CS348 INFORMATION SYSTEMS

Hisham Benotman

Most slides are adopted with permission from D. Maier, and L. Delcambre.

PURDUE UNIVERSITY | Department of Computer Science

Announcements

- HW1 solution is posted
- HW1 grade are expected by the end of this weekend.

- HW2 is due tomorrow
  - Make sure you submit the right file (hw2.py).
  - Run your hw2.py to make sure all queries print correctly.

Overview of Database Design

- Data requirements collection (e.g., using use cases)
- Conceptual design: (ER Model is used for this.)
  - What are the **entities** and **relationships** we need?
- Logical design:
  - Transform ER design to Relational Schema
- Schema Refinement: (Normalization)
  - Check relational schema for redundancies and related anomalies.
- Physical Database Design and Tuning:
  - Consider typical workloads; (sometimes) modify the database design; select indexes.

After DB design (and other parts of app. design) is complete, create the database and build the application (or use database for other purposes, such as data analysis).

Entity-Relationship Model is a different model than the Relational Model

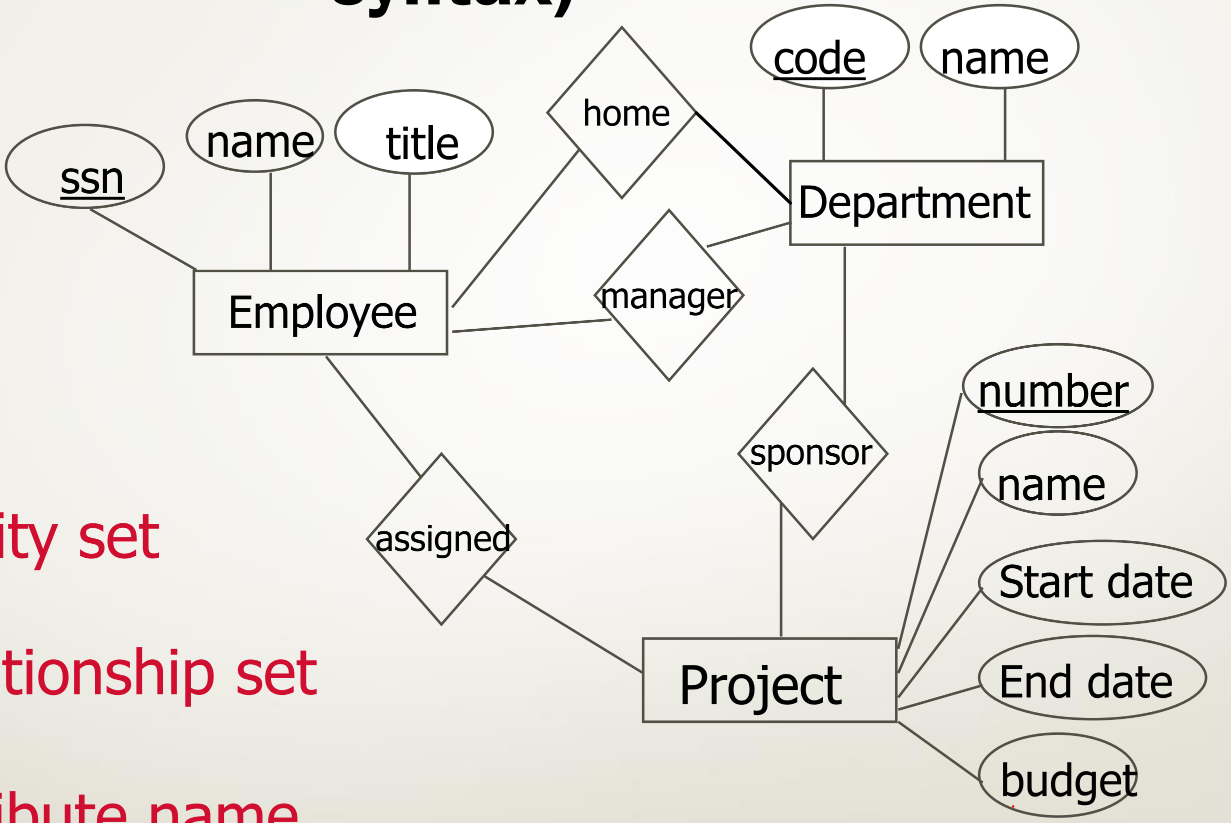
- Relational model has:
  - tables (relations) with attributes, keys, foreign keys, domain definitions for attributes
- Entity-Relationship model has:
  - Entity sets with attributes, keys, and domain definitions for attributes
  - Relationship sets among entity sets with cardinality constraints

Entity Relationship Model

ER Model

- Proposed by Peter Chen in 1976
- Gives us a notation to specify
  - What information the database must hold
  - How the bits of information relate to one another

Entity-Relationship Diagram (original syntax)



Legend:

□ Entity set

◇ Relationship set

○ Attribute name

Definitions

- *Entity*: Real-world object distinguishable from other objects.

  An entity is described using a set of *attributes*.

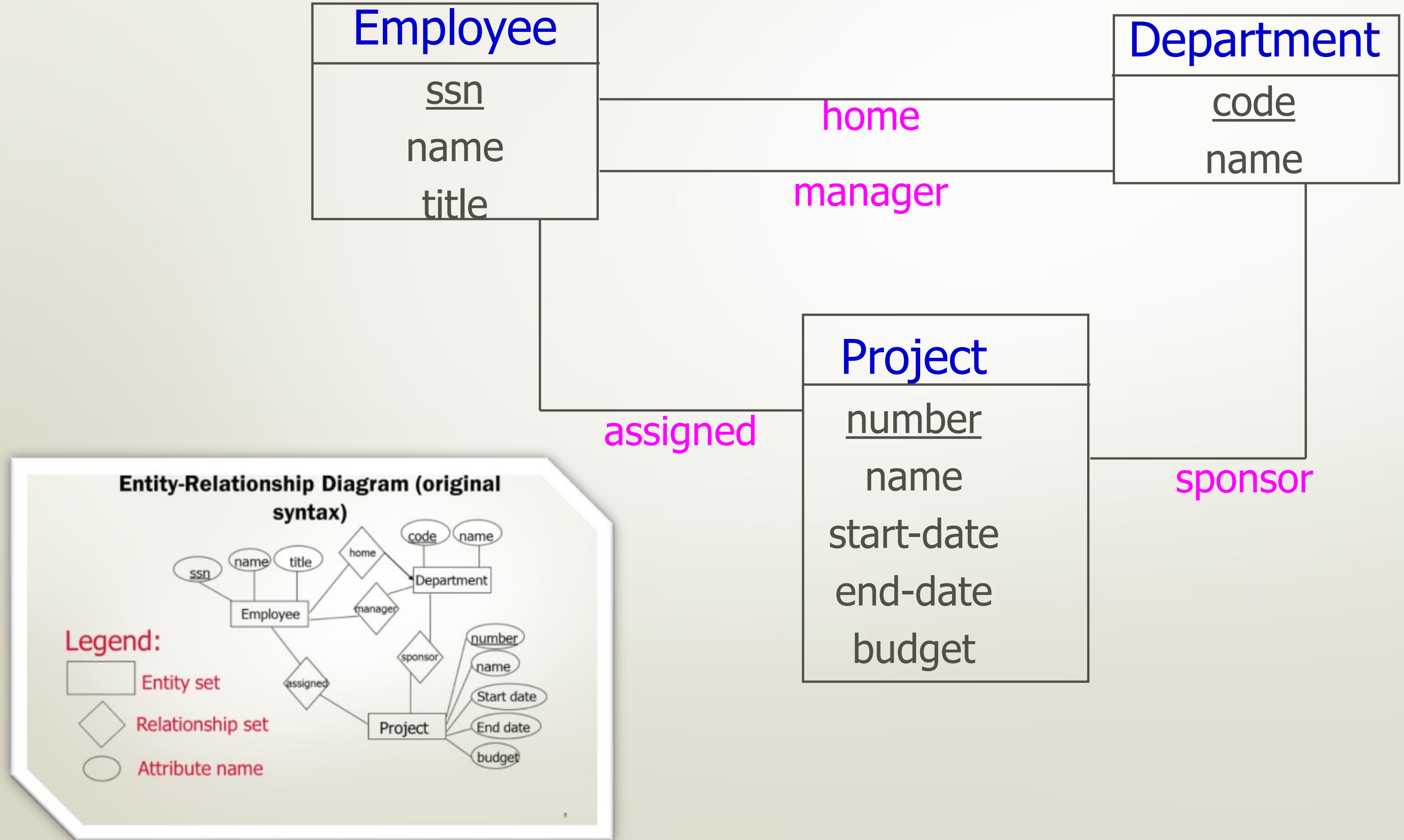
- *Entity Set*: A collection of similar entities. E.g., all employees.

