CS 5200
Database Systems

Conceptual Database Design
The Entity-Relationship Model

Reference:

Database Management Systems, by Ramakrishnan and Gehrke (Chapter 2 of the Textbook)

Dr. Hazra Imran

Databases: the continuing saga…

- We discussed that databases were great because they:
  - Store large amounts of data
  - Handle transactions
  - Allow efficient querying
  - *And many, many more classic favorites!*

- Before we can do all of these, we must design the database

Learning Goals for Chapter 2

1. Explain the purpose of an ER diagram and list the major components.
2. Given a problem description, create an ER diagram given a specification. Justify the decisions you make for entities, relationships, keys, key constraints, participation constraints, weak entities, is-a relationships, and aggregations.
3. Given a problem description, identify alternative representations of the problem concepts and evaluate the choices

Power of the ER model

- One of the most cited articles in Computer Science
  - "*The Entity-Relationship model – toward a unified view of data*" Peter Chen, 1976 (Cited by 11254)

Note : There are many variations of ER diagram

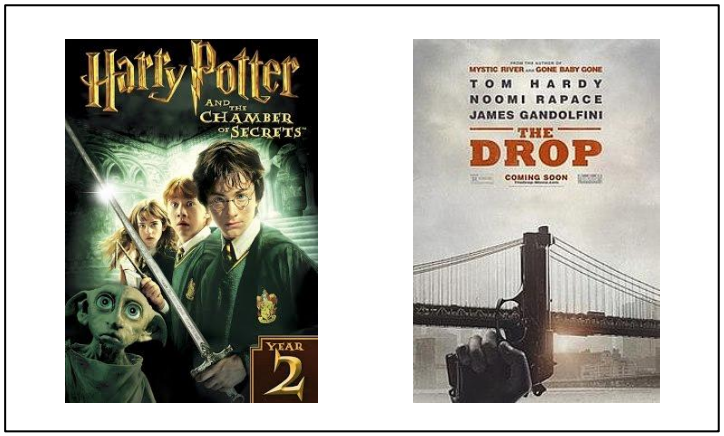
- Follow the same ER notation

Entities vs. Entity Sets

- An ***entity*** is a distinguishable object.
  - Example : Movie, Customer, Student

- An ***entity set*** is a collection of similar entities
  - Example : All of the movies (Movies), All of the customers

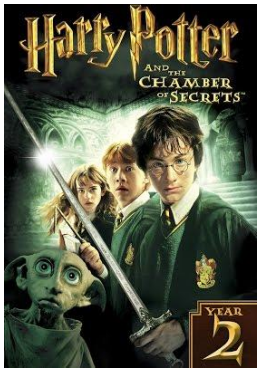
Movies



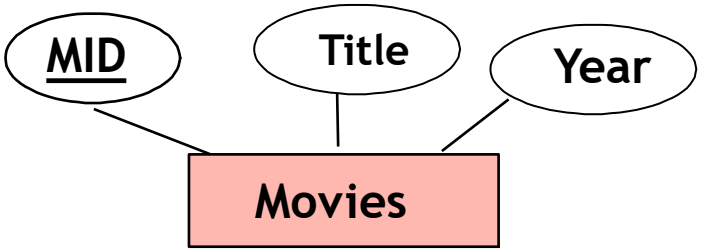
- An ***attribute*** is the property of (the entities of) an entity set

ER Model Basics: Entities

- Entities are ***not*** explicitly represented in E/R diagrams!
- An entity is described using a set of ***attributes***.



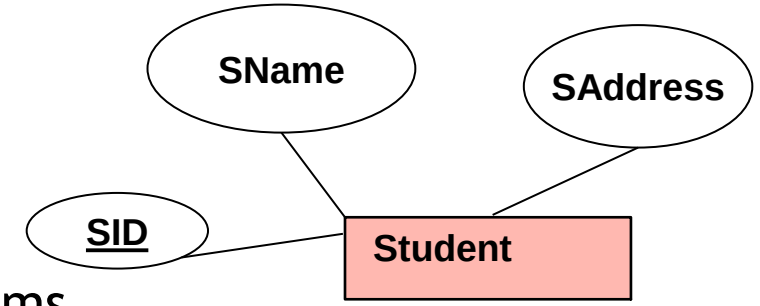
ID : M01
Title: Harry Potter and the Chamber of Secrets
Year : 2002



- Each attribute has a **domain**. (e.g.., float, date, string)
- Each entity set has a **key**.

Keys

- A *primary key* is the key chosen as the principal means to identify entities in an entity set
- *A superkey* is a key plus zero or more other attributes in the entity set



- Only primary keys are shown in ER diagrams
- The E/R model forces us to designate a single **primary** key, though there may be multiple candidate keys

Super key, Candidate key and Primary key

Super Keys : Super key stands for superset of a key. A Super Key is a set of one or more attributes that are taken collectively and can identify all other attributes uniquely.

For example Movie (MovieID, MovieName, drname) (drname is a director name)

We can have following super keys :

- (MovieID)
- (MovieID, MovieName)
- (MovieID, MovieName, drname)
- (MovieID, drname)
- (Moviename, drname)

Super keys as each one can identify the entities uniquely

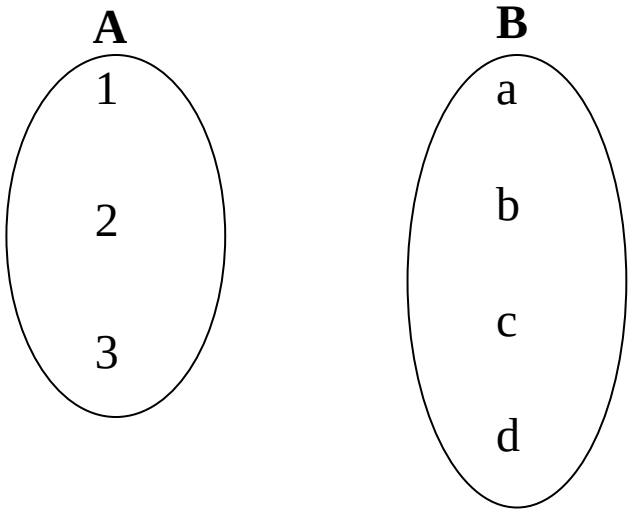
Candidate Keys

- **Candidate Keys :** Candidate keys are super keys which are not having any redundant attributes. In other words candidate keys are minimal super keys. E.g.

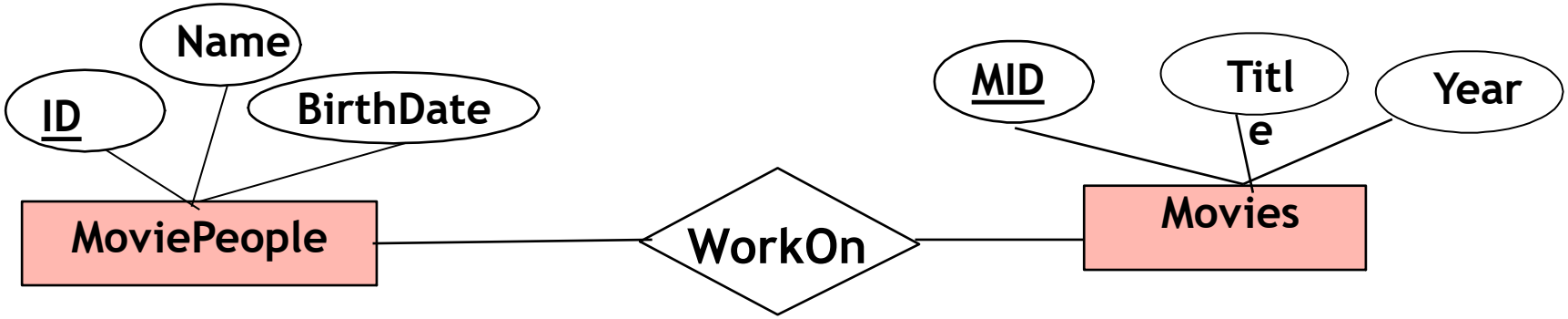
(MovieID)	CK
(MovieID, MovieName)	
(MovieID, MovieName, drname)	
(MovieID, drname)	
(Moviename, drname)	CK

