Designing Database

CSC 365

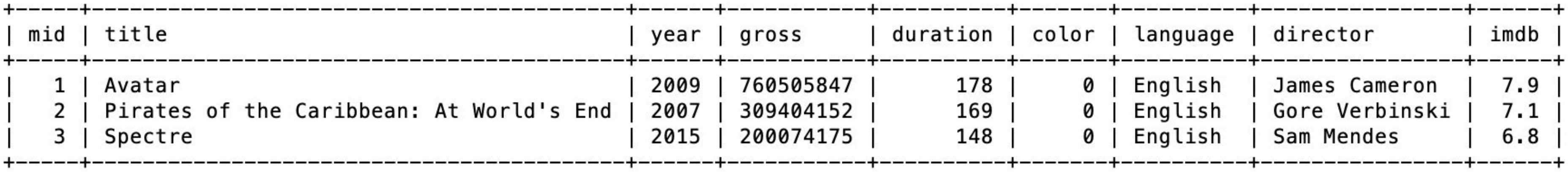
Let me introduce some terminologies and concepts first.

Functional Dependency (FD)

- Functional Dependency (FD)
  - $A_1, A_2, \ldots, A_n \longrightarrow B_1, B_2, \ldots, B_m$
  - Attributes $A_1, A_2, \ldots, A_n$ functionally determine attributes $B_1, B_2, \ldots, B_m$

Functional Dependency (FD)

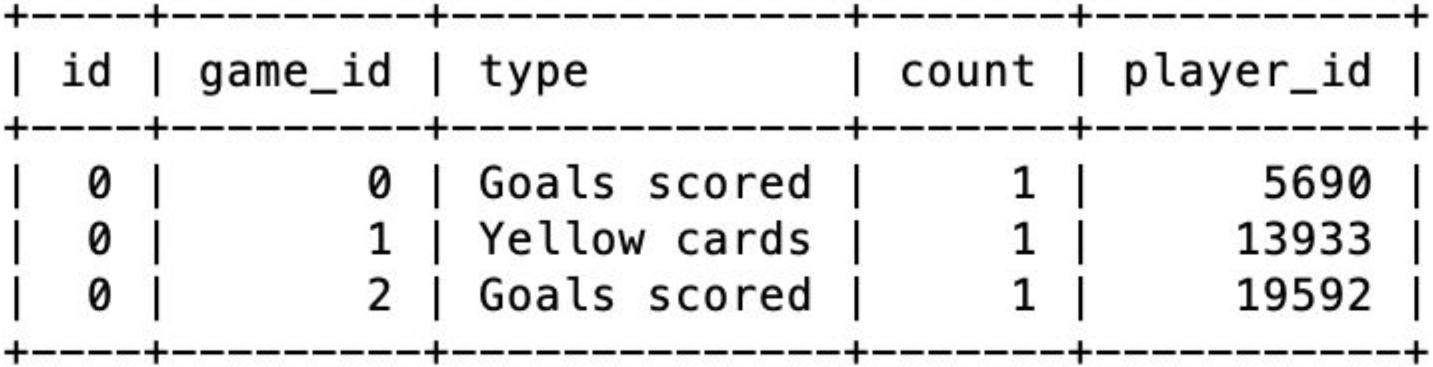
- Functional Dependency (FD)
  - $A_1, A_2, \ldots, A_n \longrightarrow B_1, B_2, \ldots, B_m$
  - Attributes $A_1, A_2, \ldots, A_n$ functionally determine attributes $B_1, B_2, \ldots, B_m$
  - FD identifies a key
  -
- mid $\longrightarrow$ title, year, gross, duration, color, language, director, imdb



```
+-----+------------------------------------------+------+-----------+----------+-------+----------+----------------+------+
| mid | title                                    | year | gross     | duration | color | language | director       | imdb |
+-----+------------------------------------------+------+-----------+----------+-------+----------+----------------+------+
|   1 | Avatar                                   | 2009 | 760505847 |      178 |     0 | English  | James Cameron  |  7.9 |
|   2 | Pirates of the Caribbean: At World's End | 2007 | 309404152 |      169 |     0 | English  | Gore Verbinski |  7.1 |
|   3 | Spectre                                  | 2015 | 200074175 |      148 |     0 | English  | Sam Mendes     |  6.8 |
+-----+------------------------------------------+------+-----------+----------+-------+----------+----------------+------+
```

Functional Dependency (FD)

- Functional Dependency (FD)
  - $A_1, A_2, \ldots, A_n \longrightarrow B_1, B_2, \ldots, B_m$
  - Attributes $A_1, A_2, \ldots, A_n$ functionally determine attributes $B_1, B_2, \ldots, B_m$
  - FD identifies a key
  -
- id, game_id → type, count, player_id



```
+----+---------+---------------+-------+-----------+
| id | game_id | type          | count | player_id |
+----+---------+---------------+-------+-----------+
|  0 |       0 | Goals scored  |     1 |      5690 |
|  0 |       1 | Yellow cards  |     1 |     13933 |
|  0 |       2 | Goals scored  |     1 |     19592 |
+----+---------+---------------+-------+-----------+
```

