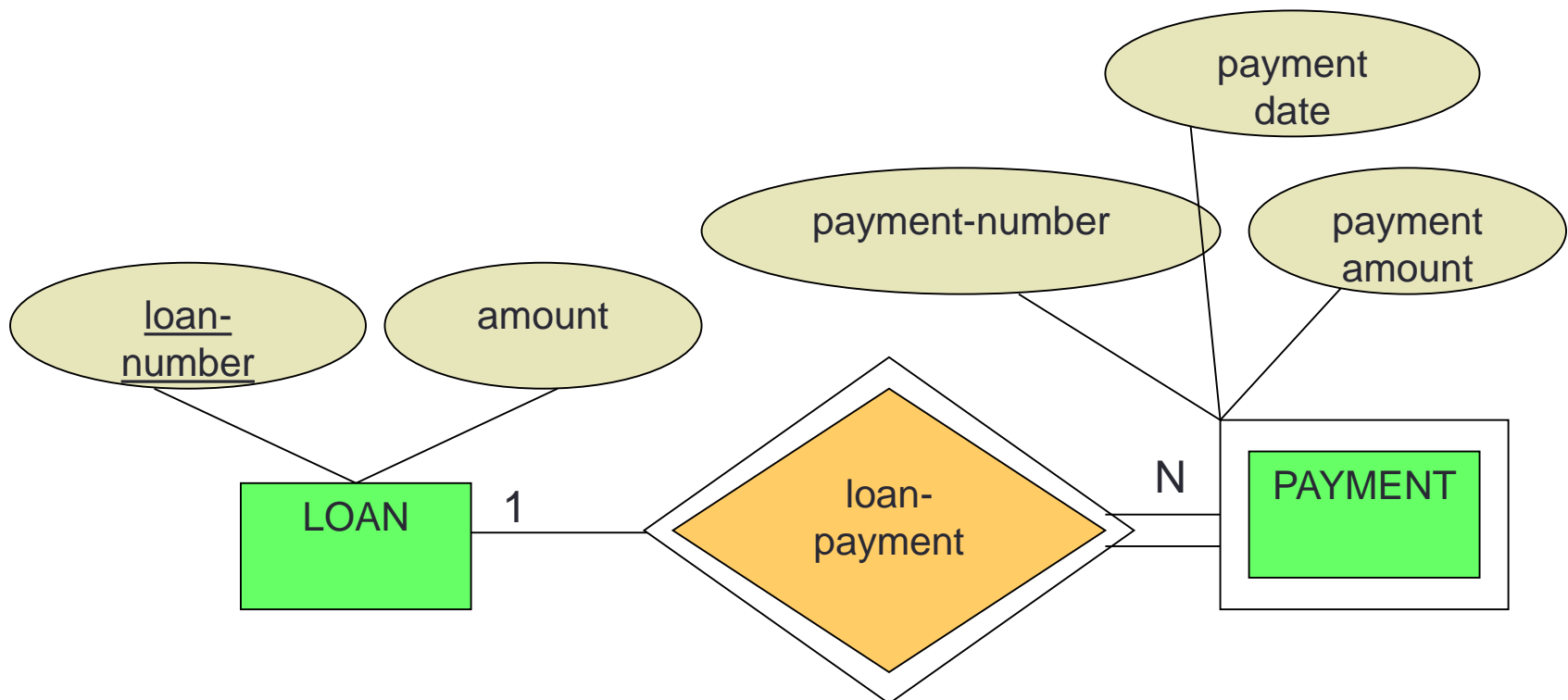
Sinh viên tự bổ sung các thuộc tính cho các tập thực thể và các liên kết để đảm bảo mỗi tập thực thể có ít nhất 3 thuộc tính.

Tập thực thể yếu

- Tập thực thể không đủ các thuộc tính để hình thành một khóa chính gọi là **tập thực thể yếu**. Tập thực thể có khóa chính được gọi là **tập thực thể mạnh**.
- Để một tập thực thể yếu có ý nghĩa, nó phải liên hệ với một tập thực thể khác, được gọi là **tập thực thể xác định** hay **tập thực thể sở hữu**.
- Mỗi quan hệ giữa tập thực thể yếu với tập thực thể xác định được gọi là **mối quan hệ xác định**.
- Mỗi quan hệ xác định là quan hệ M:1 từ tập thực thể yếu tới tập xác định và sự tham gia của tập thực thể yếu trong quan hệ là đầy đủ.