Relationship

- *A mathematical definition:*
  - Let A, B be sets
  - A x B (the ***cross-product***) is the set of all pairs
  - A relationship is a subset of A x B



Relationship



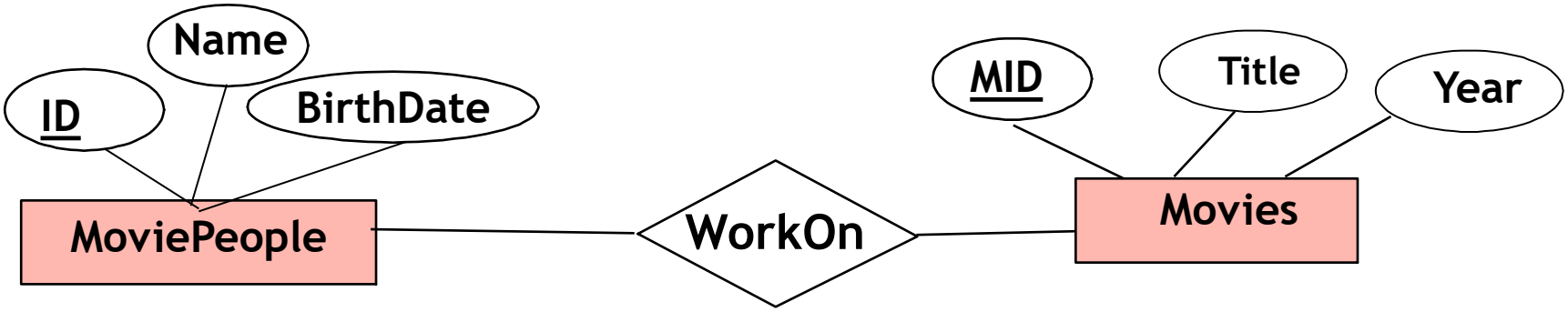
E.g., Robert Downey Jr. worked on Avengers Infinity War

Robert Downey Jr.

Image Credits:
https://goo.gl/images/3s9Htb
https://goo.gl/images/q8mejw

Relationship

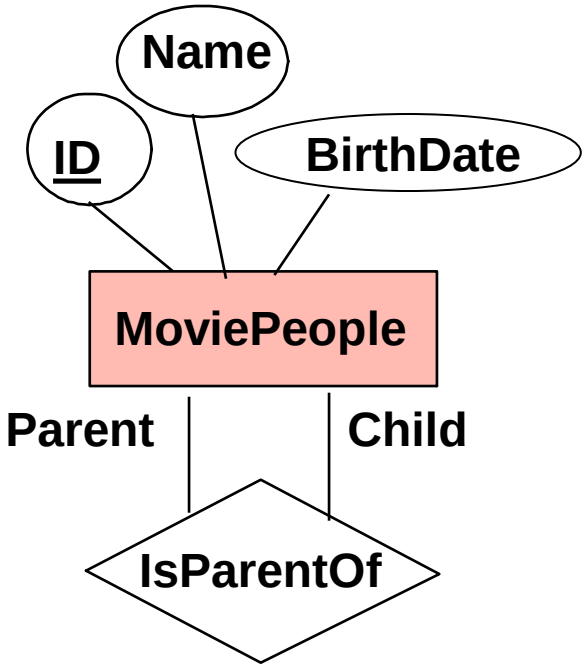
- ***Relationship Set:*** Collection of similar relationships.
  - Collection of all MoviePeople that have worked in Movies.



Self Referencing Relationship

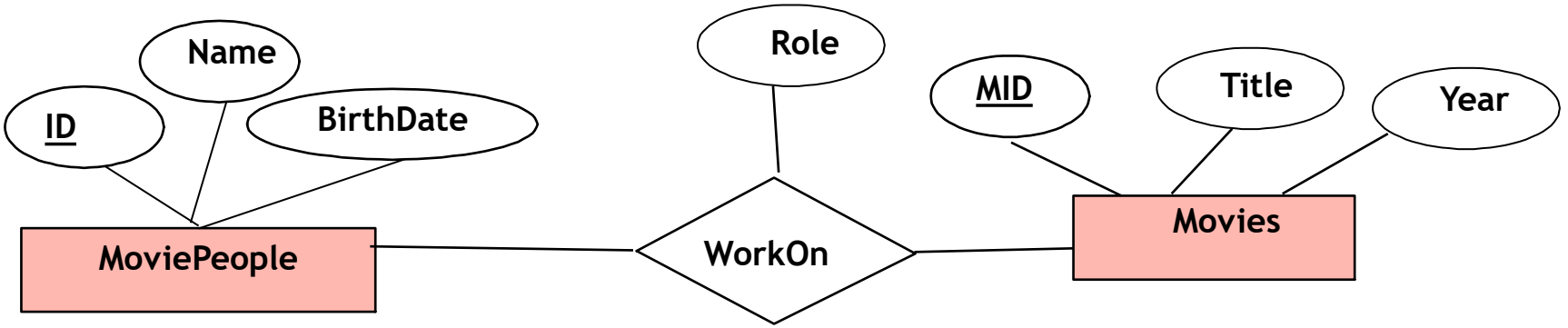
- Same entity set could participate in different relationship sets, or in different "roles" in same set.

- E.g. Will Smith isParentOf Jaden Smith



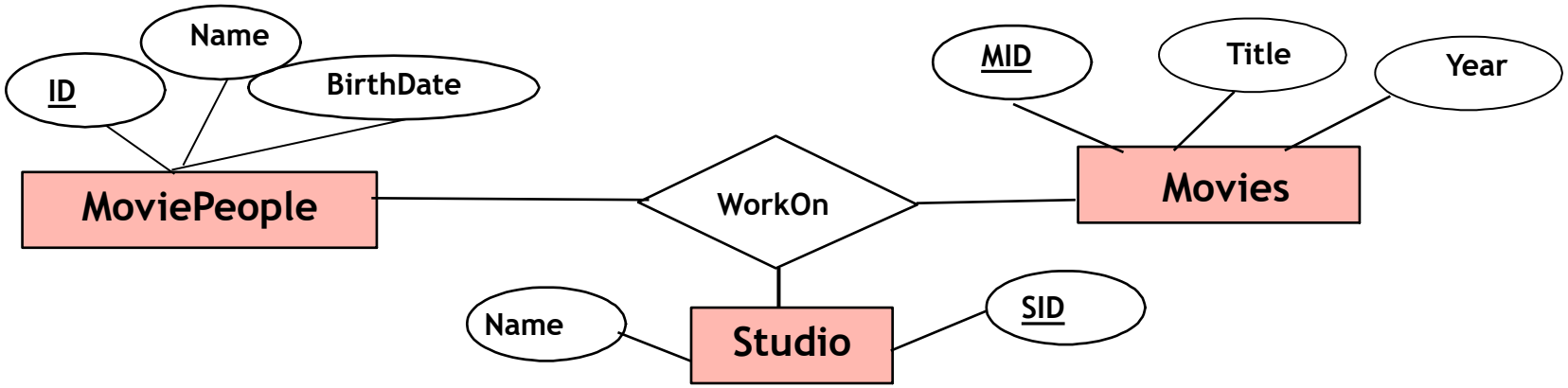
Relationship

- A relationship set may have ***descriptive attributes*** (like Role).



Relationship

- Degree of the relationship is the number of participating entity types (binary, ternary, etc.)