Trivial Functional Dependency

- FD is trivial if it has a right side that is a subset of its left side.
- $A_1, A_2, \ldots, A_n \longrightarrow B_1, B_2, \ldots, B_m$ such that

$$\{B_1, B_2, \ldots, B_m\} \subseteq \{A_1, A_2, \ldots, A_n\}$$

Examples:

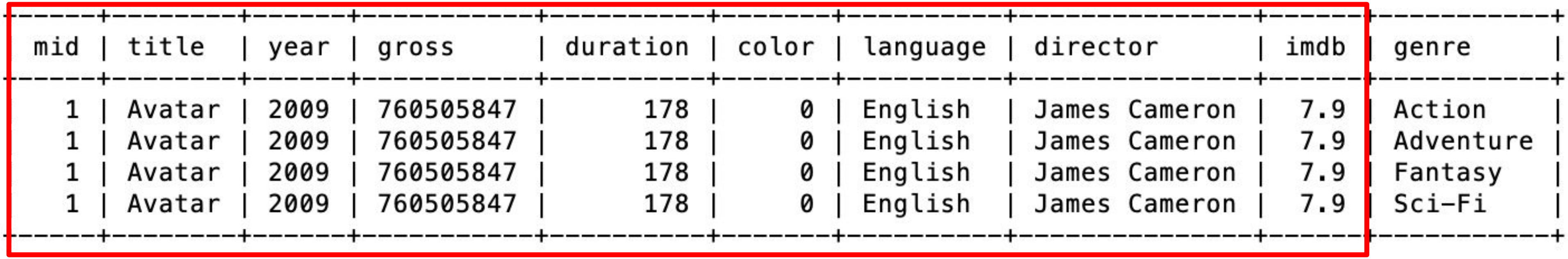
- mid ⟶ mid
- id, game_id ⟶ game_id

Anomalies

- Redundancy
  - Information is repeated unnecessarily in several tuples.
- Update Anomalies
  - Information is not updated in all tuples that contain the same value.
- Deletion Anomalies
  - Other information is lost as a side effect when a tuple is deleted.

Anomalies Example

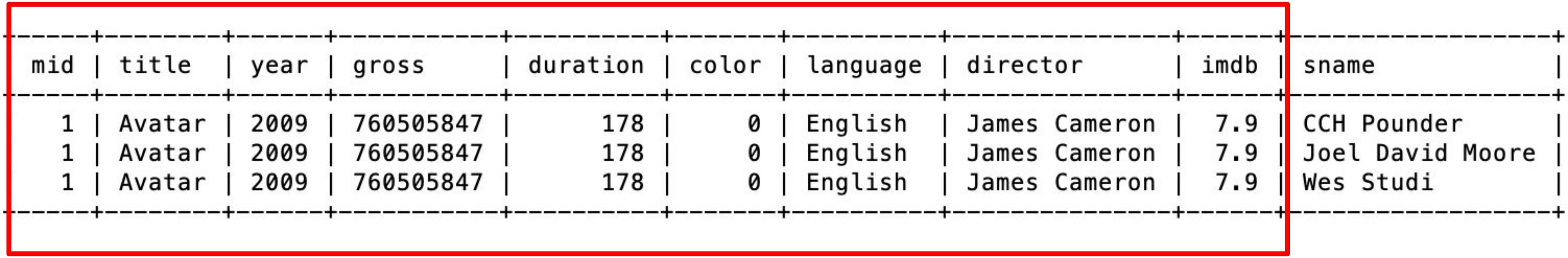
- Redundancy
  - Information is repeated unnecessarily in several tuples.
- Update Anomalies
  - Information is not updated in all tuples that contain the same value.
- Deletion Anomalies
  - Other information is lost as a side effect when a tuple is deleted.



```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+-----------+
| mid | title  | year | gross     | duration | color | language | director      | imdb | genre     |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+-----------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Action    |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Adventure |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Fantasy   |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Sci-Fi    |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+-----------+
```

Anomalies Example

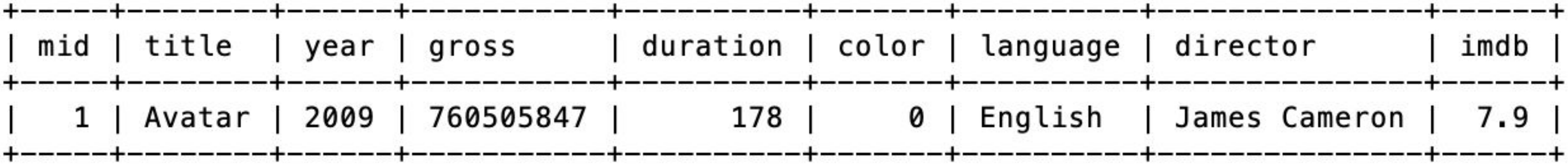
- Redundancy
  - Information is repeated unnecessarily in several tuples.
- Update Anomalies
  - Information is not updated in all tuples that contain the same value.
- Deletion Anomalies
  - Other information is lost as a side effect when a tuple is deleted.



