Databases 2011
Lectures 04 – 05

Timothy G. Griffin

Computer Laboratory
University of Cambridge, UK

Databases, Easter 2011

Lectures 04 and 5 : Modeling with Entities and Relationships

Outline

- Entities
- Relationships
- Their relational implementations
- On the importance of SCOPE
- n-ary relationships
- Weak entity sets
- Generalization

Some real-world data ...

... from the Internet Movie Database (IMDb).

Title	Year	Actor
Austin Powers: International Man of Mystery	1997	Mike Myers
Austin Powers: The Spy Who Shagged Me	1999	Mike Myers
Dude, Where's My Car?	2000	Bill Chott
Dude, Where's My Car?	2000	Marc Lynn

Entities diagrams and Relational Schema

These diagrams represent relational schema

Movie(*MovieID*, *Title*, *Year*)

Person(*PersonID*, *FirstName*, *LastName*)

Yes, this ignores types ...

Entity sets (relational instances)

Movie

MovieID	**Title**	**Year**
55871	Austin Powers: International Man of Mystery	1997
55873	Austin Powers: The Spy Who Shagged Me	1999
171771	Dude, Where's My Car?	2000

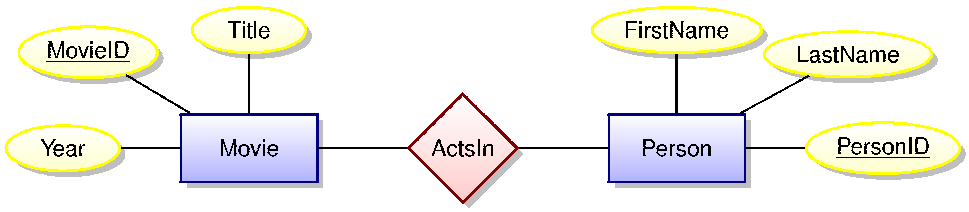
(I used line number from IMDb raw file movies.list as MovieID.)

Person

PersonID	**FirstName**	**LastName**
6902836	Mike	Myers
1757556	Bill	Chott
5882058	Marc	Lynn

(I used line number from IMDb raw file actors.list as PersonID)

Relationships



Foreign Keys and Referential Integrity

Foreign Key

Suppose we have $R(\underline{\mathbf{Z}}, \mathbf{Y})$. Furthermore, let $S(\mathbf{W})$ be a relational schema with $\mathbf{Z} \subseteq \mathbf{W}$. We say that $\mathbf{Z}$ represents a Foreign Key in S for R if for any instance we have $\pi_{\mathbf{Z}}(S) \subseteq \pi_{\mathbf{Z}}(R)$. This is a semantic assertion.

Referential integrity

A database is said to have referential integrity when all foreign key constraints are satisfied.

A relational representation

A relational schema

$$\mathit{ActsIn}(\underline{\mathit{MovieID}}, \underline{\mathit{PersonID}})$$

With referential integrity constraints

$$\pi_{\mathit{MovieID}}(\mathit{ActsIn}) \subseteq \pi_{\mathit{MovieID}}(\mathit{Movie})$$

$$\pi_{\mathit{PersonID}}(\mathit{ActsIn}) \subseteq \pi_{\mathit{PersonID}}(\mathit{Person})$$

ActsIn

PersonID	**MovieID**
6902836	55871
6902836	55873
1757556	171771
5882058	171771

Foreign Keys in SQL

```
create table ActsIn
(  MovieID int not NULL,
   PersonID int not NULL,
   primary key (MovieID, PersonID),
   constraint actsin_movie
         foreign key (MovieID)
         references Movie(MovieID),
   constraint actsin_person
         foreign key (PersonID)
         references Person(PersonID))
```

Yes, we could do a better job on the date type ...

Relational representation of relationships, in general?

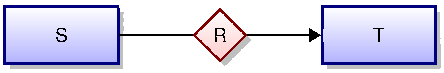
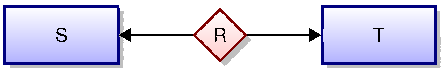
That depends ...