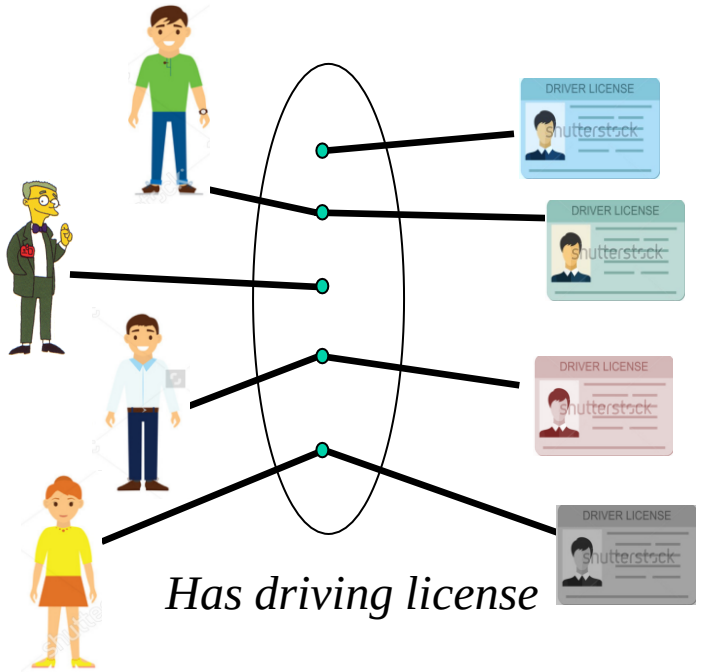
Cardinalities

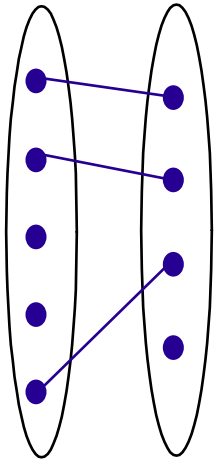
- A **cardinality ratio** for a relationship set specifies the number of relationships in the set that an entity can participate in.

- Let R be a relationship set between sets A and B. R can have 1 of 4 cardinalities:
  1. ***one-to-one*** from A to B
  2. ***one-to-many*** from A to B
  3. ***many-to-one*** *from A to B*
  4. ***many-to-many*** *from A to B*

Cardinalities

1. ***one-to-one*** from A to B:
   - an entity in A is associated with **at most one** entity in B and vice versa
   - e.g. A: driver, B: driver's license
   - Employee manages department

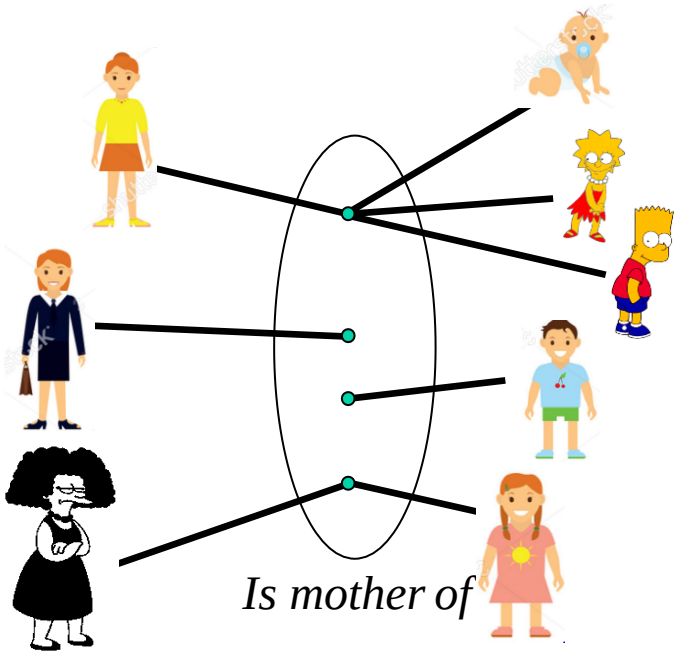


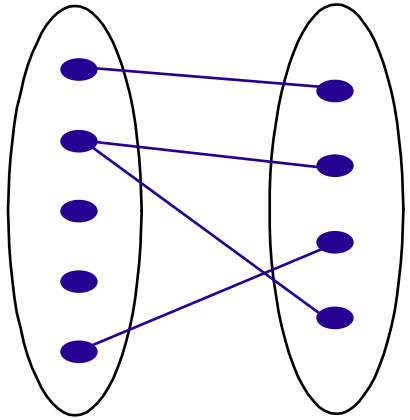


1-to-1

Cardinalities

2. ***one-to-many*** from A to B:
   - an entity in A is associated with **any number** of entities in B
   - an entity in B is associated with at most one entity in A
   - e.g. A: biological-mother, B: children
   - Employee works_for department



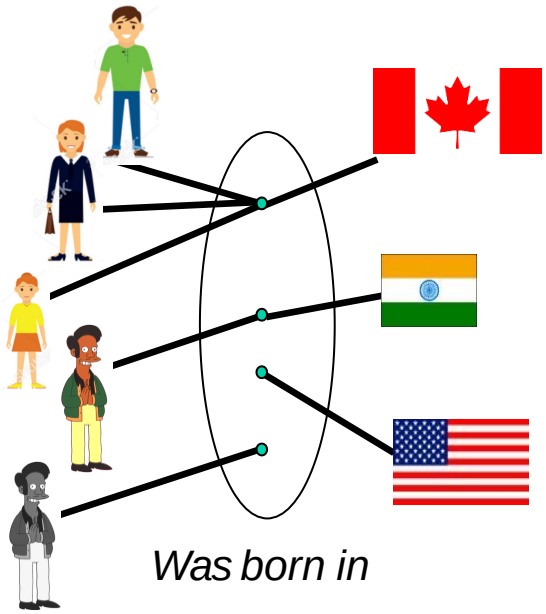


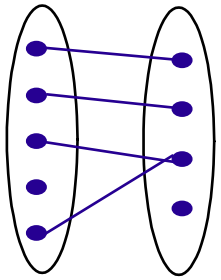
1-to Many

Cardinalities (cont')

***3. many-to-one** from A to B:*

- an entity in A is associated with at most one entity in B. An entity in B is associated with any number in A.
- Many people can be born in any county, but each person is born in at most one country.



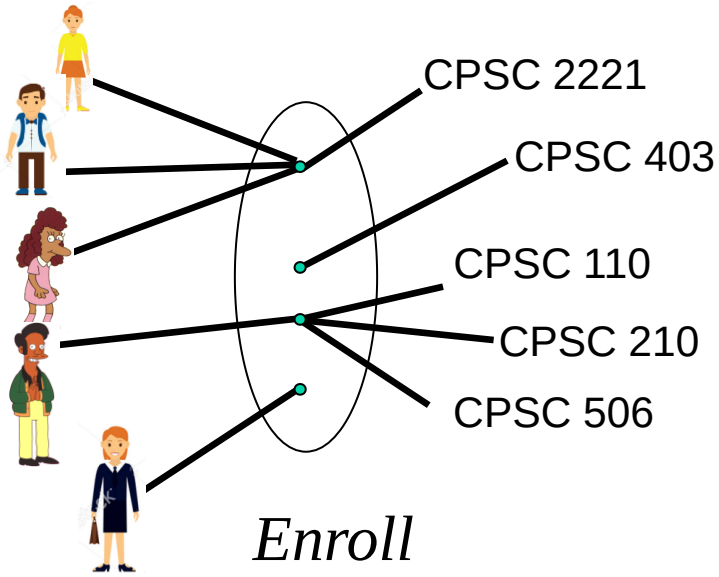


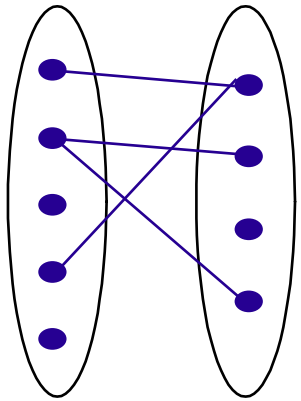
Many-to-1

Cardinalities (cont')

4. many-to-many from A to B:

- an entity in A is associated with any number of entities in B and vice versa
- e.g. A: students, B: courses
- Employee works_on project