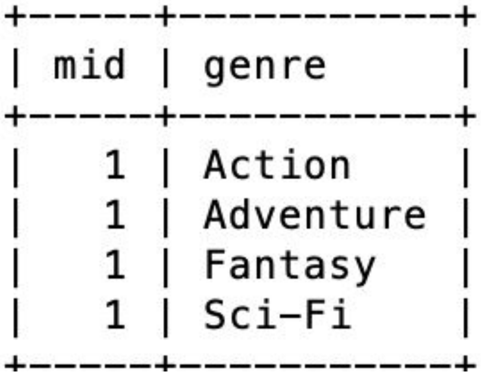
```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+------------------+
| mid | title  | year | gross     | duration | color | language | director      | imdb | sname            |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+------------------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | CCH Pounder      |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Joel David Moore |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Wes Studi        |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+------------------+
```

Solution to the Anomalies

Decompose the Relation because too much information is crammed in it.



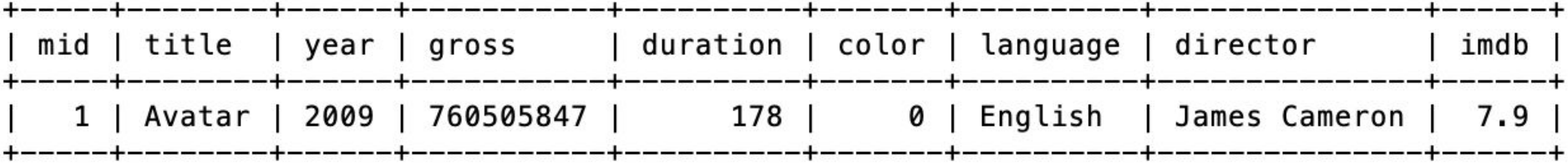
```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
| mid | title  | year | gross     | duration | color | language | director      | imdb |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
```



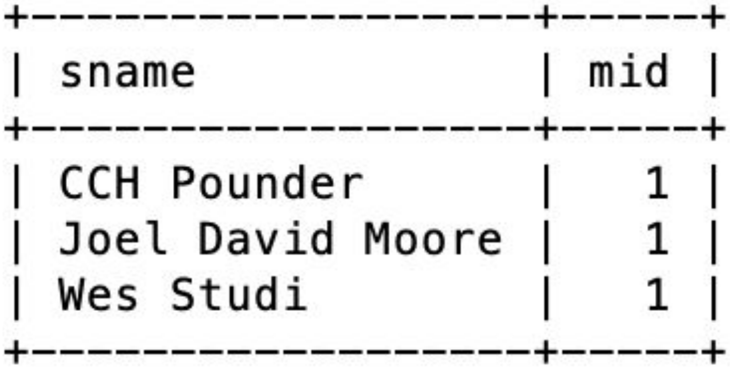
```
+-----+-----------+
| mid | genre     |
+-----+-----------+
|   1 | Action    |
|   1 | Adventure |
|   1 | Fantasy   |
|   1 | Sci-Fi    |
+-----+-----------+
```

Solution to the Anomalies

Decompose the Relation because too much information is crammed in it.



```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
| mid | title  | year | gross     | duration | color | language | director      | imdb |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
```



```
+------------------+-----+
| sname            | mid |
+------------------+-----+
| CCH Pounder      |   1 |
| Joel David Moore |   1 |
| Wes Studi        |   1 |
+------------------+-----+
```

Decomposition

- Objective
  - Elimination of Anomalies
- Properties to be achieved:
  - Recoverability of Information
    - Can we recover the original relation by JOIN?
  - Preservation of Dependencies
    - Can we preserve original FD in some individual relations resulting after decomposition?

Decomposition

- Objective
  - Elimination of Anomalies
- Properties to be achieved:
  - Recoverability of Information
    - Can we recover the original relation by JOIN?
  - Preservation of Dependencies
    - Can we preserve original FD in some individual relations resulting after decomposition?

It is difficult to achieve both properties always.

This is more important

Schema Design Principles #1

Each component in each tuple needs to be ***atomic***

- No structure inside an attribute
- Domain must be atomic

Schema Design Principles #1

Each component in each tuple needs to be ***atomic***

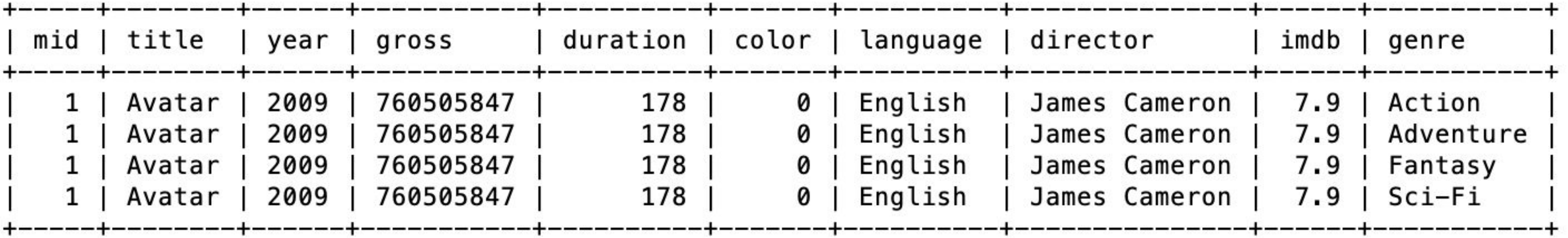
mid	title	year	...	genre
1	Avatar	2009		Action/Adventure/Fantasy/Sci-Fi