  (often referred to as just entity, which blurs the distinction between type and instance)

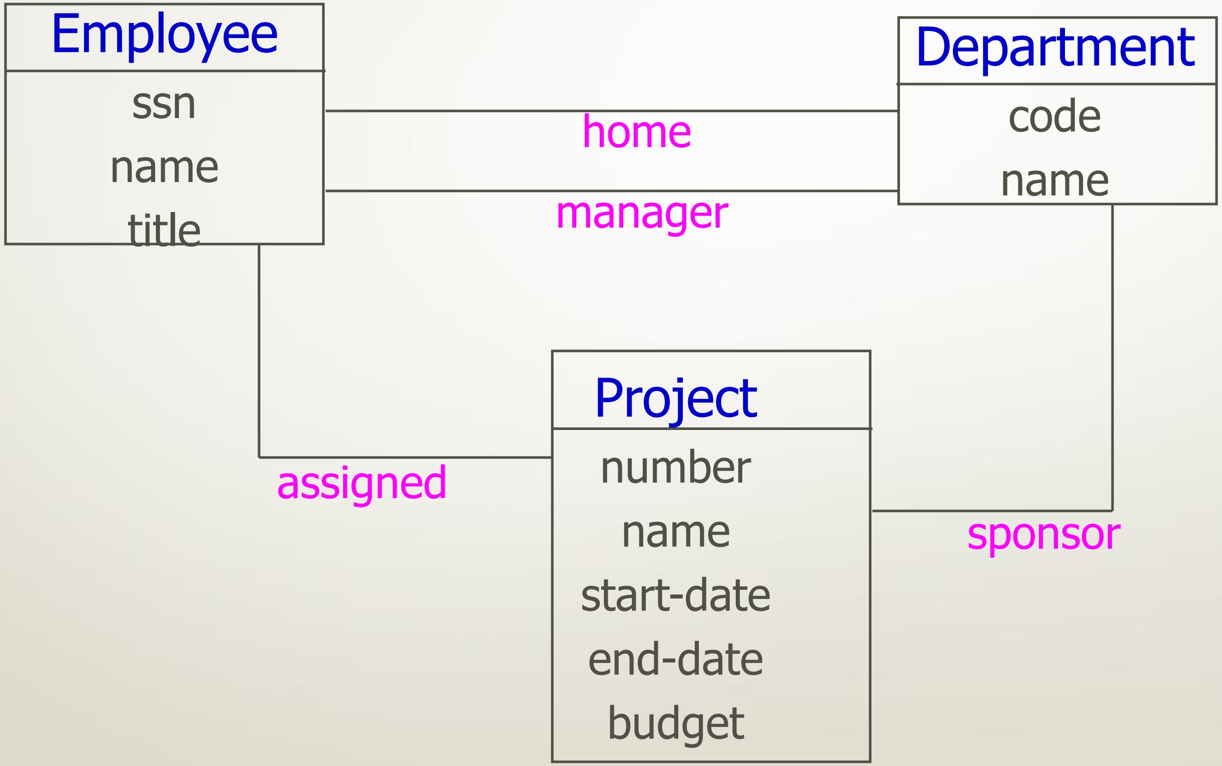
- *Relationship*: Association among 2 or more entities. E.g., Kristin's home department is Research & Development.

- *Relationship Set*: Collection of similar relationships. E.g., Home (often referred to as just relationship)

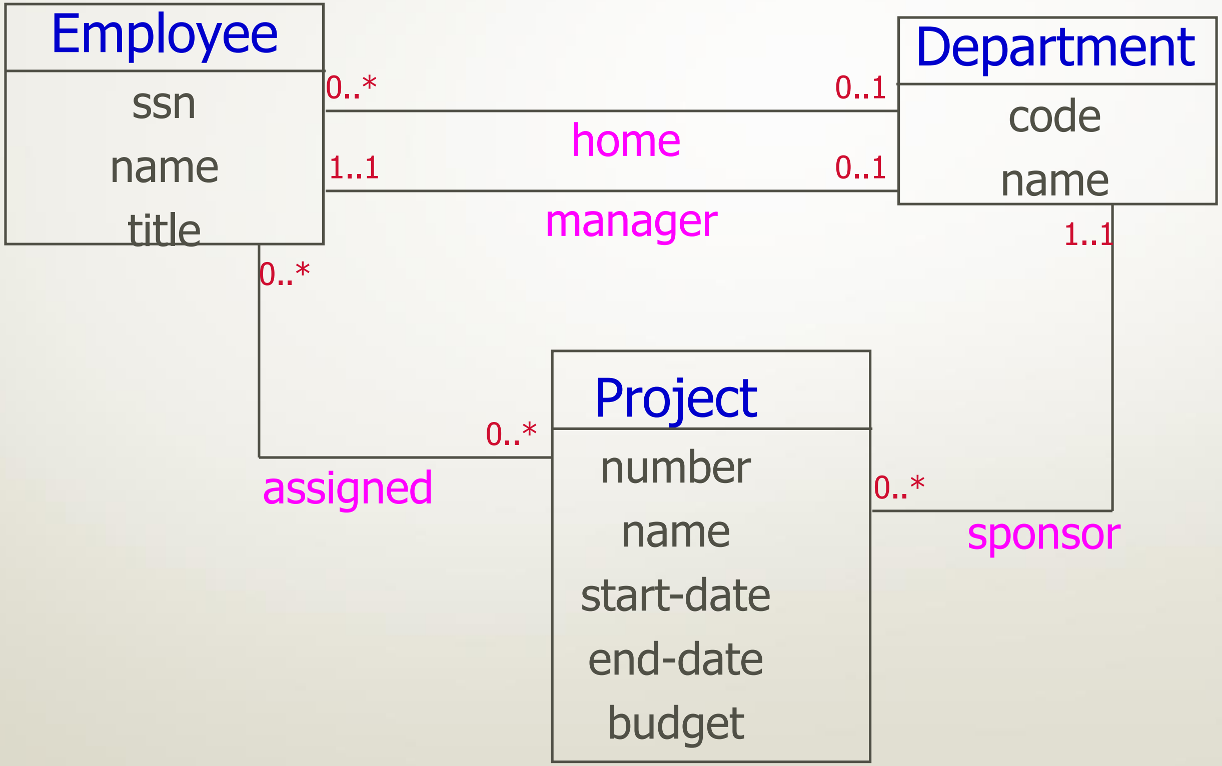
UML version of the same E-R Diagram



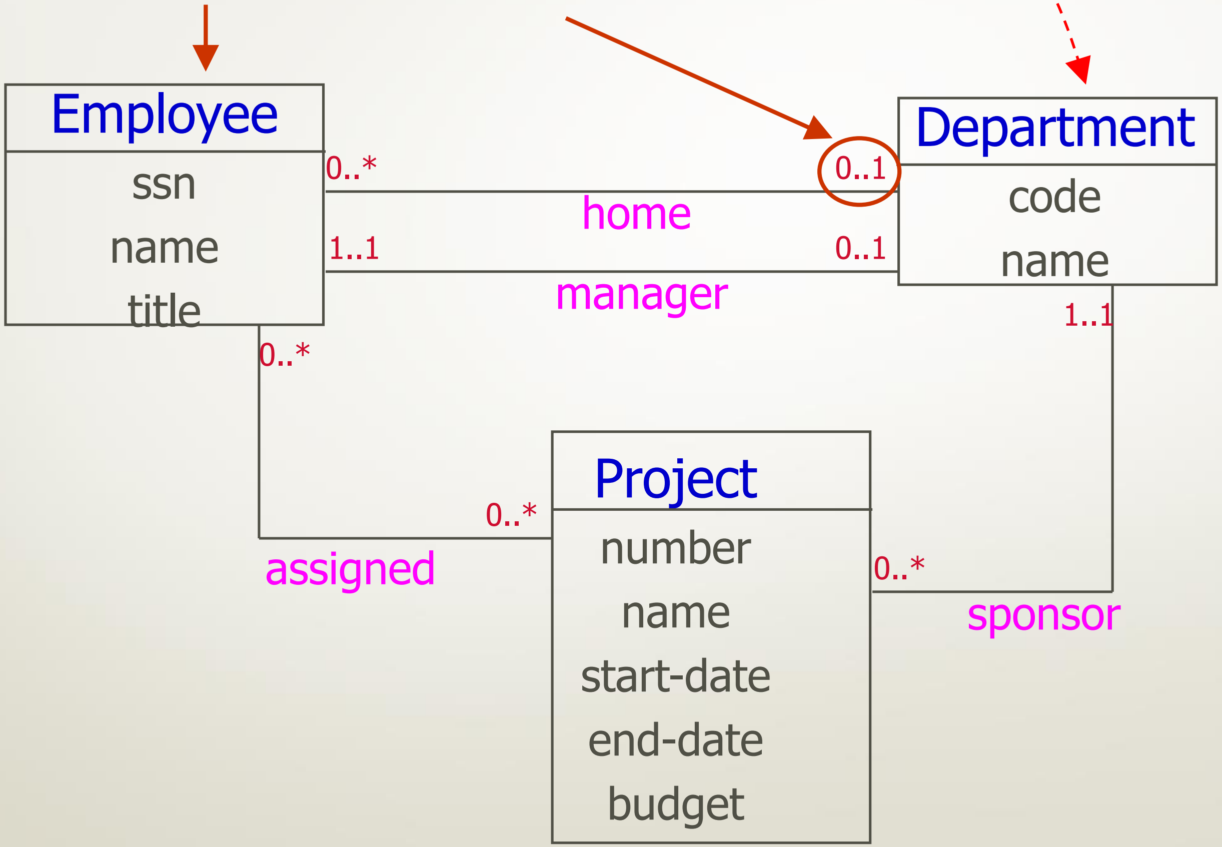
Cardinality Constraints on Relationship sets: How many entities can participate?