Mapping Cardinalities for binary relations, $R \subseteq S \times T$

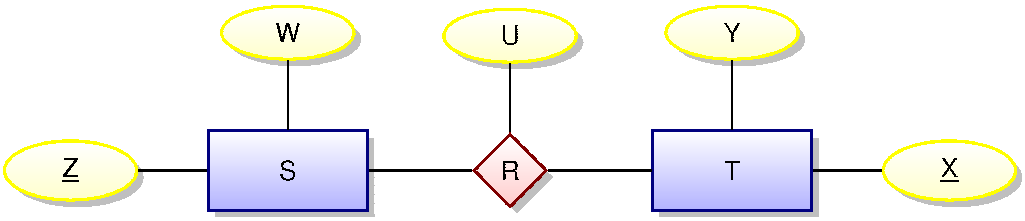
Relation R is	meaning
many to many	no constraints
one to many	$\forall t \in T, s_1, s_2 \in S.(R(s_1, t) \wedge R(s_2, t)) \implies s_1 = s_2$
many to one	$\forall s \in S, t_1, t_2 \in T.(R(s, t_1) \wedge R(s, t_2)) \implies t_1 = t_2$
one to one	one to many and many to one

Note that the database terminology differs slightly from standard mathematical terminology.

Diagrams for Mapping Cardinalities

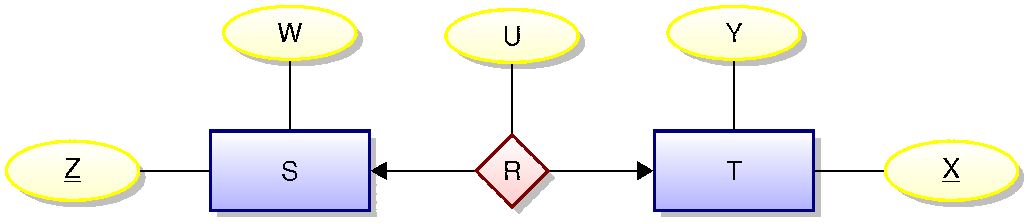
ER diagram	Relation R is
	many to many ($M : N$)
	one to many ($1 : M$)
	many to one ($M : 1$)
	one to one ($1 : 1$)

Relationships to Relational Schema



Relation R is	Schema
many to many ($M : N$)	$R(\underline{X}, \underline{Z}, U)$
one to many ($1 : M$)	$R(X, \underline{Z}, U)$
many to one ($M : 1$)	$R(\underline{X}, Z, U)$
one to one ($1 : 1$)	$R(\underline{X}, Z, U)$ or $R(X, \underline{Z}, U)$

“one to one” does not mean a "1-to-1 correspondence”



This database instance is OK

S

Z	W
z_1	w_1
z_2	w_2
z_3	w_3

R

Z	X	U
z_1	$x_2 1$	u_1

T

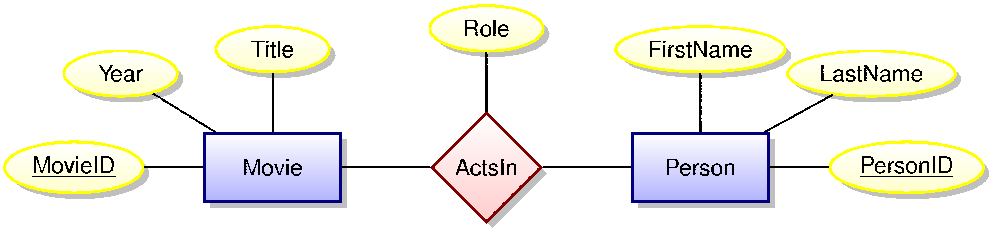
X	Y
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4

Some more real-world data ... (a slight change of SCOPE)

Title	Year	Actor	Role
Austin Powers: International Man of Mystery	1997	Mike Myers	Austin Powers
Austin Powers: International Man of Mystery	1997	Mike Myers	Dr. Evil
Austin Powers: The Spy Who Shagged Me	1999	Mike Myers	Austin Powers
Austin Powers: The Spy Who Shagged Me	1999	Mike Myers	Dr. Evil
Austin Powers: The Spy Who Shagged Me	1999	Mike Myers	Fat Bastard
Dude, Where's My Car?	2000	Bill Chott	Big Cult Guard 1
Dude, Where's My Car?	2000	Marc Lynn	Cop with Whips