Schema Design Principles #1

Each component in each tuple needs to be ***atomic***

mid	title	year	...	genre
1	Avatar	2009		Action/Adventure/Fantasy/Sci-Fi

Normalization ⇩ First Normal Form



```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+-----------+
| mid | title  | year | gross     | duration | color | language | director      | imdb | genre     |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+-----------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Action    |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Adventure |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Fantasy   |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Sci-Fi    |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+-----------+
```

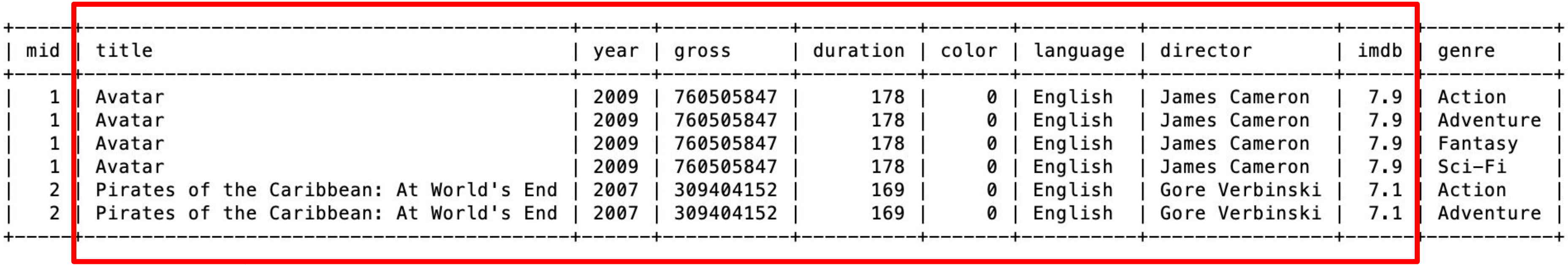
Schema Design Principles #2

No partial dependencies

Schema Design Principles #2

No partial dependencies

FD = mid, genre ? because they seem to uniquely identify each tuple

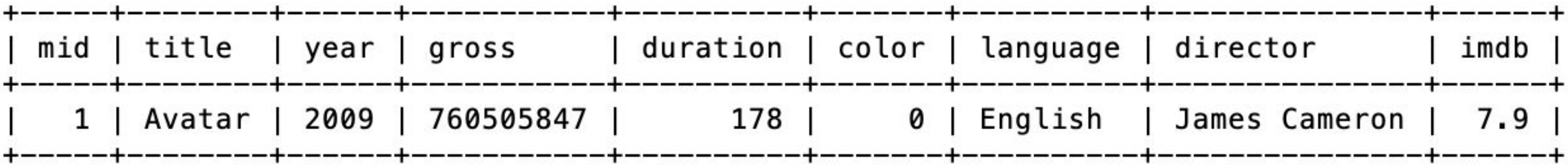


```
+-----+------------------------------------------+------+-----------+----------+-------+----------+----------------+------+-----------+
| mid | title                                    | year | gross     | duration | color | language | director       | imdb | genre     |
+-----+------------------------------------------+------+-----------+----------+-------+----------+----------------+------+-----------+
|   1 | Avatar                                   | 2009 | 760505847 |      178 |     0 | English  | James Cameron  |  7.9 | Action    |
|   1 | Avatar                                   | 2009 | 760505847 |      178 |     0 | English  | James Cameron  |  7.9 | Adventure |
|   1 | Avatar                                   | 2009 | 760505847 |      178 |     0 | English  | James Cameron  |  7.9 | Fantasy   |
|   1 | Avatar                                   | 2009 | 760505847 |      178 |     0 | English  | James Cameron  |  7.9 | Sci-Fi    |
|   2 | Pirates of the Caribbean: At World's End | 2007 | 309404152 |      169 |     0 | English  | Gore Verbinski |  7.1 | Action    |
|   2 | Pirates of the Caribbean: At World's End | 2007 | 309404152 |      169 |     0 | English  | Gore Verbinski |  7.1 | Adventure |
+-----+------------------------------------------+------+-----------+----------+-------+----------+----------------+------+-----------+
```

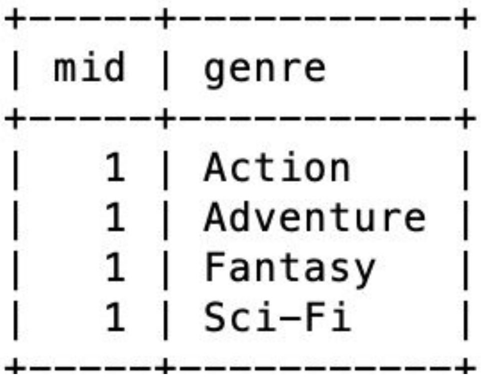
But this part is partially dependent on the part of FD, mid.