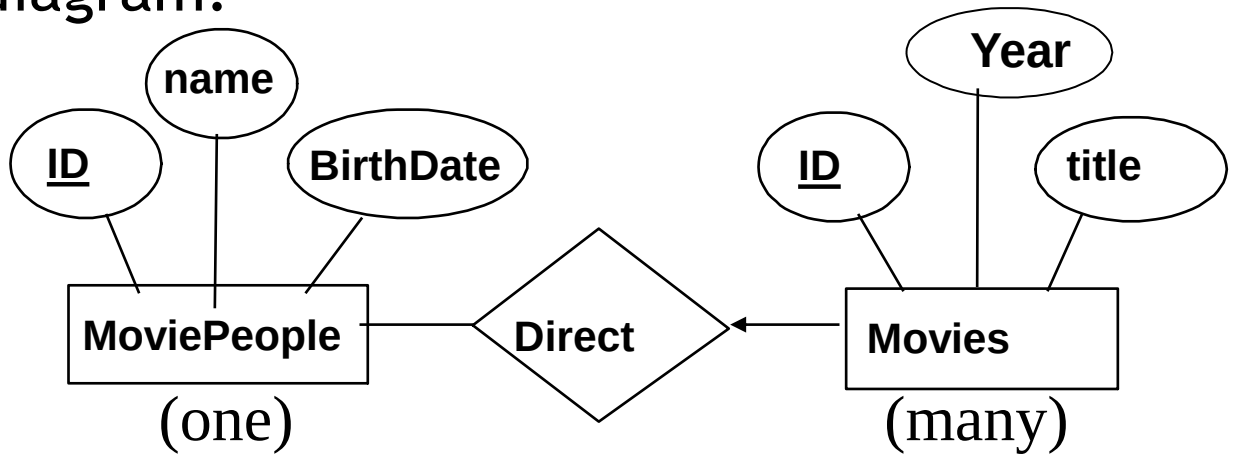


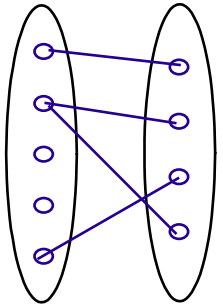


Many-to-Many

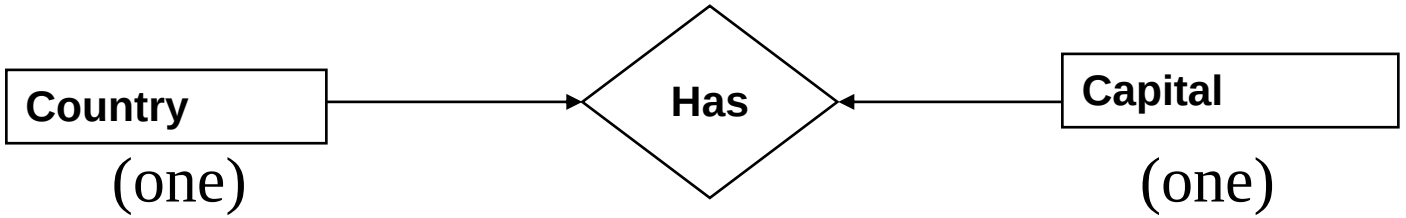
Key Constraints

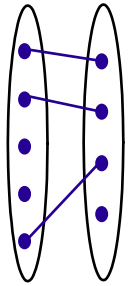
- The restriction imposed by a 1-to-1 and 1-to-many ratios are examples of ***key constraints***.
- A key constraint is shown with an arrow in the ER diagram.





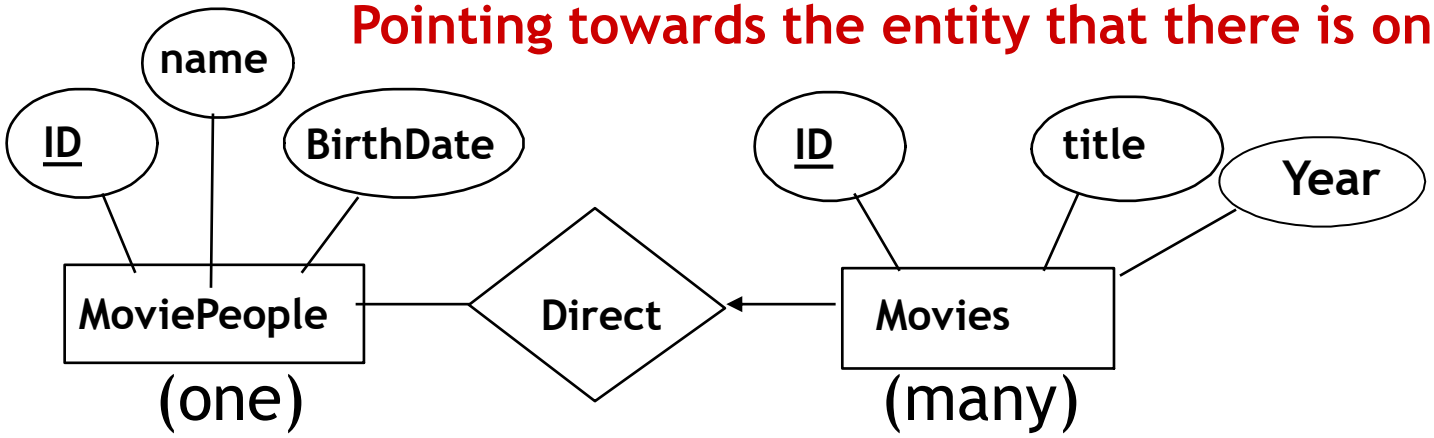
or



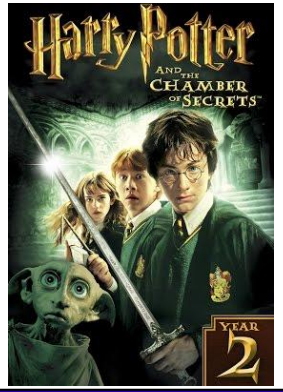


A brief detour

- The ER notation we use can be read: “if you know the entity at the tail of the arrow, then you know the relationship (and the other entities involved)”

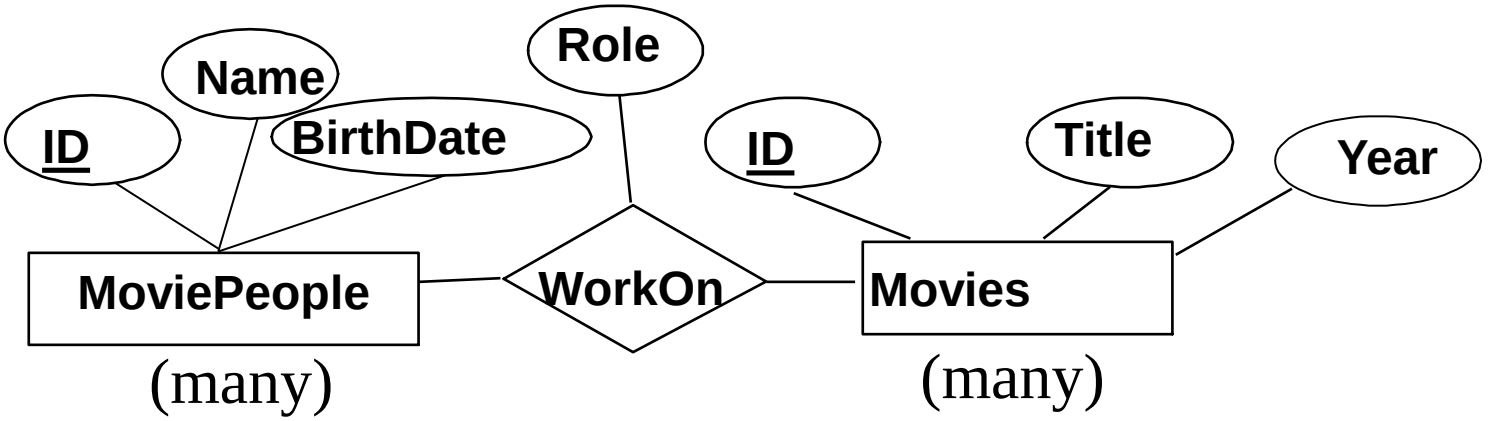


ID : MP001
Name: Chris Columbus
BirthDate: September 10, 1958



ID : M01
Title: Harry Potter and the Chamber of Secrets
Year : 2002

How can we uniquely identify a relationship (M:M)?



- How can we identify the role of a specific MoviePeople in a specific Movie?

ID : MP002
Name: Tom Hardy
BirthDate: September 15, 1977

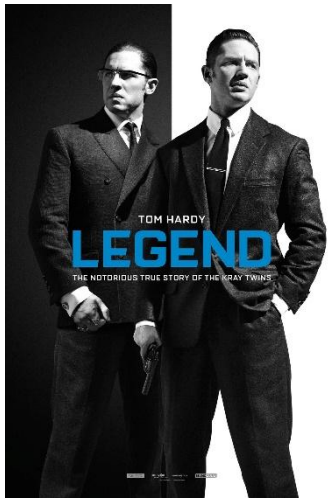
Tom Hardy
as
Bob Saginowski

ID : M02
Title: The Drop
Year : 2014

WorksOn
Need
MPID
and
MID

How can we uniquely identify a relationship (M:M)?