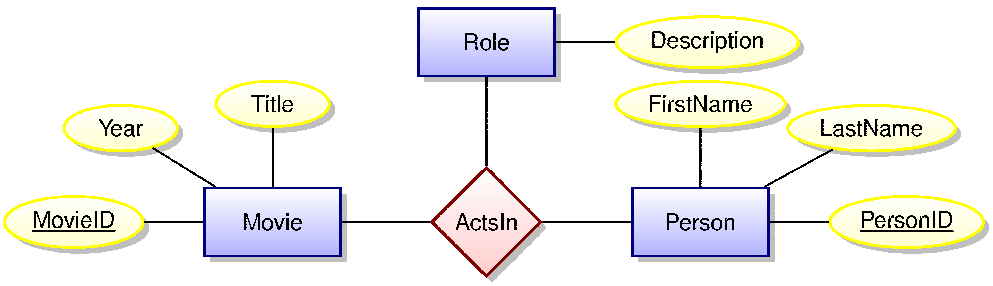
How will this change our model?

Will **ActsIn** remain a binary Relationship?



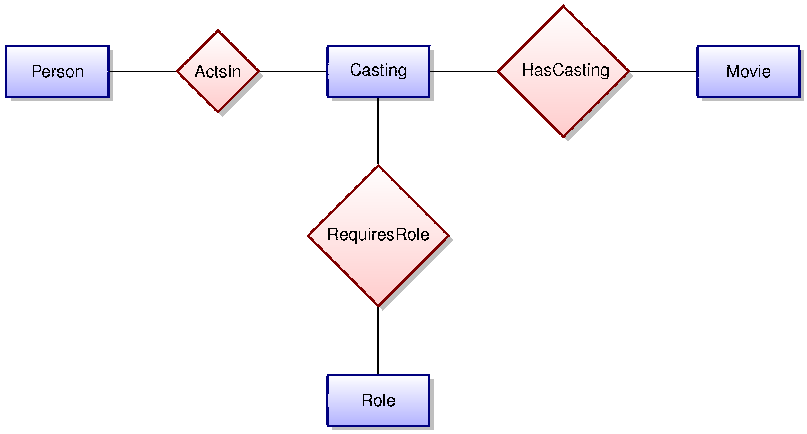
No! An actor can have many roles in the same movie!

Could **ActsIn** be modeled as a Ternary Relationship?



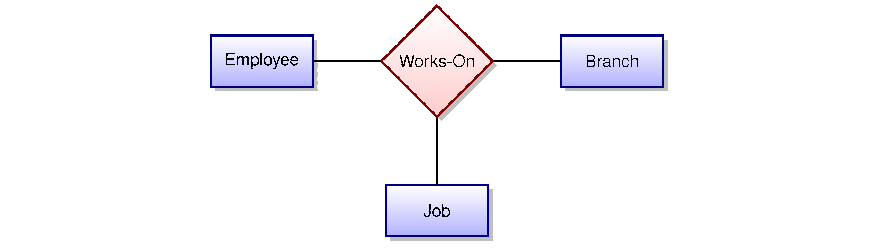
Yes, this works!

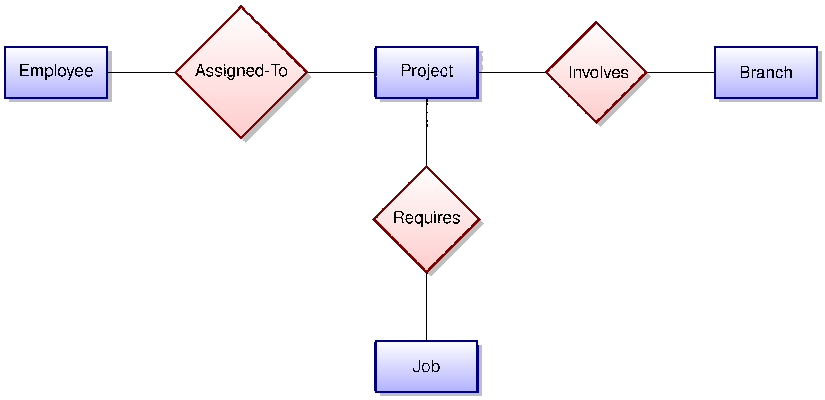
Can a ternary relationship be modeled with multiple binary relationships?