Schema Design Principles #2

Remove partial dependencies



```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
| mid | title  | year | gross     | duration | color | language | director      | imdb |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+
```



```
+-----+-----------+
| mid | genre     |
+-----+-----------+
|   1 | Action    |
|   1 | Adventure |
|   1 | Fantasy   |
|   1 | Sci-Fi    |
+-----+-----------+
```

Second Normal Form

Schema Design Principle #3

No Transitive Dependencies

X -> Y is a transitive dependency

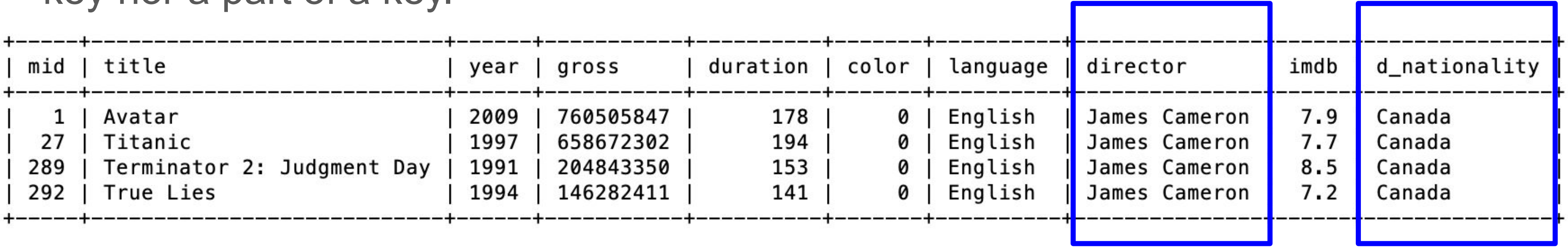
If both X -> Z and Z -> Y hold and Z is not a key nor a part of a key.

Schema Design Principle #3

Remove Transitive Dependencies

mid -> d_nationality is a transitive dependency

because both mid -> director and director -> d_natonality hold and director is not a key nor a part of a key.

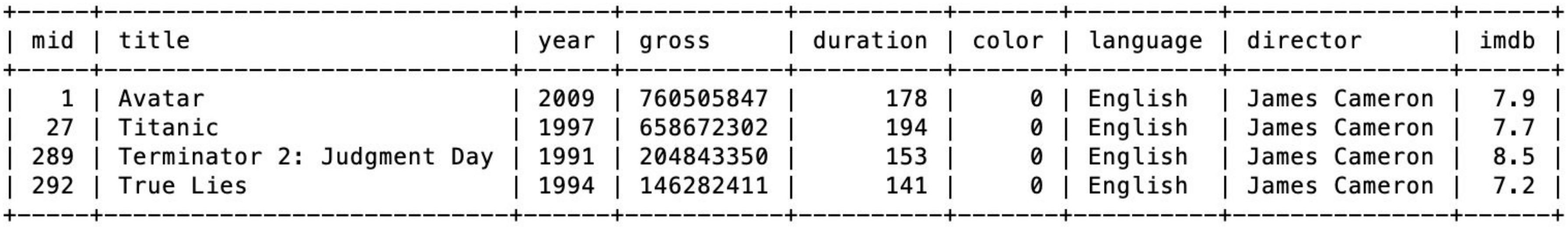


```
+-----+----------------------------+------+-----------+----------+-------+----------+--------------------------------------------------
| mid | title                      | year | gross     | duration | color | language | director         imdb    d_nationality    |
+-----+----------------------------+------+-----------+----------+-------+----------+--------------------------------------------------
|   1 | Avatar                     | 2009 | 760505847 |      178 |     0 | English  | James Cameron     7.9    Canada           |
|  27 | Titanic                    | 1997 | 658672302 |      194 |     0 | English  | James Cameron     7.7    Canada           |
| 289 | Terminator 2: Judgment Day | 1991 | 204843350 |      153 |     0 | English  | James Cameron     8.5    Canada           |
| 292 | True Lies                  | 1994 | 146282411 |      141 |     0 | English  | James Cameron     7.2    Canada           |
+-----+----------------------------+------+-----------+----------+-------+----------+--------------------------------------------------
```

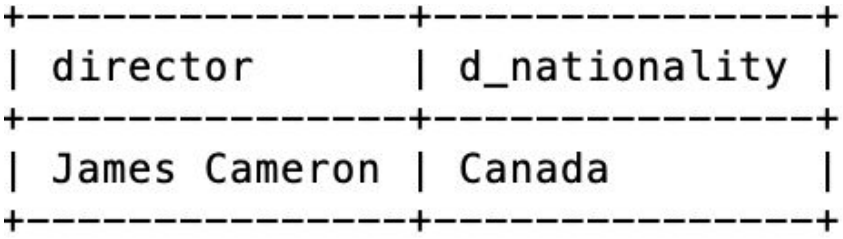
Schema Design Principle #3

Remove Transitive Dependencies

Third Normal Form



```
+-----+----------------------------+------+-----------+----------+-------+----------+---------------+------+
| mid | title                      | year | gross     | duration | color | language | director      | imdb |
+-----+----------------------------+------+-----------+----------+-------+----------+---------------+------+
|   1 | Avatar                     | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 |
|  27 | Titanic                    | 1997 | 658672302 |      194 |     0 | English  | James Cameron |  7.7 |
| 289 | Terminator 2: Judgment Day | 1991 | 204843350 |      153 |     0 | English  | James Cameron |  8.5 |
| 292 | True Lies                  | 1994 | 146282411 |      141 |     0 | English  | James Cameron |  7.2 |
+-----+----------------------------+------+-----------+----------+-------+----------+---------------+------+
```



```
+---------------+---------------+
| director      | d_nationality |
+---------------+---------------+
| James Cameron | Canada        |
+---------------+---------------+
```

Informal Design Guidelines for Relation Schema

Guideline 1:

Design a relation schema so that it is easy to explain its meaning.