Tom Hardy
as
Reggie and Ronnie Cray



ID : MP002
Name: Tom Hardy
BirthDate: September 15, 1977

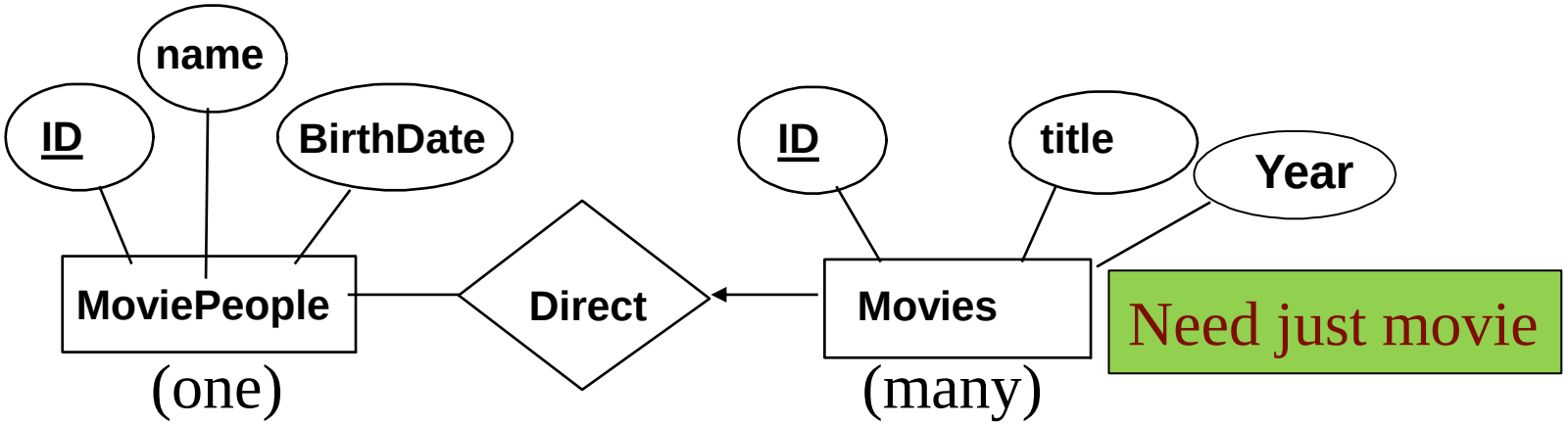
ID : M03
Title: Legend
Year : 2015

Same actor having two roles in the same movie? What do I need now?

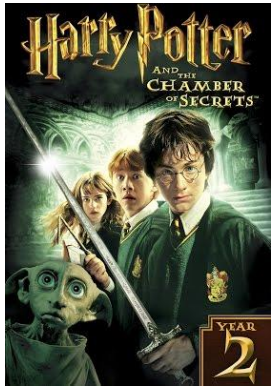
WorksOn

Need MPID, MID and Role

How can we uniquely identify a relationship (1:M)?

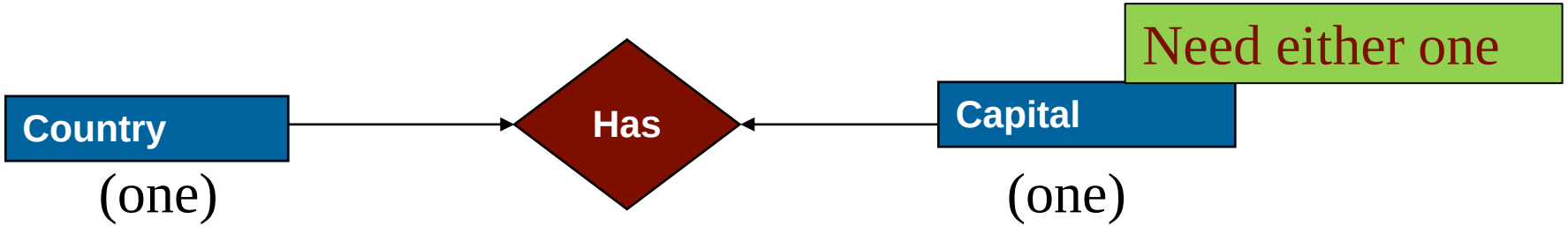


ID : MP001
Name: Chris Columbus
BirthDate: September 10, 1958



ID : M01
Title: Harry Potter and the Chamber of Secrets
Year : 2002

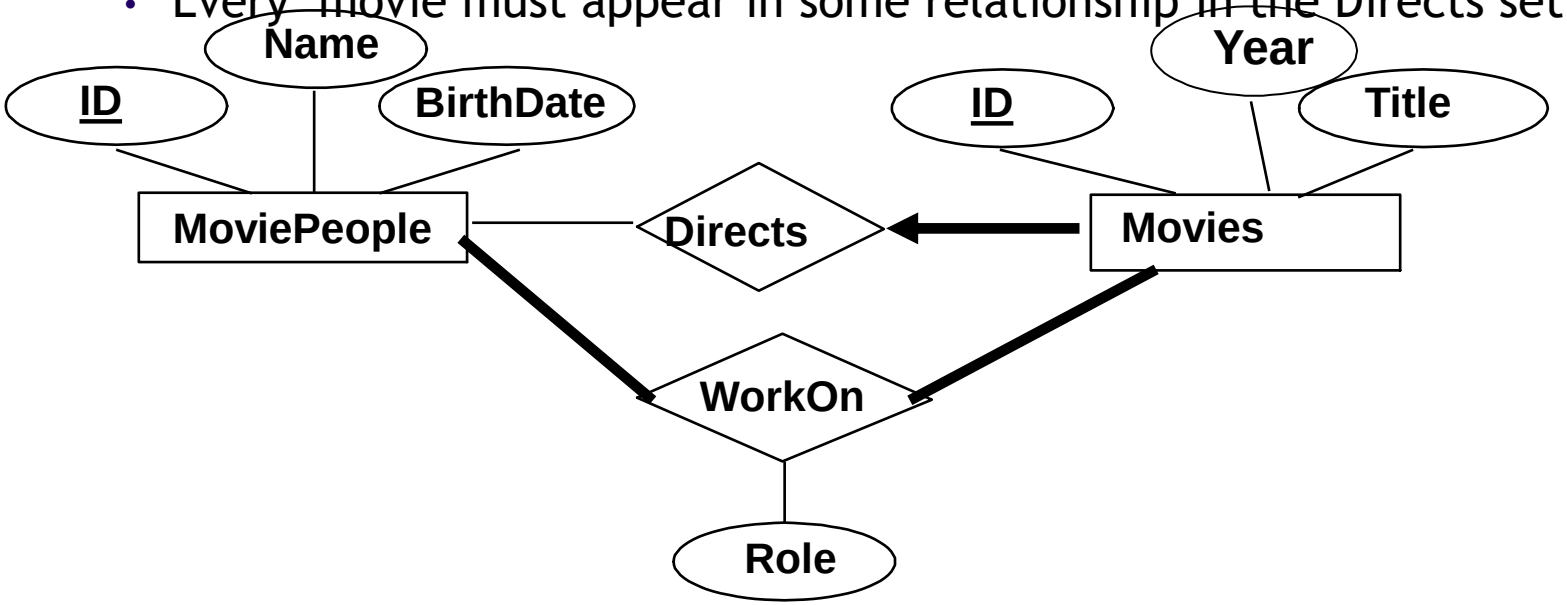
How can we uniquely identify a relationship (1:1)?