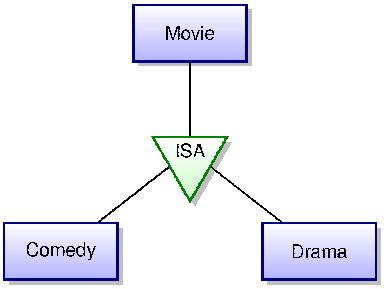
The **Casting** entity seems artificial. What attributes would it have?

Sometimes ternary to multiple binary makes more sense ...





Generalization



Questions

- Is every movie either comedy or a drama?
- Can a movie be a comedy and a drama?

But perhaps this isn't a good model ...

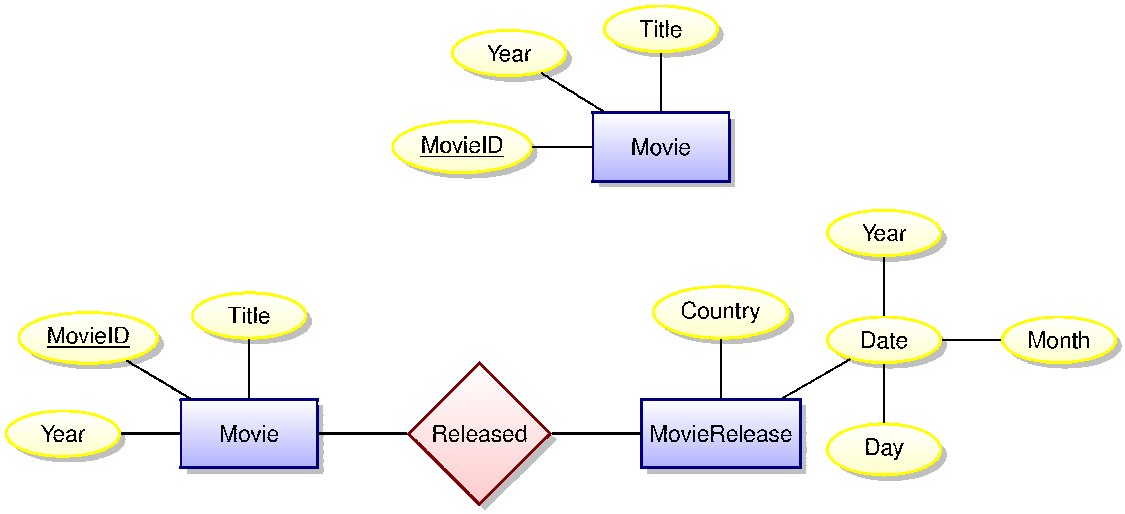
- What attributes would distinguish Drama and Comedy entities?
- What abound **Science Fiction**?
- Perhaps **Genre** would make a nice entity, which could have a relationship with **Movie**.

Another change of SCOPE ...

Movies with detailed release dates

Title	Country	Day	Month	Year
Austin Powers: International Man of Mystery	USA	02	05	1997
Austin Powers: International Man of Mystery	Iceland	24	10	1997
Austin Powers: International Man of Mystery	UK	05	09	1997
Austin Powers: International Man of Mystery	Brazil	13	02	1998
Austin Powers: The Spy Who Shagged Me	USA	08	06	1999
Austin Powers: The Spy Who Shagged Me	Iceland	02	07	1999
Austin Powers: The Spy Who Shagged Me	UK	30	07	1999
Austin Powers: The Spy Who Shagged Me	Brazil	08	10	1999
Dude, Where's My Car?	USA	10	12	2000
Dude, Where's My Car?	Iceland	9	02	2001
Dude, Where's My Car?	UK	9	02	2001
Dude, Where's My Car?	Brazil	9	03	2001
Dude, Where's My Car?	Russia	18	09	2001

... and an attribute becomes an entity with a connecting relation.



Is this really a good model?

- What is (natural) key of MovieRelease?
- Would a ternary relationship be better?

Question: What is the right model?

Answer: The question doesn't make sense!

- There is no "right" model ...
- It depends on the intended use of the database.
- What activity will the DBMS support?
- What data is needed to support that activity?

The issue of SCOPE is missing from most textbooks

- **Suppose** that all databases begin life with beautifully designed schemas.
- **Observe** that many operational databases are in a sorry state.
- **Conclude** that the scope and goals of a database continually change, and that schema evolution is a difficult problem to solve, in practice.