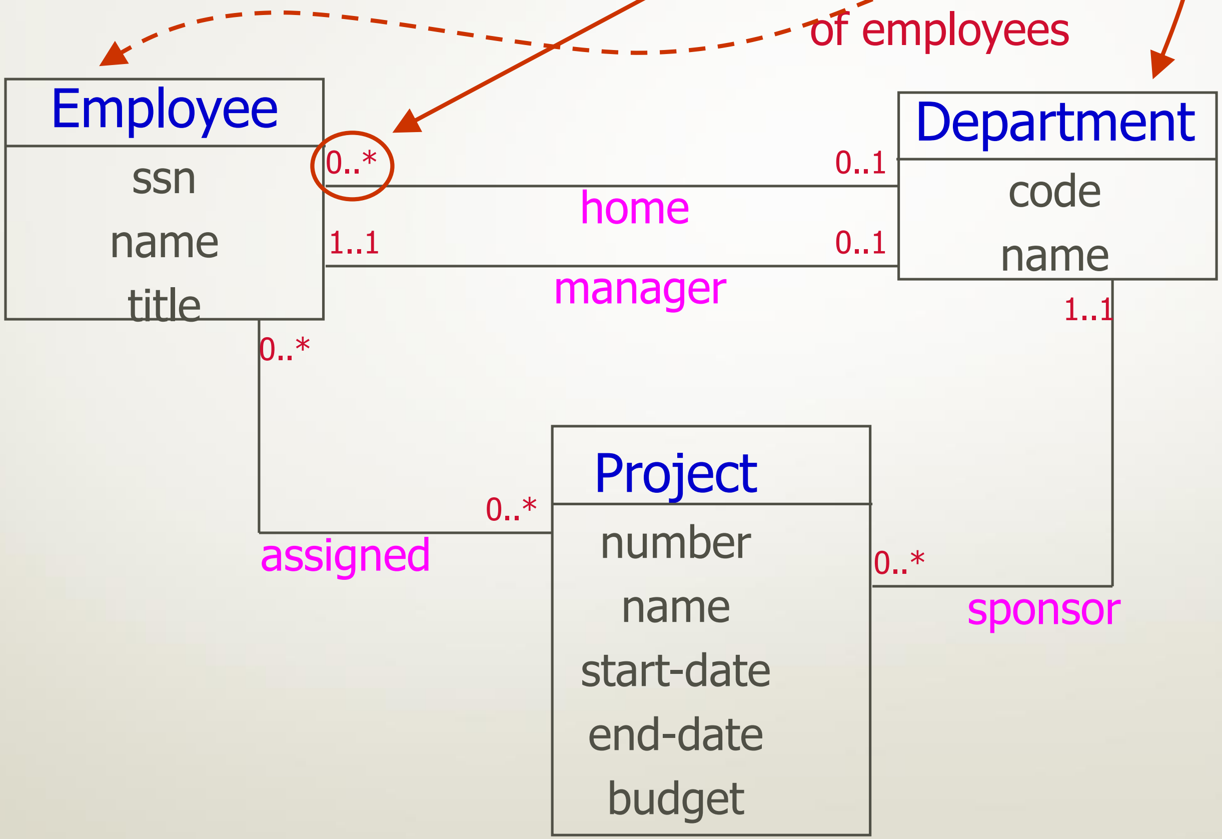
Cardinality Constraints on Relationship sets: How many entities can participate?



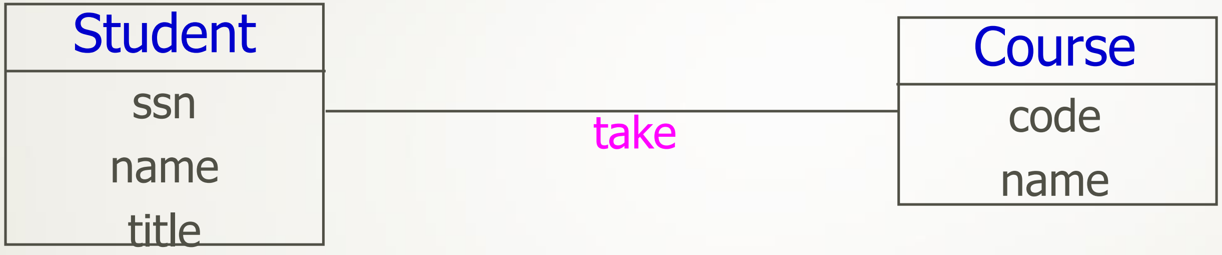
An employee can have **<u>0 or 1</u>** home departments



A department can be home to **0 to any number** of employees



Cardinality Constraints Exercise



Student

ssn

name

title

take

Course

code

name

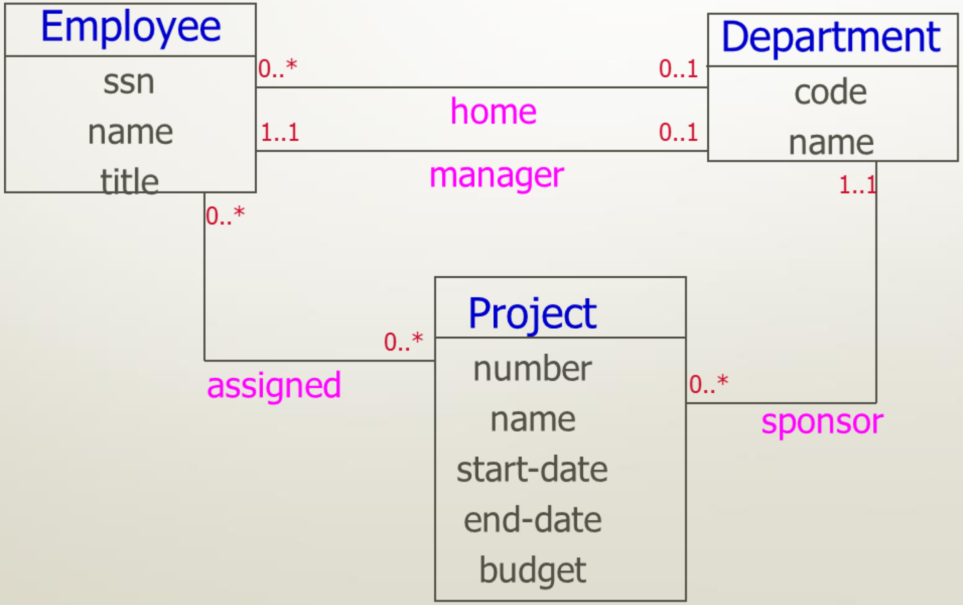
Equivalent Relational Schema

Employee (ssn, name, title, home-dept)

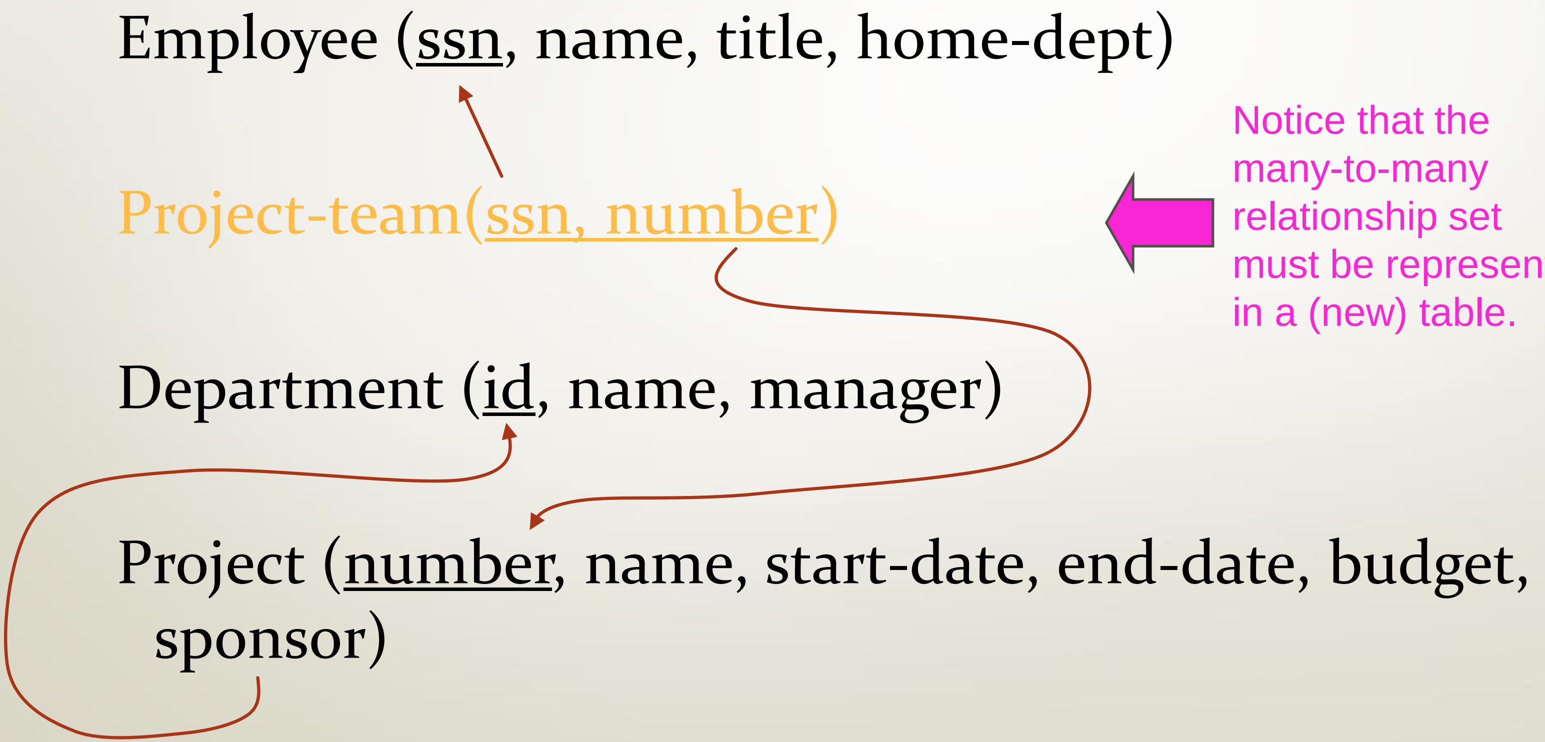
Project-team(ssn, number)