Tập thực thể yếu

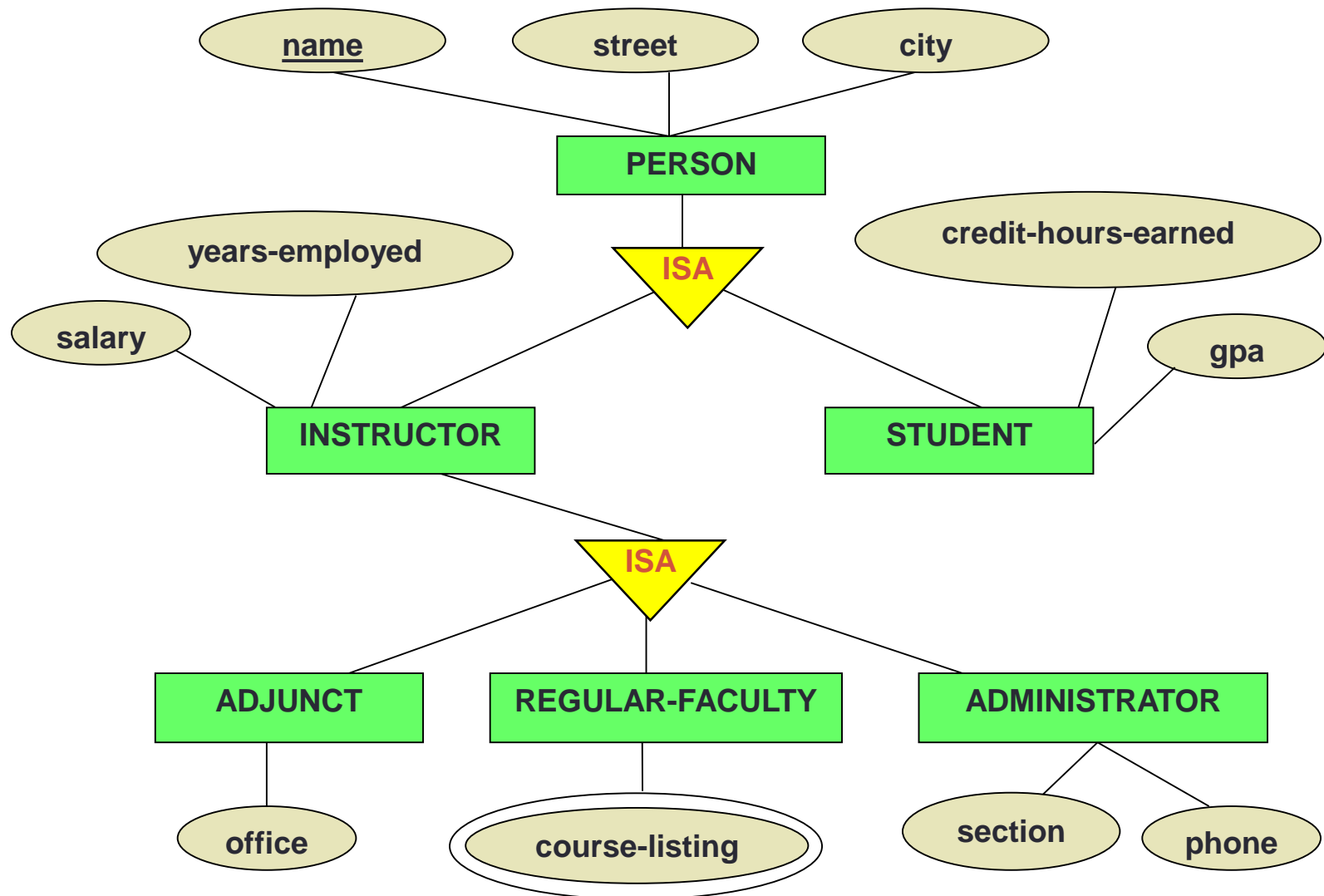
- Khóa chính của một tập thực thể yếu được cấu thành bởi khóa chính của tập thực thể xác định và thuộc tính phân biệt của tập thực thể yếu.



Cụ thể hoá – Tổng quát hoá

- **Cụ thể hóa** là quá trình thiết kế các phân nhóm trong một tập thực thể. Nghĩa là, phân biệt các tập thực thể con trong một tập thực thể khi chúng có các thuộc tính không giống nhau.
- **Tổng quát hóa** là quá trình nhiều tập thực thể được đồng bộ vào một tập thực thể ở mức cao hơn trên cơ sở các thuộc tính chung.
- Cụ thể hóa và tổng quát hóa là 2 khái niệm ngược và có thể hoán đổi cho nhau trong việc thiết kế các lược đồ CSDL.
- Trong lược đồ E-R, không có sự khác biệt giữa tổng quát hóa và cụ thể hóa mà chỉ khác nhau ở cách xem sơ đồ từ dưới lên hay từ trên xuống.