Ottawa

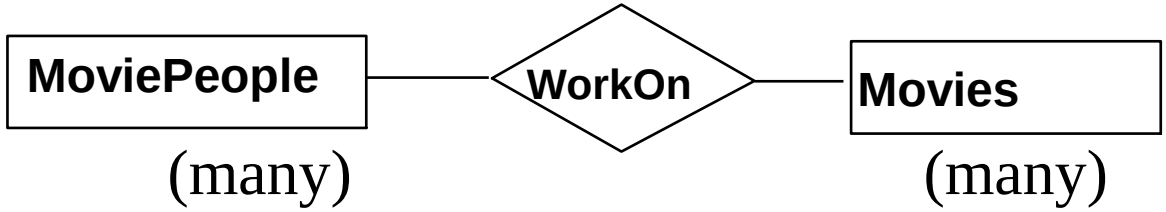
Participation Constraints

- Participation : Indicates if all entities participate in the relationship.
- An entity's participation can be ***total*** or ***partial*** .
- Requiring total participation is a ***participation constraint*** and it is shown with a thick line
  - Important on deletions
  - i.e., participation of Movie in Directs is total (thick line)
    - Every movie must appear in some relationship in the Directs set

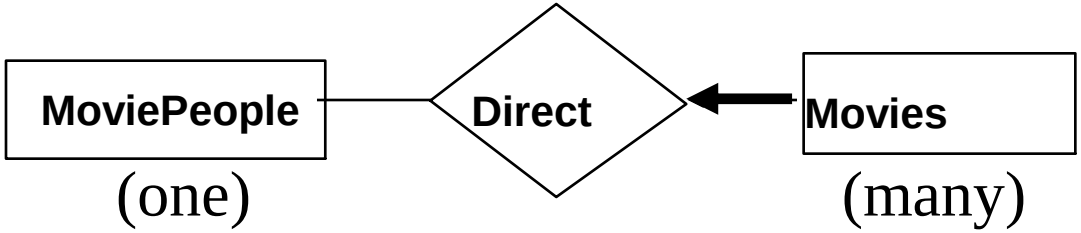


Line types summarized

- Plain lines mean many to many:



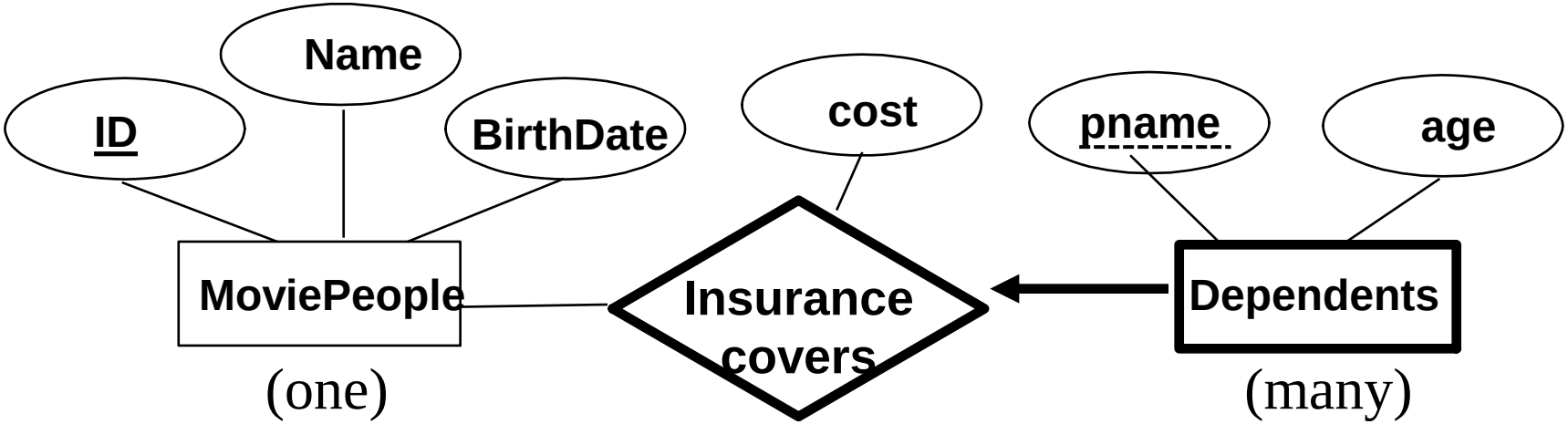
- Arrows mean the other side has a cardinality of one



- A thick line requires total participation and can be added to any line, arrow or not

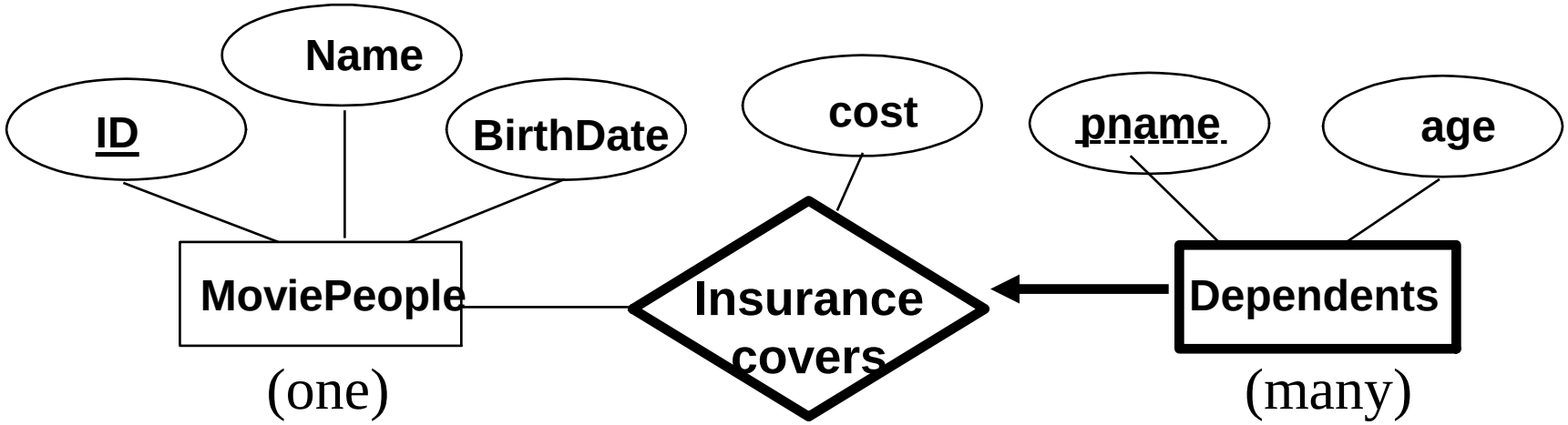
Weak Entities

- A ***weak entity*** can be identified uniquely only by considering the primary key of another (*owner*) entity.
  - Think of this as a "belongs to" relationship.



- Weak entity sets and their identifying relationship sets are shown with thick lines
- The partial identifier(------) is shown as dashed underline