Department (id, name, manager)

Project (number, name, start-date, end-date, budget, sponsor)



Equivalent Relational Schema - with foreign keys shown



Employee (ssn, name, title, home-dept)

Project-team(ssn, number)

Department (id, name, manager)

Project (number, name, start-date, end-date, budget, sponsor)

Notice that the many-to-many relationship set must be represented in a (new) table.

Equivalent Relational Schema

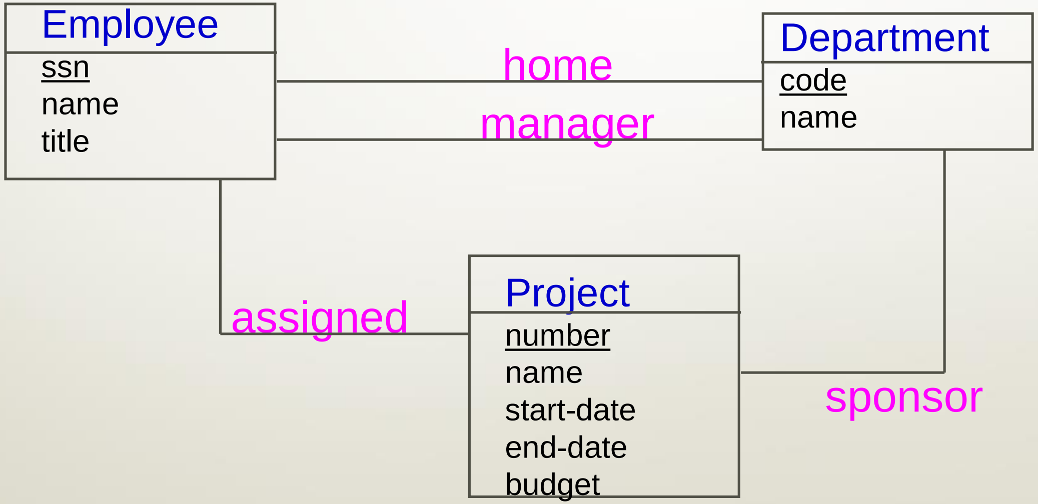
Employee (ssn, name, title, home-dept)

Project-team(ssn, number)

Department (id, name, manager)

Project (number, name, start-date, end-date, budget, sponsor)

Which representation is more readable to end users?

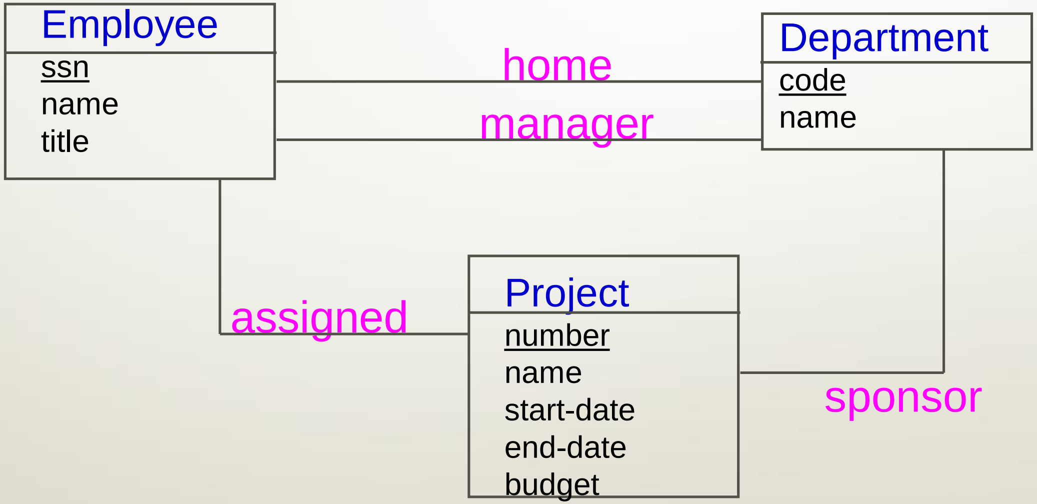


Exercise

How would you add the following data items (item names can be ambiguous, so make your own assumptions)?

office number, number of employees, email, total expenditure, building address, building number of floors, building code,

department is located in a building, employee is located in a building.

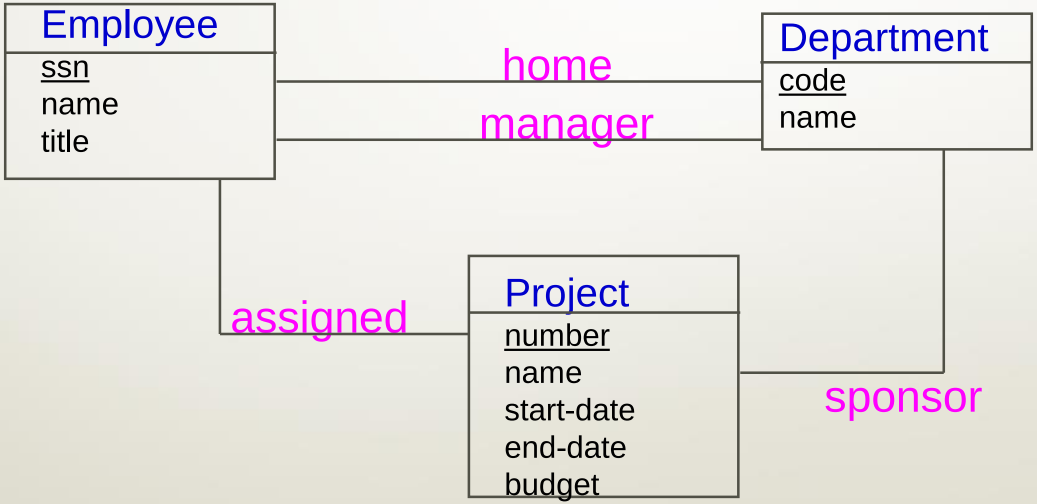


Exercise

How would you add the following data items (item names can be ambiguous, so make your own assumptions)?

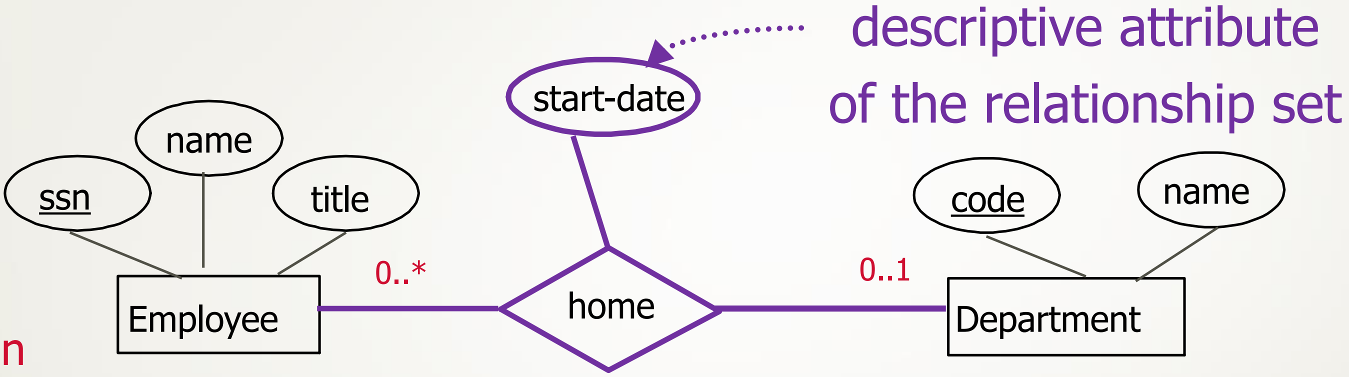
office number, number of employees, email, total expenditure, building address, building number of floors, building code,

department is located in a building, employee is located in a building.