- Do not combine attributes from multiple kinds of relations.

Informal Design Guidelines for Relation Schema

Guideline 2:

Design the base relation schemas so that no anomalies are present in the relations.

- No redundant information.

Informal Design Guidelines for Relation Schema

Guideline 3:

Avoid placing attributes in a base relation whose values may frequently be null as much as possible.

- No frequently empty columns.

Informal Design Guidelines for Relation Schema

Guideline 3:

Avoid placing attributes in a base relation whose values may frequently be null as much as possible.

- No frequently empty columns.

mid	title	year	...	Star1	Star2	Star3	Star4	Star5
1	Avatar	2009		CCH Pounder	Joel David Moore	Wes Studi	Null	Null

Informal Design Guidelines for Relation Schema

Guideline 3:

Avoid placing attributes in a base relation whose values may frequently be null as much as possible.

- No frequently empty columns.

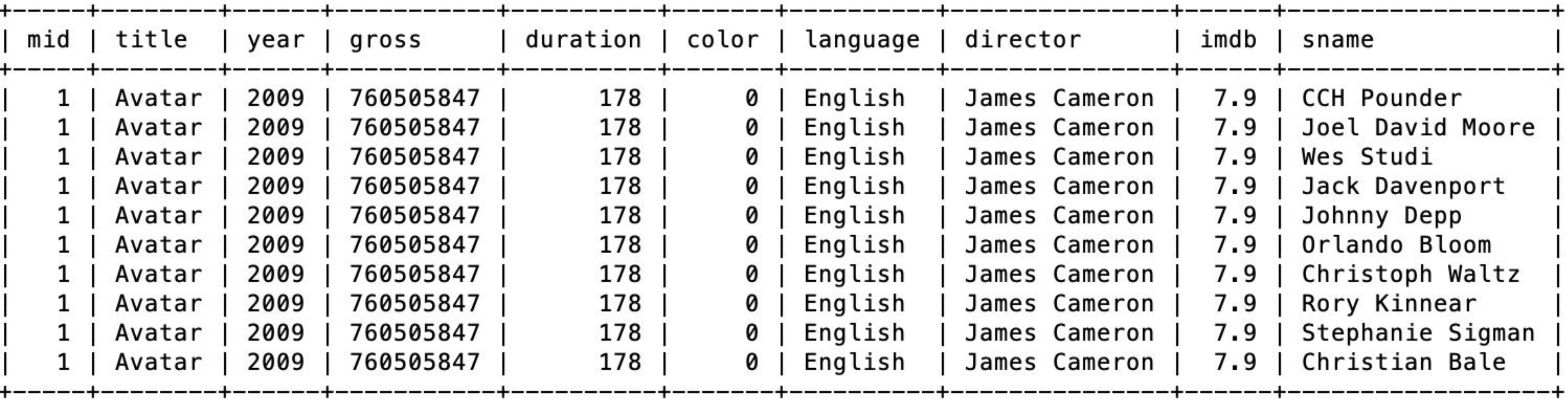
mid	title	year	...	Adventure	Action	Comedy	Darama	...
1	Avatar	2009		True	True	False	False	False

Informal Design Guidelines for Relation Schema

Guideline 4:

Design relation schemas so that they can be JOINED with equality conditions on attributes that are either primary keys or foreign keys in a way that guarantees that no spurious tuples (unrelated tuples joined) are generated

Guideline 4: Prevent Spurious tuples



```
+-----+--------+------+-----------+----------+-------+----------+---------------+------+------------------+
| mid | title  | year | gross     | duration | color | language | director      | imdb | sname            |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+------------------+
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | CCH Pounder      |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Joel David Moore |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Wes Studi        |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Jack Davenport   |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Johnny Depp      |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Orlando Bloom    |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Christoph Waltz  |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Rory Kinnear     |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Stephanie Sigman |
|   1 | Avatar | 2009 | 760505847 |      178 |     0 | English  | James Cameron |  7.9 | Christian Bale   |
+-----+--------+------+-----------+----------+-------+----------+---------------+------+------------------+
```

High Level Database Design

- High level design helps you create a blueprint for your database

High Level Database Design

- E-R Model
- UML

Entity-Relationship (ER) Model

ER Model: A representation of the data for an organization, business area, etc. Expressed in terms of entities, relationships and attributes.

ER Diagram: A graphical representation of an entity-relationship model. Sometimes abbreviated as ERD.

ER Definitions

Entity: Principal data object about which information is to be collected. Usually a person, place, thing, or event. Analogous to object in OOP.

Entity Set: Set of similar entities. Analogous to class in OOP.