Weak Entities - Constraints

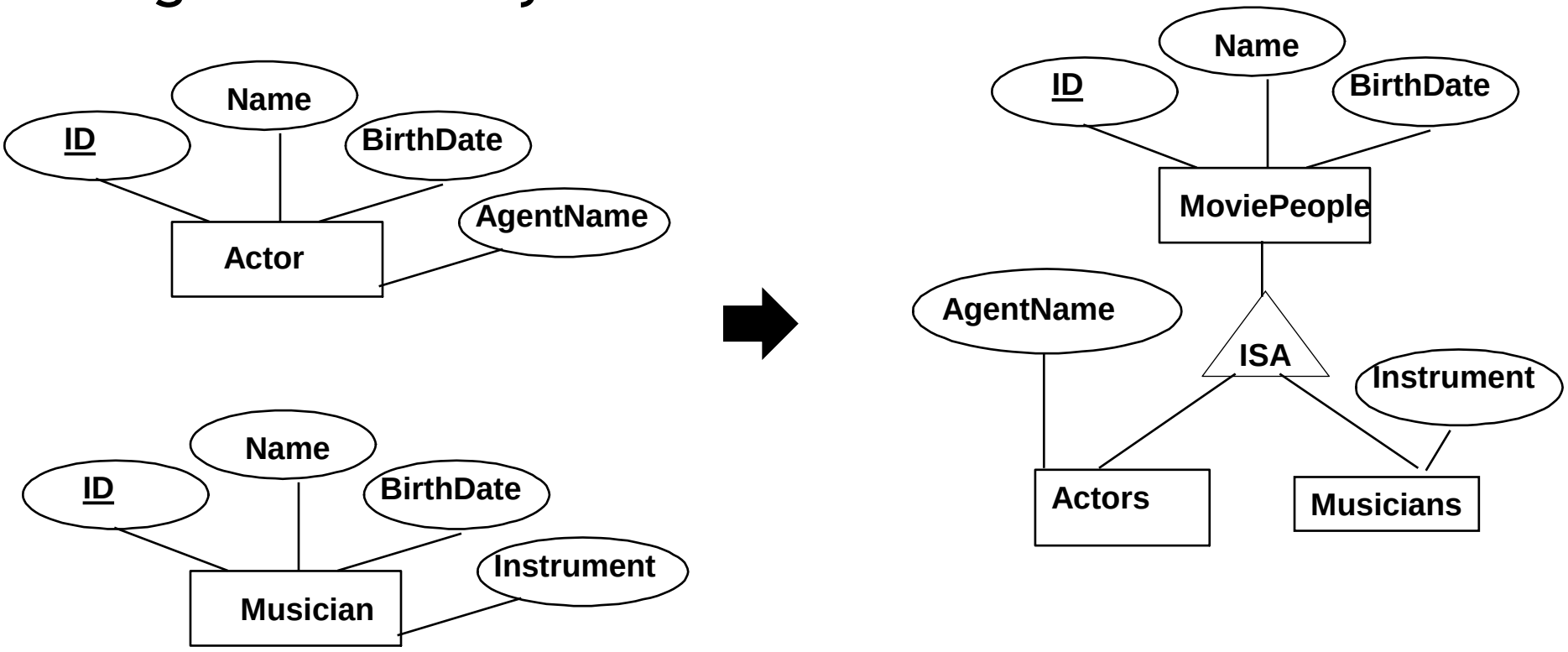


- Owner entity set and weak entity set must participate in a one-to-many or one-to-one relationship set (one owner, many/one weak entities).

- Weak entity set must have total participation in this identifying relationship set.

Class Hierarchies (ISA relationships)

- Several entity sets can be abstracted by a more general entity set



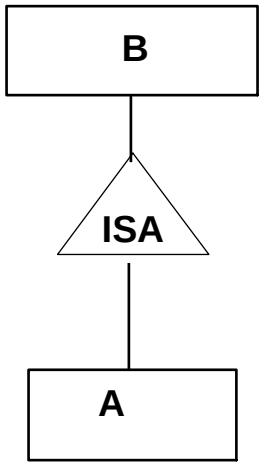
Class Hierarchies (ISA relationships)

- Sometimes, an entity set contains some entities that do share many, but not all properties with the entity set. In this case, we want to define class (entity set) hierarchies.

• A **ISA** B: every A entity is also considered to be a B entity. A ***specializes*** B, B ***generalizes*** A.

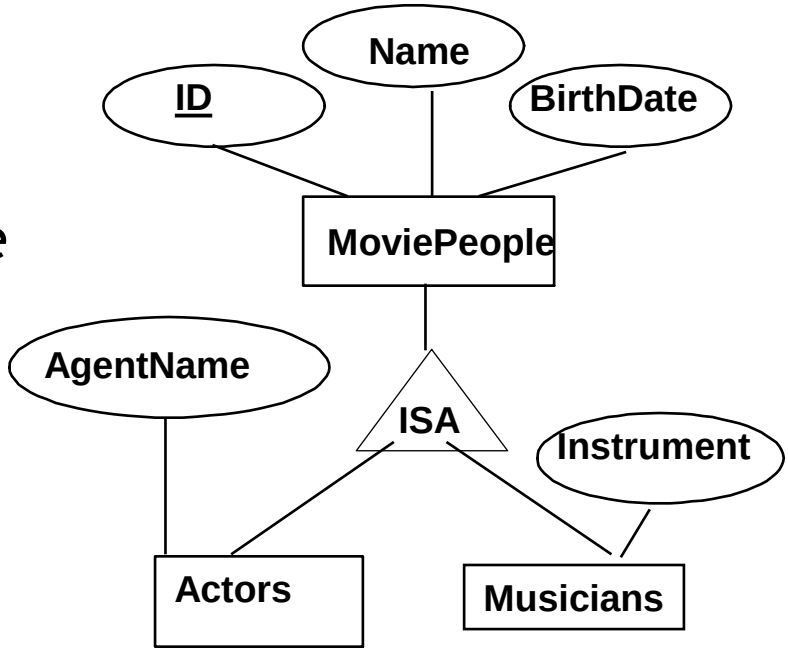
• A is called ***subclass (specialized)***, B is called ***superclass***.

• A subclass ***inherits*** the attributes of a superclass, and may define additional attributes.



ISA - Overlap Constraints

- ✓ Example : Can Mike be both actor and Musician?
- ✓ If OVERLAP is allowed this should be noted on ERD;



Actors **OVERLAPS** Musicians

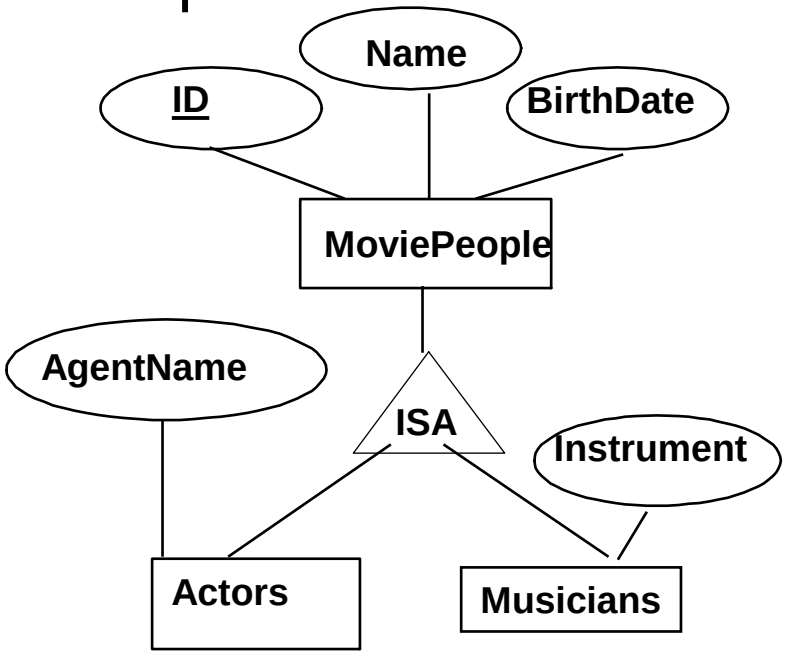
ISA - Covering Constraints

- Determine whether the entities in subclasses collectively include all entities in the superclass.

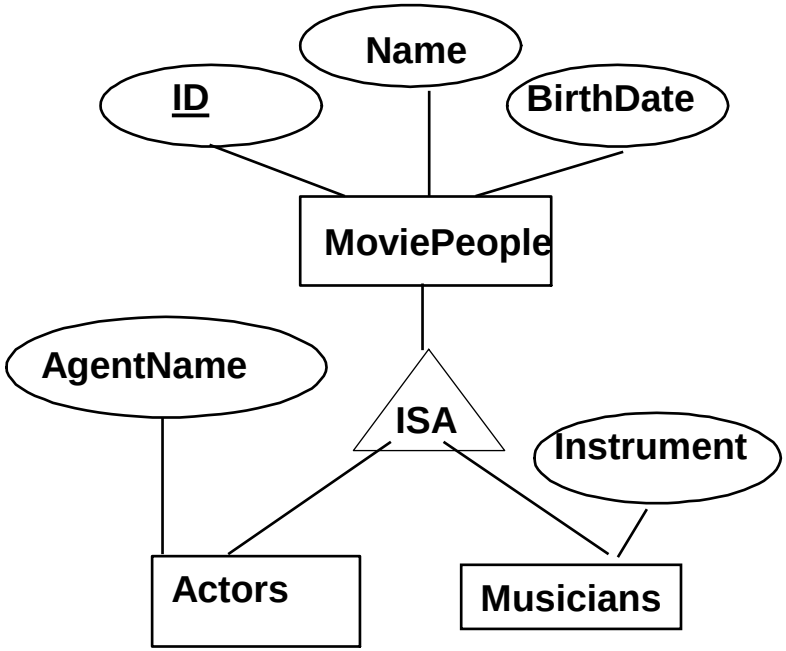
• Does every MoviePeople entity have to be either an Actor or a Musician entity?