Cụ thể hoá – Tổng quát hoá



Kế thừa thuộc tính

- Các tập thực thể ở mức thấp hơn kế thừa các thuộc tính từ tập thực thể ở mức cao hơn.
- Tập thực thể mức thấp hơn cũng kế thừa các mối quan hệ thuộc về tập thực thể mức cao hơn định nghĩa nó.
- **Cây phân cấp** các tập thực thể:
 - Thực thể ở mức cao nhất nằm trên đỉnh của cây phân cấp.
 - **Kế thừa đơn**: Tập thực thể ở mức thấp hơn có một mối quan hệ ISA.
 - **Đa kế thừa**: Tập thực thể mức thấp có nhiều hơn một mối quan hệ ISA.

Các ràng buộc

- **Ràng buộc thứ nhất:** Xác định thực thể nào có thể là thành viên của tập thực thể mức thấp hơn.
- **Ràng buộc thứ hai:** Trong quá trình tổng quát hóa, liệu các thực thể có thuộc vào nhiều hơn một tập thực thể ở mức thấp hơn không.

Các tập thực thể ở mức thấp hơn có thể:

- **Không giao nhau:** Một thực thể không thể thuộc vào nhiều tập thực thể ở mức thấp hơn.
- **Giao nhau:** Một thực thể có thể thuộc vào nhiều tập thực thể ở mức thấp hơn trong cùng một quá trình tổng quát hóa.

Các ràng buộc

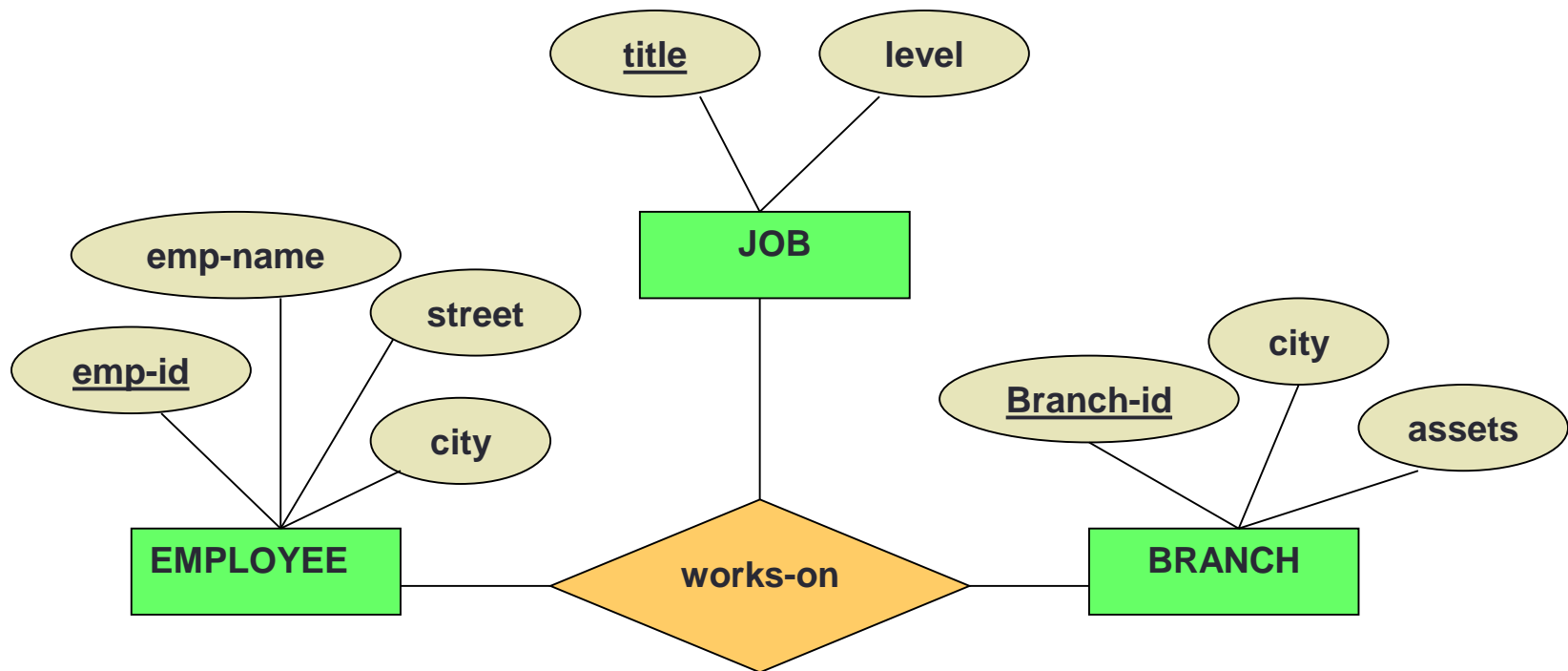
- **Ràng buộc thứ ba:** Là ràng buộc dựa trên tính toàn bộ, xác định xem một thực thể trong tập thực thể ở mức cao có thuộc vào một trong các tập thực thể ở mức thấp hơn hay không.