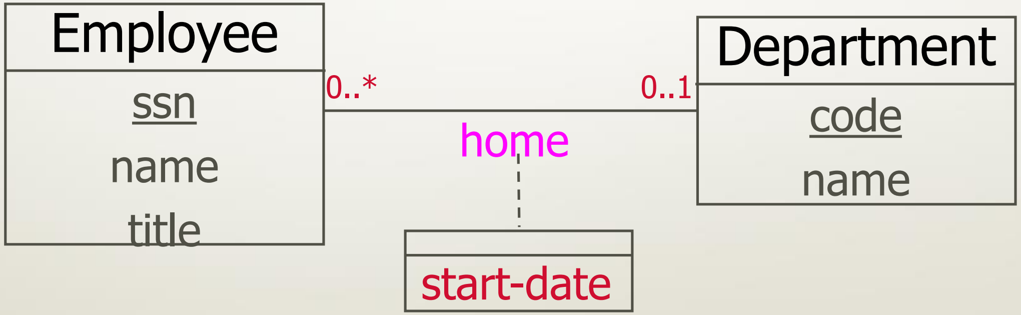


Relationship sets can have attributes

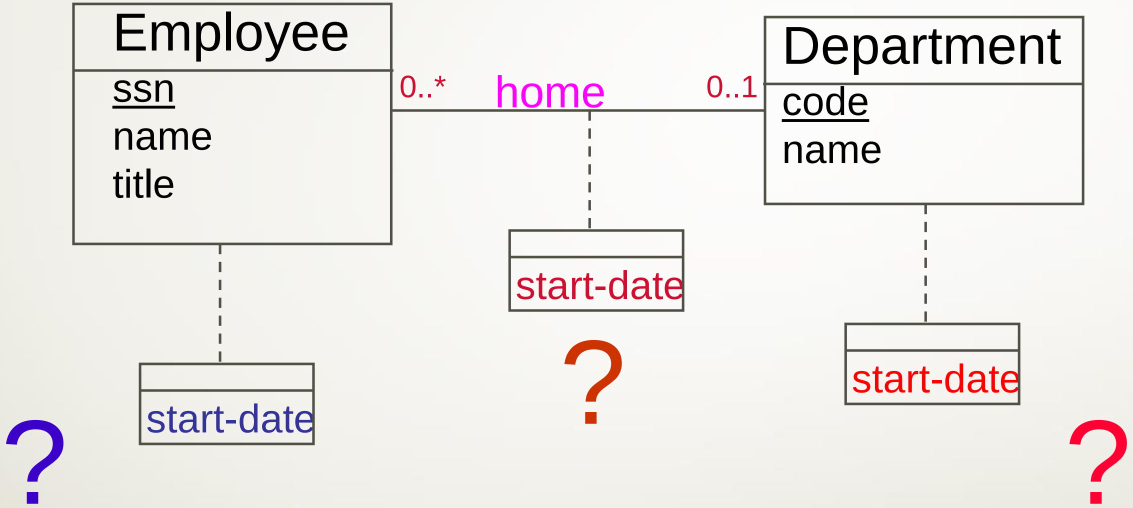
E-R notation



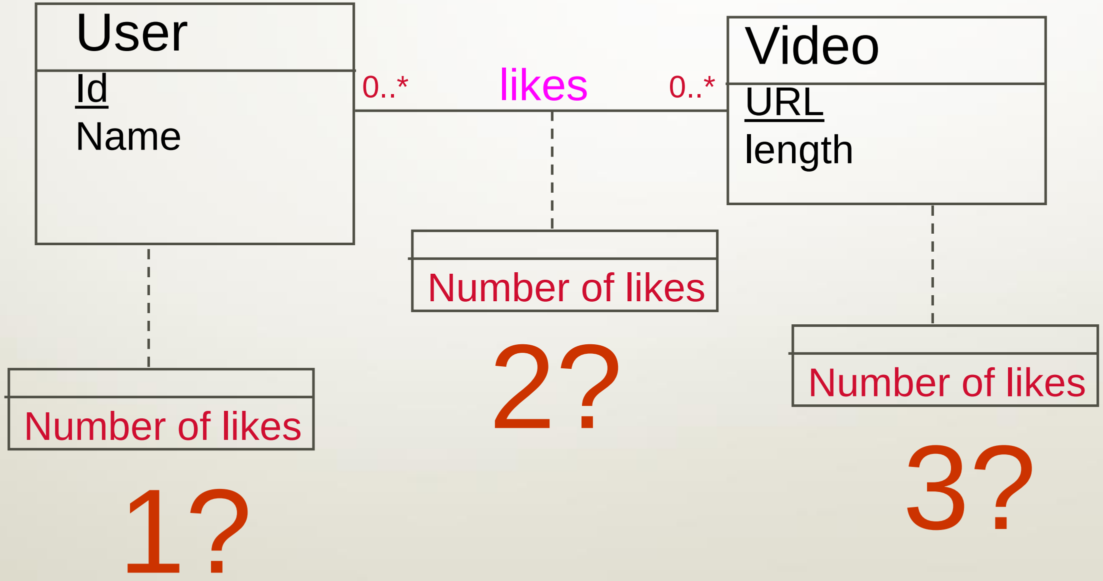
UML notation



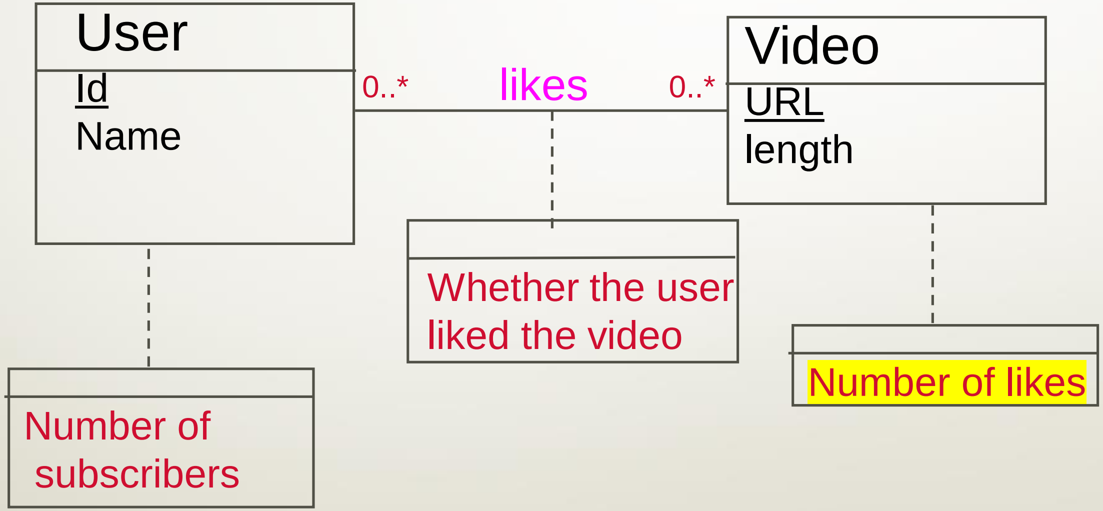
Try all three locations for the attribute: which one makes sense?



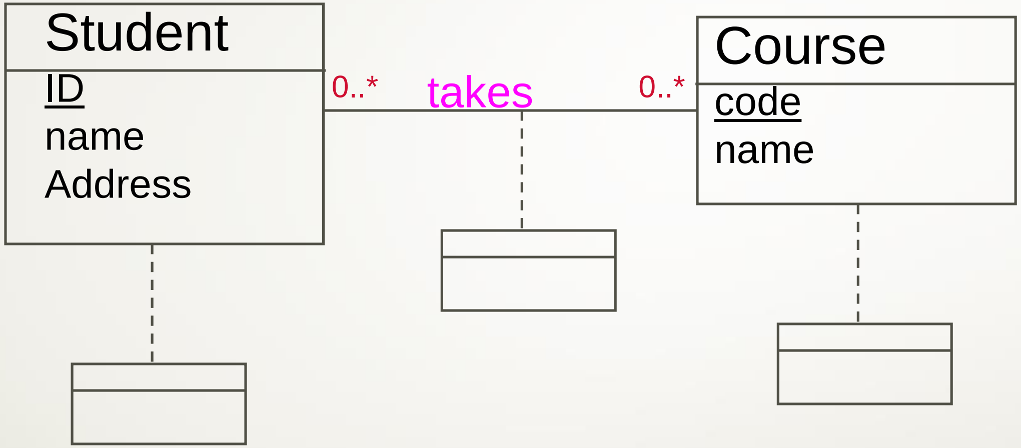
Hotseat Question
Where should we place the ‘number of likes’ attribute for a YouTube video?