• By default the entity set are constrained to have NO COVERING constraint.

• If COVERING is allowed this should be noted on ERD.



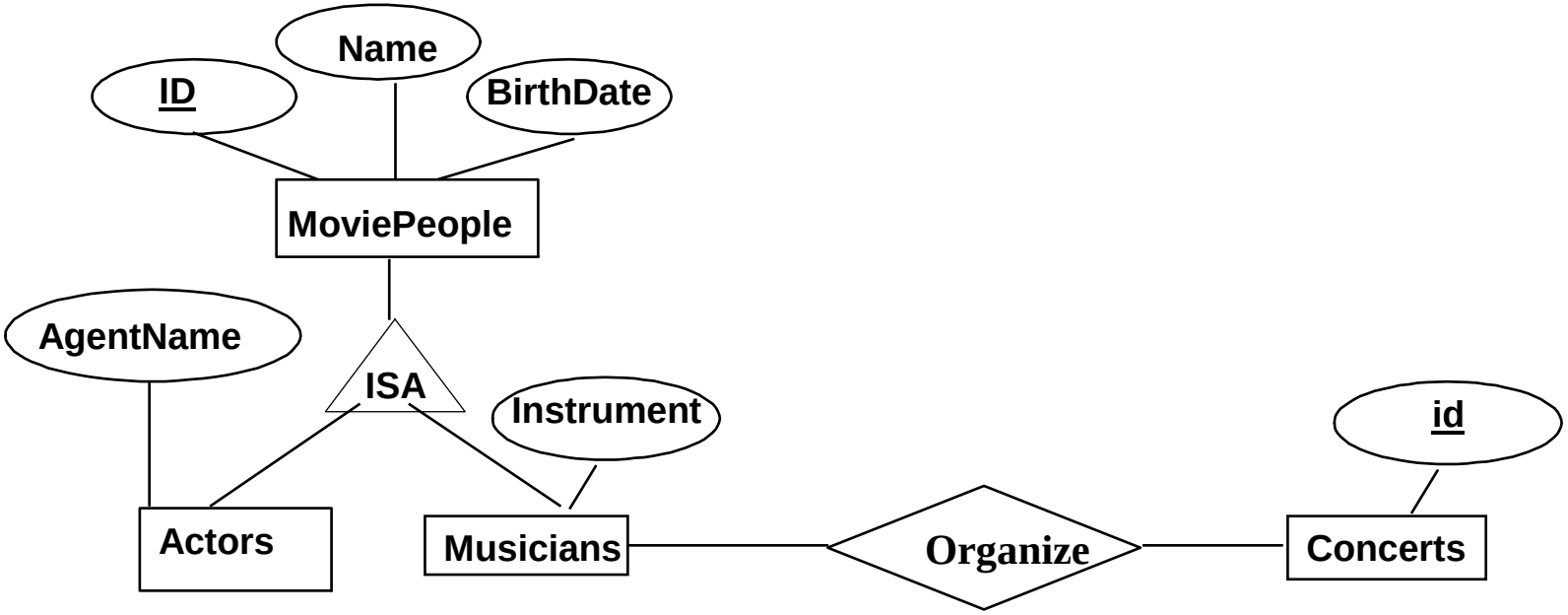
ISA - Covering Constraints



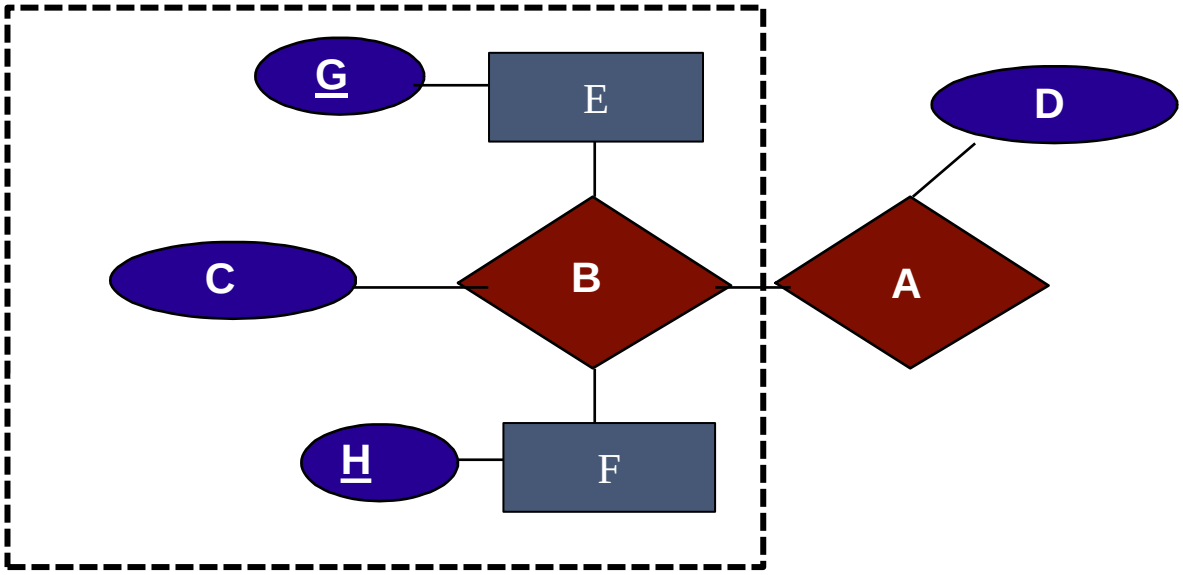
•Actor AND Musicians **COVER** MoviePeople

ISA Hierarchies

- Reasons for using ISA relationships and subclasses:
  - Do not have to redefine all the attributes.
  - Can add descriptive attributes specific to a subclass.
  - To identify entity sets that participate in a relationship set as precisely as possible.



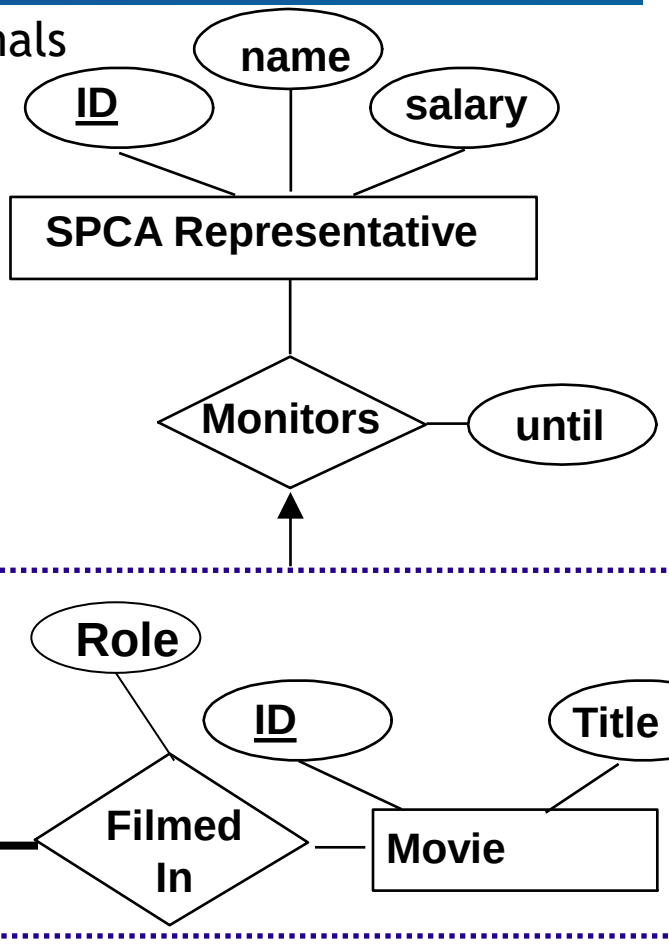
Aggregation



- Having a relationship between relationships is forbidden.
- *Aggregation* allows us to treat a relationship set as an entity set for purposes of participation in (other) relationships

Aggregation

SPCA - Society for the prevention of cruelty to animals



- Key for FilmedIn? Animal ID, movie ID
- Key for Monitors? Animal ID, movie ID

•Each sponsorship is monitored by **at most** one SPCA Representative.