Có 2 loại ràng buộc này:

- **Tổng quát hóa/cụ thể hóa toàn bộ:** mỗi thực thể ở mức cao phải thuộc vào một thực thể ở mức thấp hơn.
- **Tổng quát hóa/cụ thể hóa một phần:** Tồn tại thực thể ở mức cao không thuộc vào tập thực thể nào ở mức thấp hơn. (Đây là trường hợp mặc định).

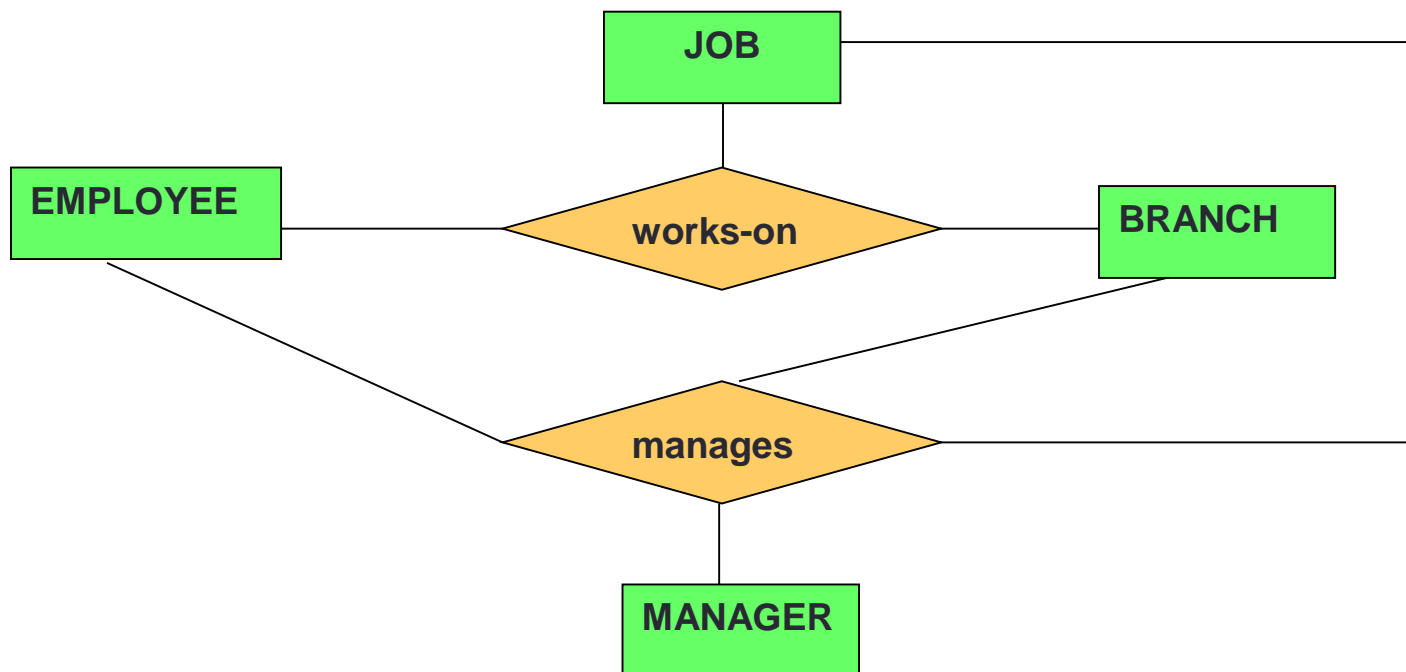
Tập thực thể kết hợp

- Hạn chế của lược đồ E-R: không thể biểu diễn rõ các mối quan hệ giữa các tập thực thể trong các liên kết.