Example

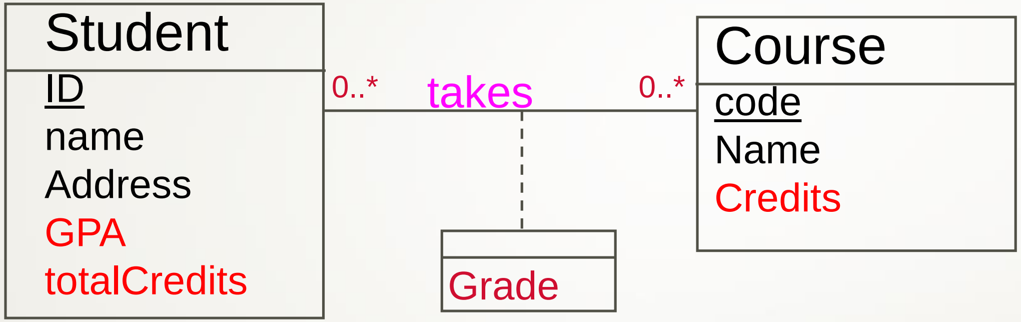

User
Id
Name
0..*
likes
0..*
Video
URL
length
Whether the user liked the video
Number of subscribers
Number of likes

Try all three locations for the attributes: which one makes sense?



Where would you place the following attributes: GPA, Grade, Credits, totalCredits?

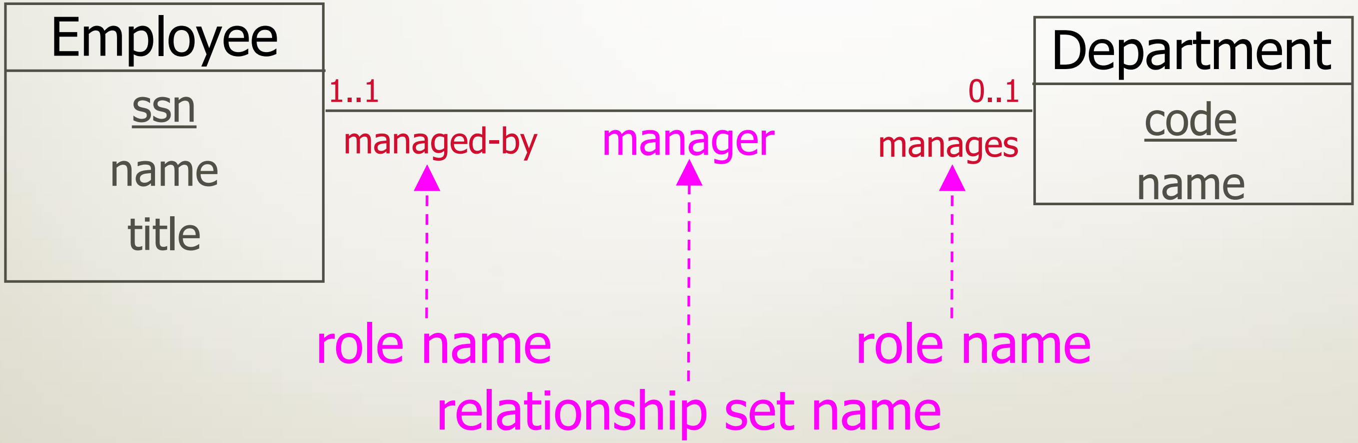
Try all three locations for the attribute: which one makes sense?



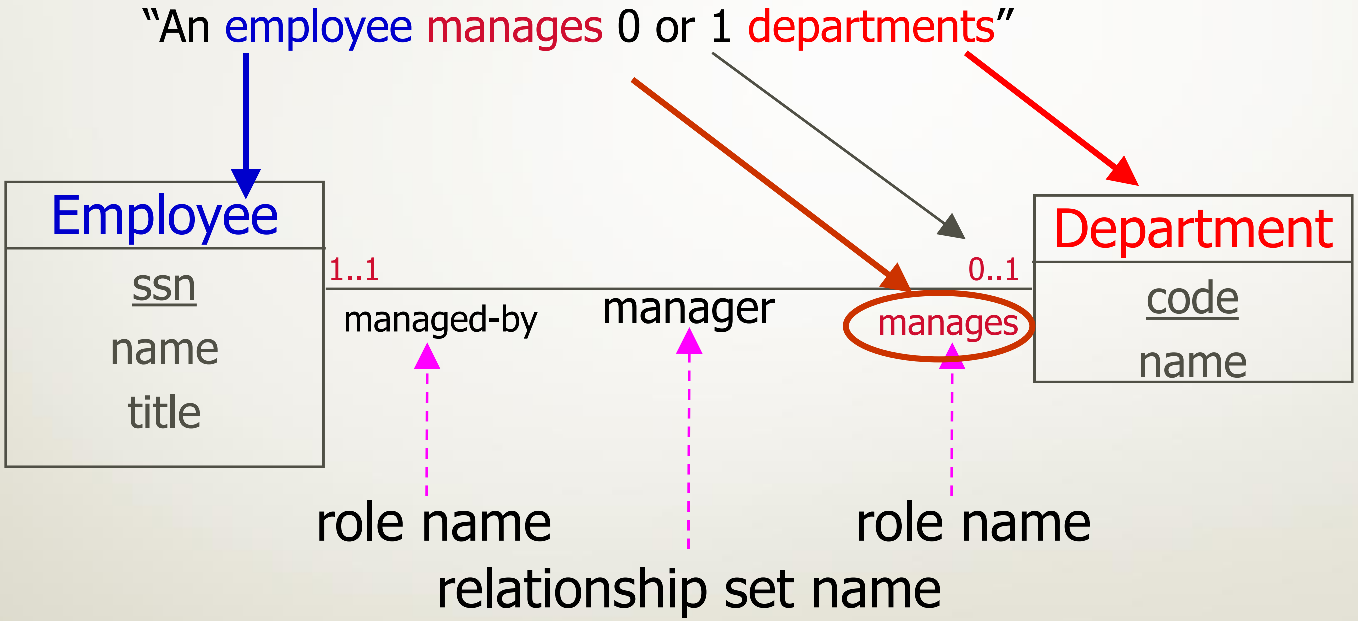
Where would you place the following attributes: GPA, Grade, Credits, totalCredits?

Relationship sets can have role names

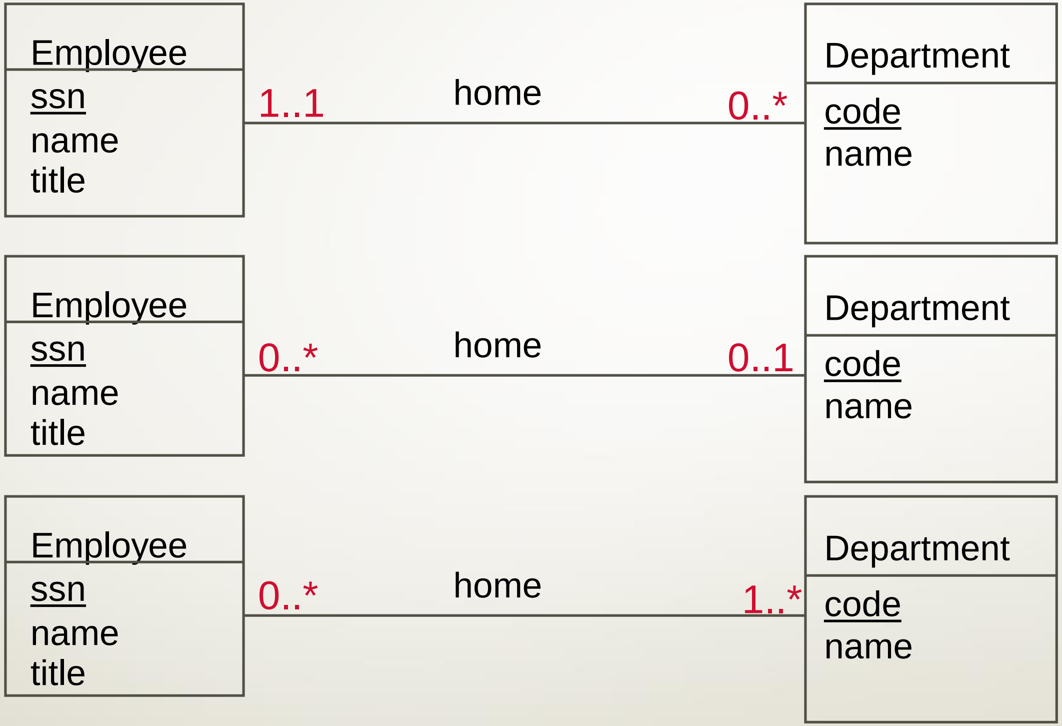
(in addition to the name of the relationship set)



Example: reading role names

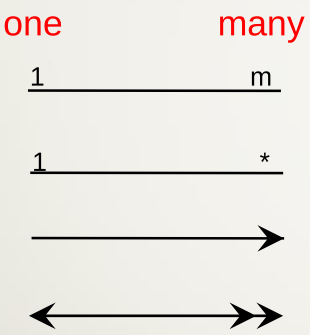


Some Alternative Constraints (in UML)

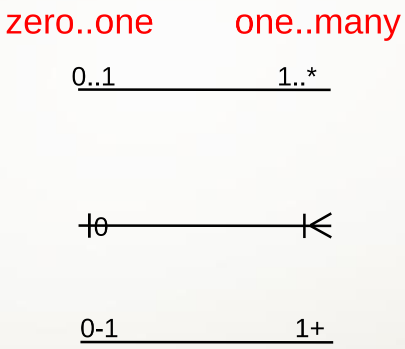


Which one is right? We must discover the semantics of the application!