Relationship: Real-world association among one or more entities

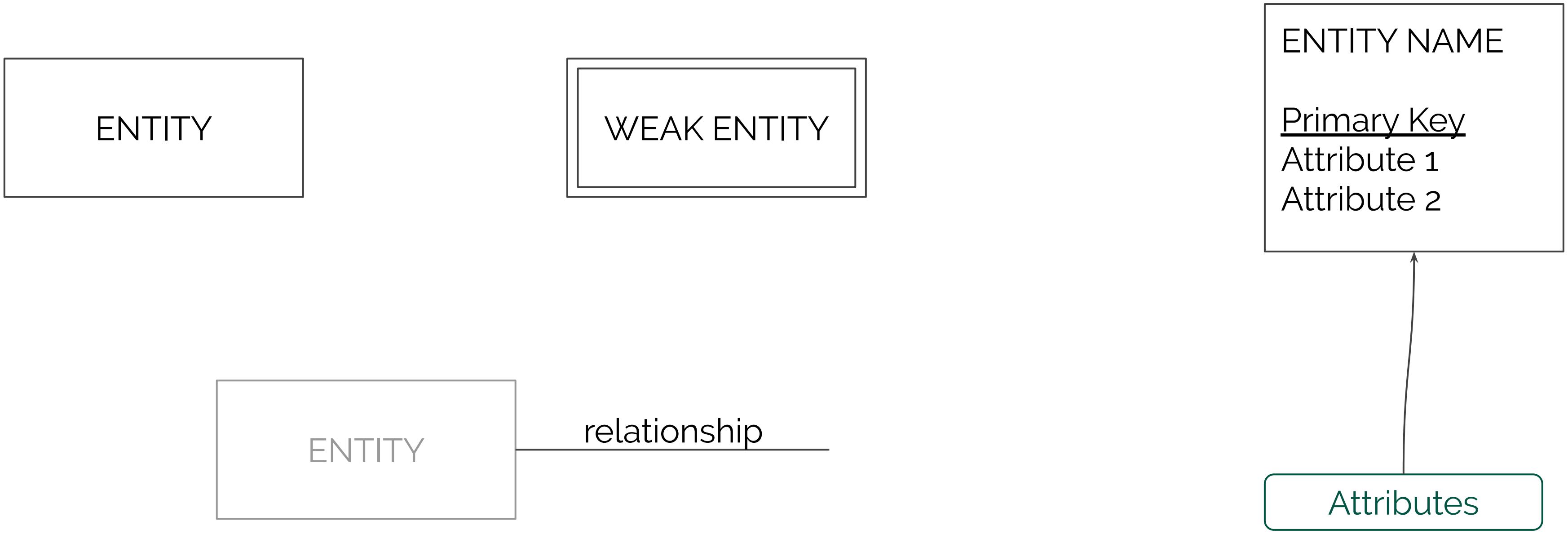
Attribute: Provides descriptive information about an entity or relationship

ER Definitions - Types of Entities

[Strong] Entity: An entity that exists independently of other entities

Weak Entity: An entity whose existence depends on some other entity type (for example: Event in the English Premier League Soccer database depends on Game because a part of its primary key, game_id, comes from Game.)

ER Diagram - Basic Notation



ENTITY

WEAK ENTITY

ENTITY NAME

Primary Key
Attribute 1
Attribute 2

ENTITY

relationship

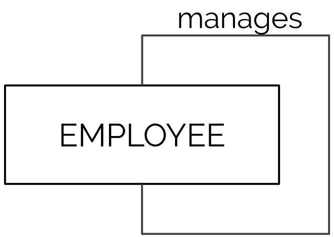
Attributes

ER - Relationship Concepts

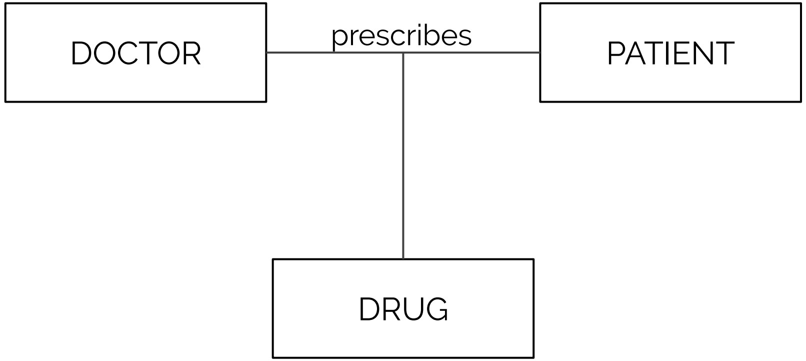
Degree: Number of entities that participate in a relationship

Entity Count	Degree of Relationship
1	Unary
2	Binary
3	Ternary
4 or more	n-ary

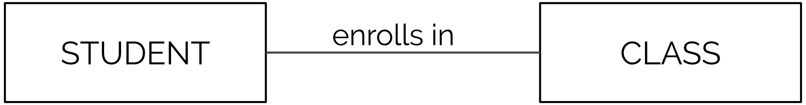
ER Diagram - Relationship Degree



Unary (Recursive)