Tập thực thể kết hợp

- Giả sử muốn ghi nhận việc quản lý từng công việc được thực hiện bởi một nhân viên tại một chi nhánh; nghĩa là muốn lưu trữ người quản lý cho quan hệ ba ngôi (**EMPLOYEE**, **BRANCH**, **JOB**).



Tập thực thể kết hợp

Câu hỏi đặt ra:

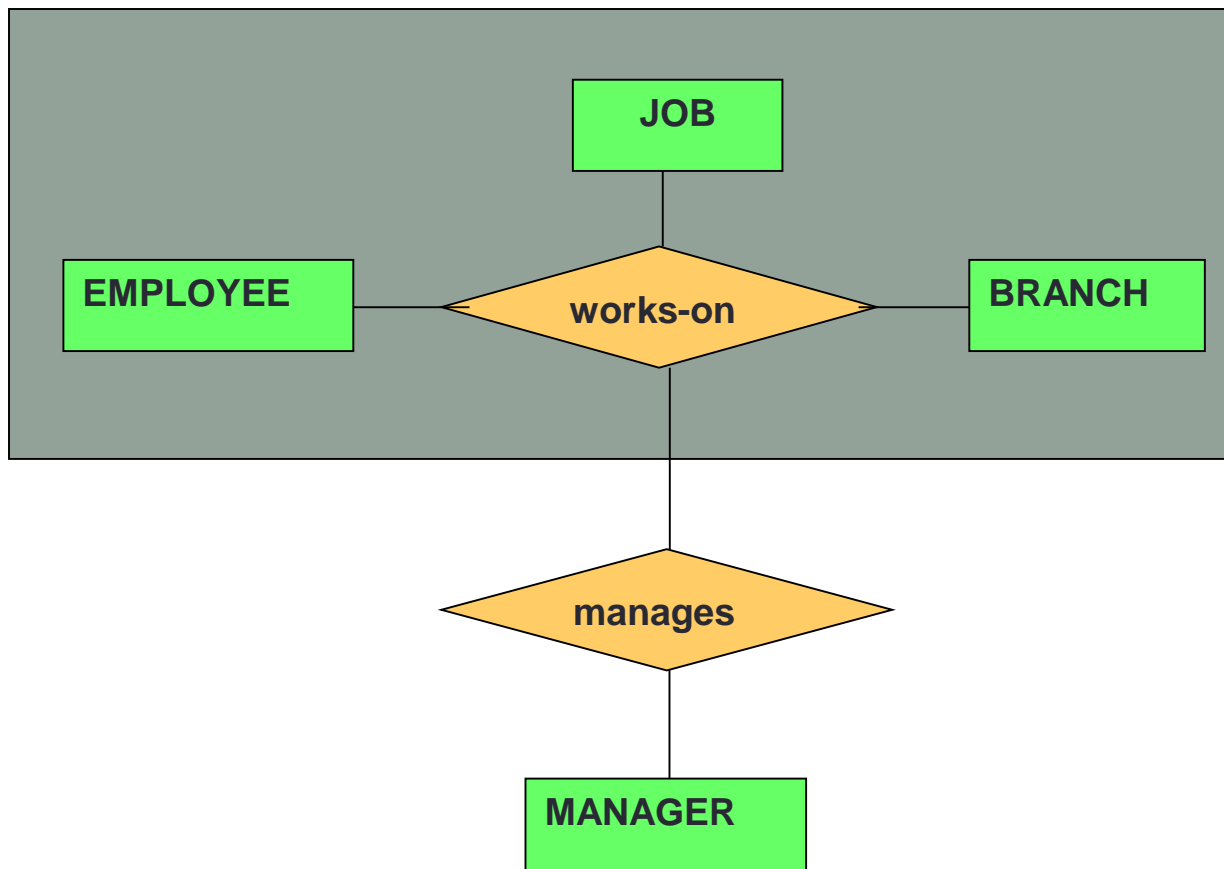
Tại sao không thể hiện mối quan hệ hai ngôi giữa MANAGER và EMPLOYEE ?