Various notations for “one-to-many”

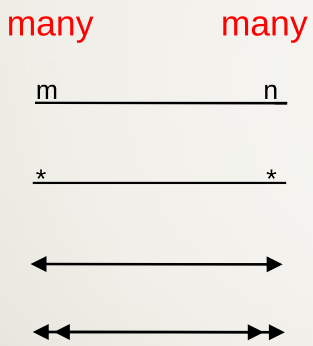


maximum cardinalities only

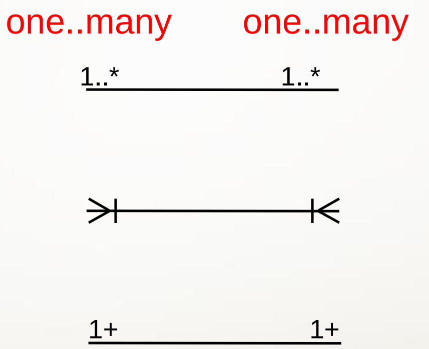


minimum and maximum cardinalities

Various notations for “many-to-many”



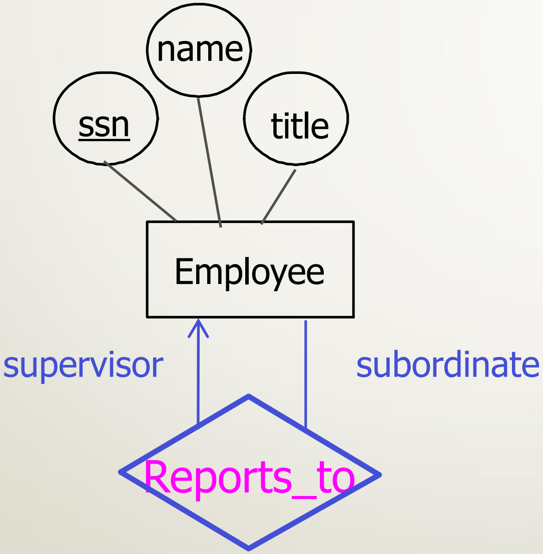
maximum cardinalities only



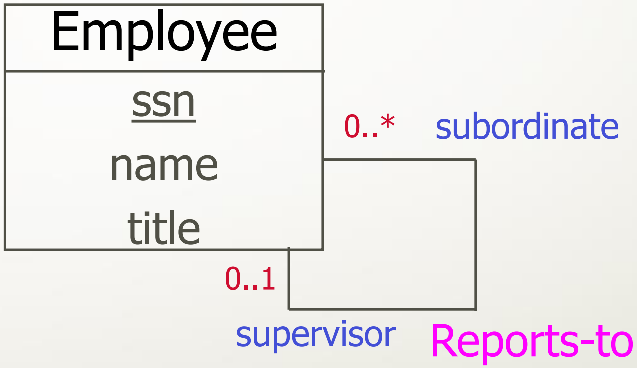
minimum and maximum cardinalities

Same entity sets can participate in different “roles” for the same relationship set

E-R notation

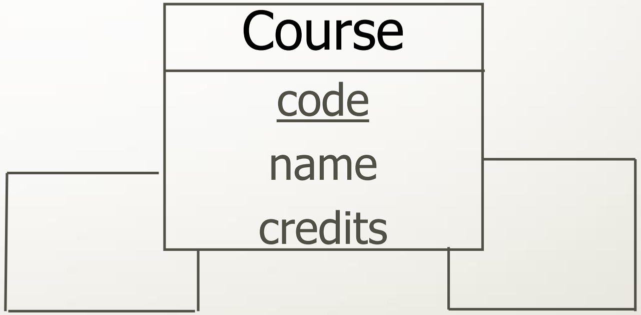


UML notation



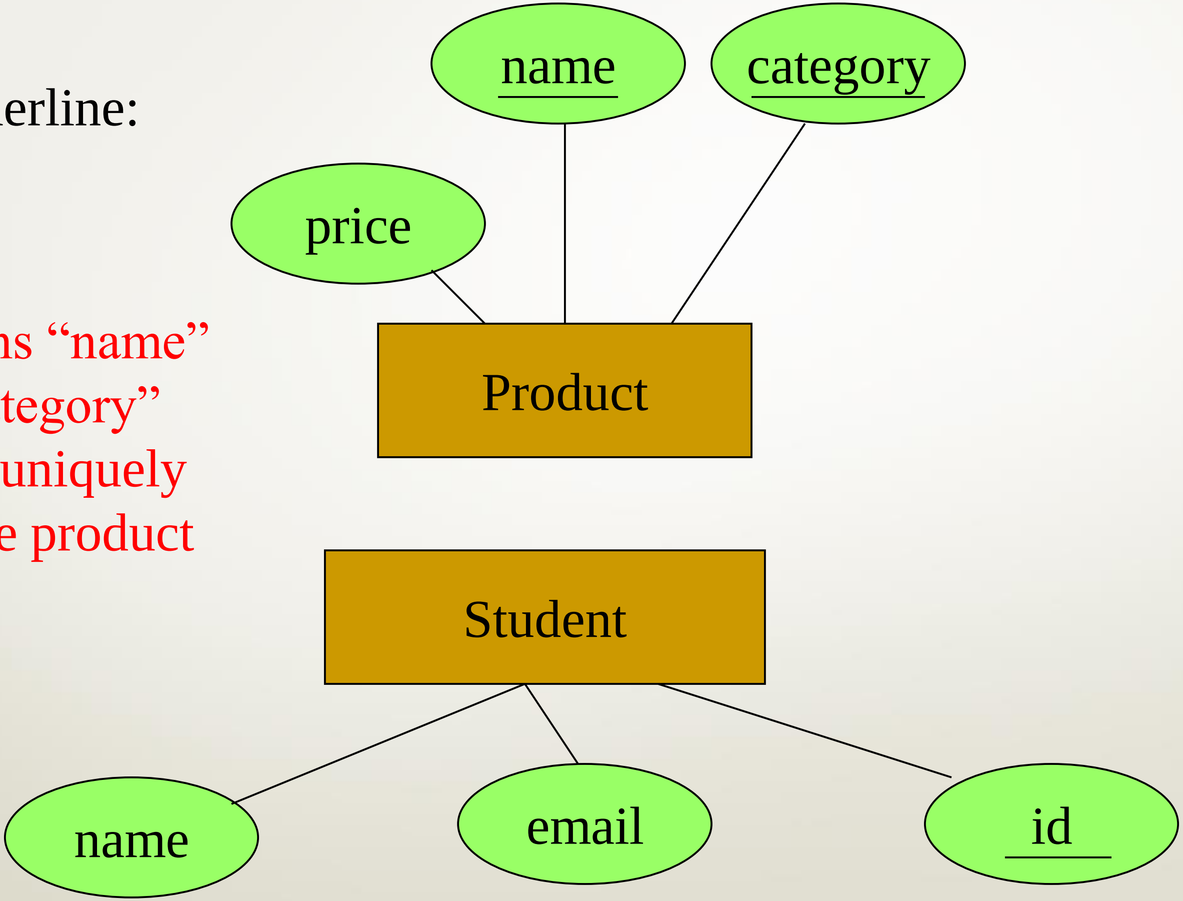
Same entity sets can participate in different "roles" for the same relationship set

Can you think of relationships between courses?