=> Câu trả lời:

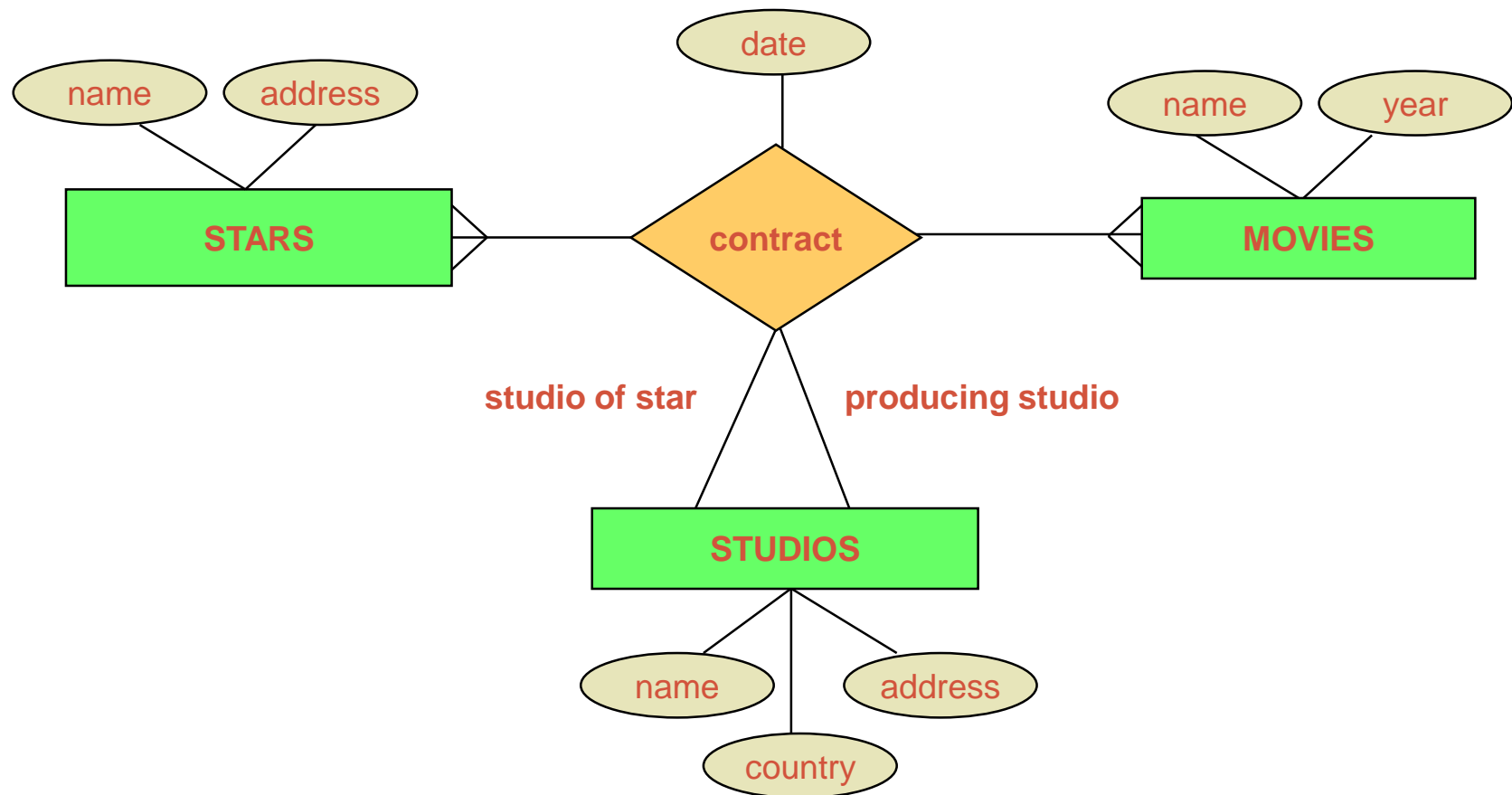
Một quan hệ hai ngôi sẽ không cho phép thể hiện liên kết (BRANCH, JOB) của một EMPLOYEE (nhân viên) được quản lý bởi người quản lý.

=> Cách tốt nhất để mô hình hóa là sử dụng **kết hợp** (hay **tích hợp**). *Kết hợp là một sự trừu tượng thông qua việc coi các mối quan hệ như là các thực thể ở mức cao.*

Tập thực thể kết hợp



Liên kết nhiều ngôi



Lược đồ ER và cách chỉ thị vai trò

- Các vai trò trong lược đồ E-R được biểu diễn bằng việc gán nhãn lên các đường kết nối giữa các tập thực thể và các tập quan hệ.
- Các vai trò này có thể được xác định cho các mối quan hệ đệ quy, nhị phân, và không nhị phân.