Keys in E/R Diagrams

Underline:



This means "name" and "category" together uniquely determine product

Weak Entity Sets

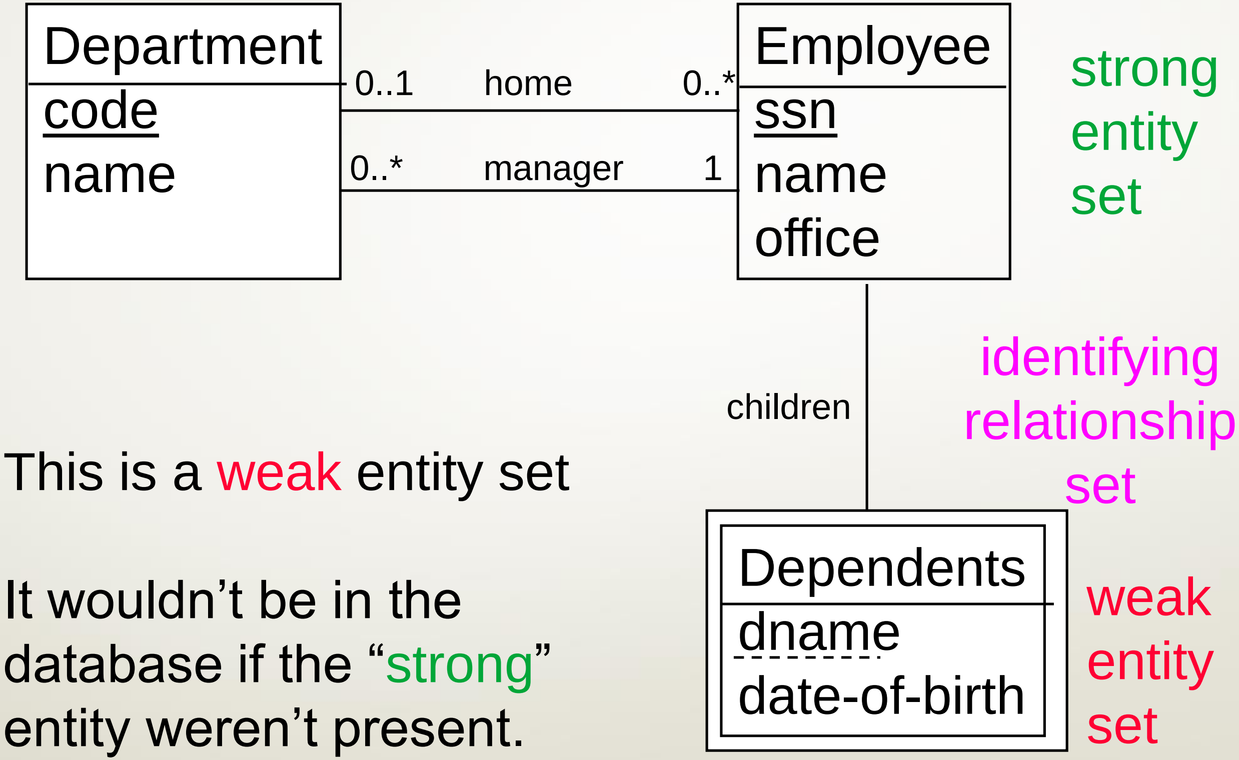
- Occasionally, entities of an entity set need “help” to identify them uniquely.

- Entity set *E* is *weak* if in order to identify entities of *E* uniquely, we need to follow one or more many-one relationships from *E* and include the key of the related entity sets.

- Note: not an is-a relationship because *E* is not a “subclass” of *F:* Univ and Team

Weak Entity Sets

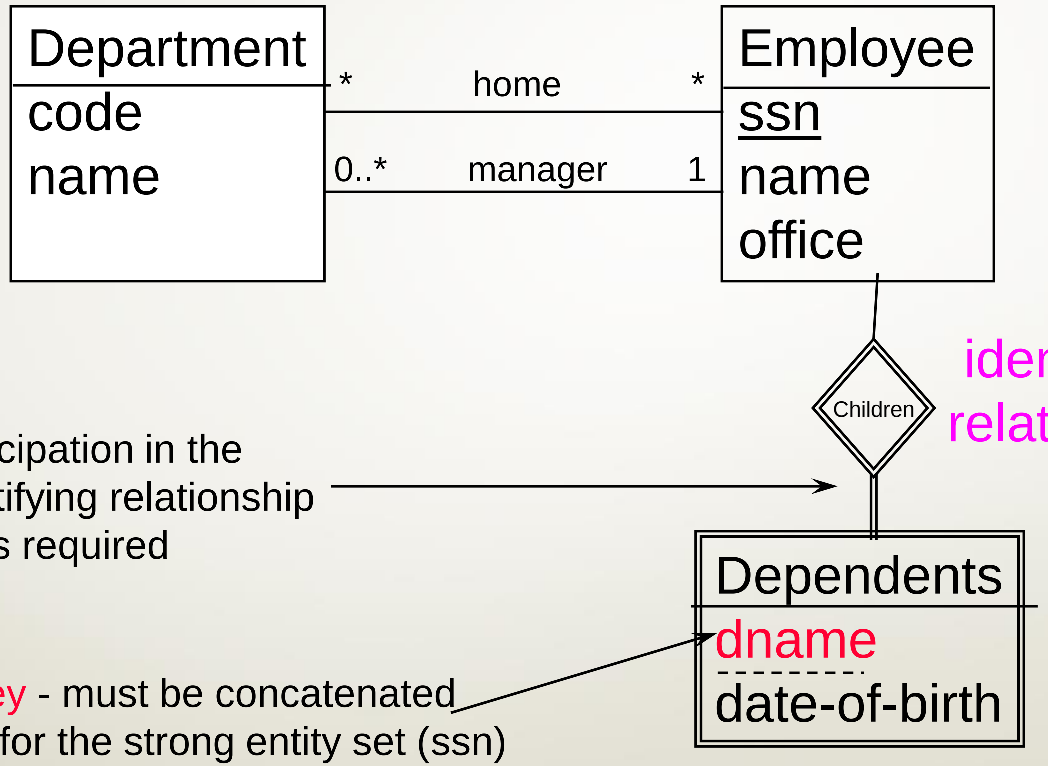
(and Identifying Relationship sets)



This is a weak entity set

It wouldn’t be in the database if the “strong” entity weren’t present.

Weak Entity Sets and Identifying Relationship sets: Alternative Notation



strong entity set

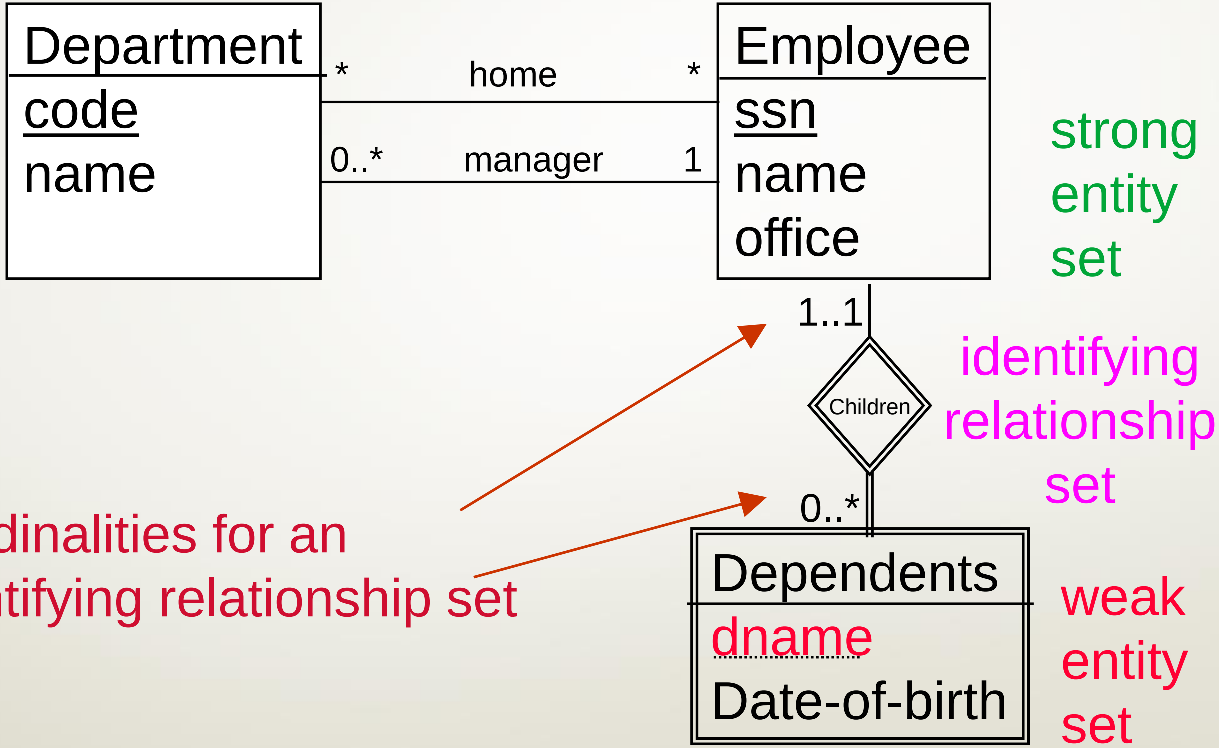
identifying relationship set

weak entity set

participation in the identifying relationship set is required

partial key - must be concatenated with key for the strong entity set (ssn)

Weak Entity Sets and Identifying Relationship sets: Alternative Notation (cont.)



Cardinalities for an identifying relationship set

Hotseat Question

- Which of the following relationships may connect a strong entity set to a weak entity set?
  1. Course(id, name, credits) to course-offering (code, year, semester, number of students).
  2. Student(id, name, …) to course(id, name, credits)
  3. Building(code, name, address, number of floors) to office (number, sq. feet, floor)

Hotseat Question

- Which of the following relationships may connect a strong entity set to a weak entity set?
    1. Course(id, name, credits) to course-offering (code, year, semester, number of students).
    2. Student(id, name, ...) to course(id, name, credits)
    3. Building(code, name, address, number of floors) to office (number, sq. feet, floor)