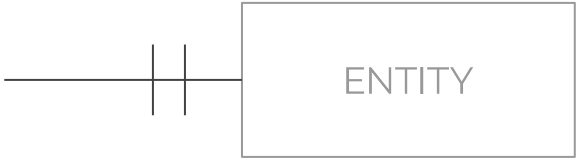


Ternary

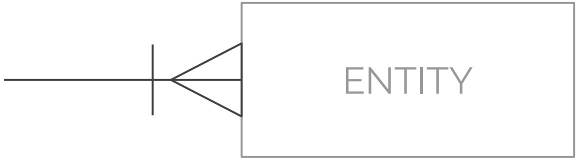


Binary

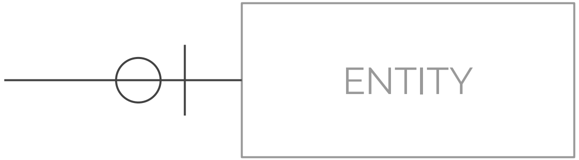
ER Diagrams - Cardinality / Existence



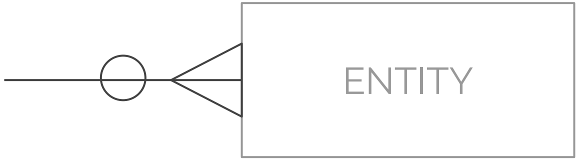
Mandatory One (1..1)



Mandatory Many (1..*)

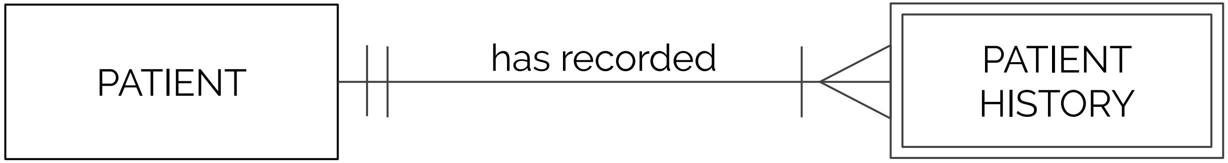


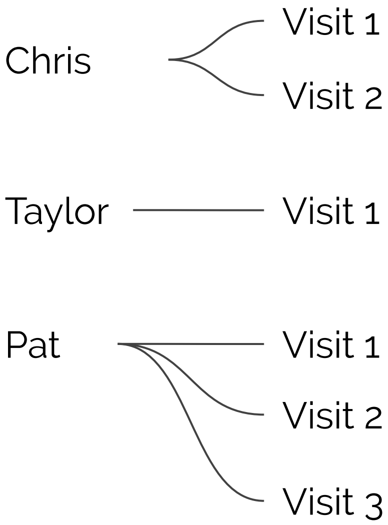
Optional One (0 .. 1)



Optional Many (0 .. *)

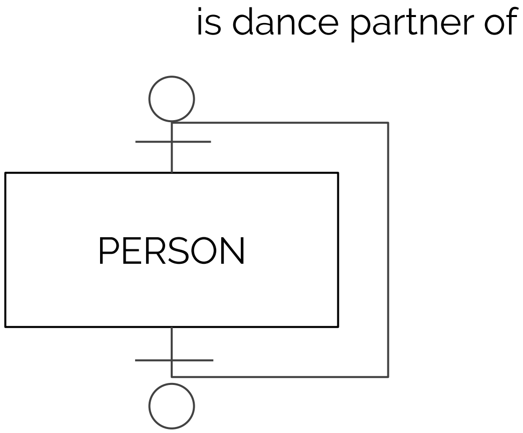
ER Diagram - One-to-Many Mandatory

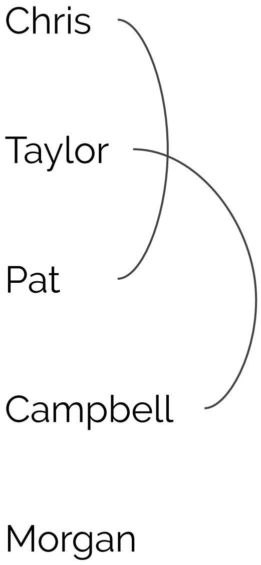




Degree?	Binary
Cardinality?	One to many (mandatory on both sides)

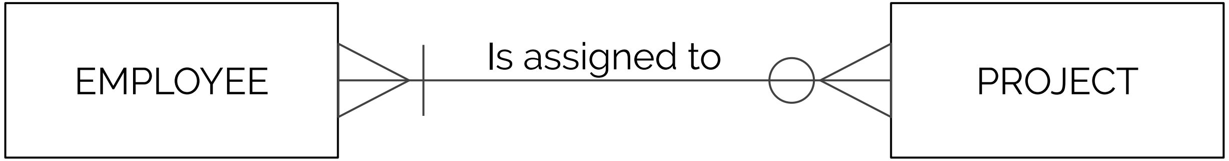
ER Diagram - Unary One-to-One Optional

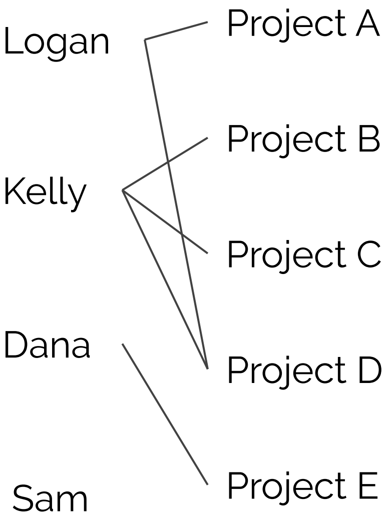




Degree?	Unary
Cardinality?	One to one (optional on both sides)

ER Diagram - Many-to-Many

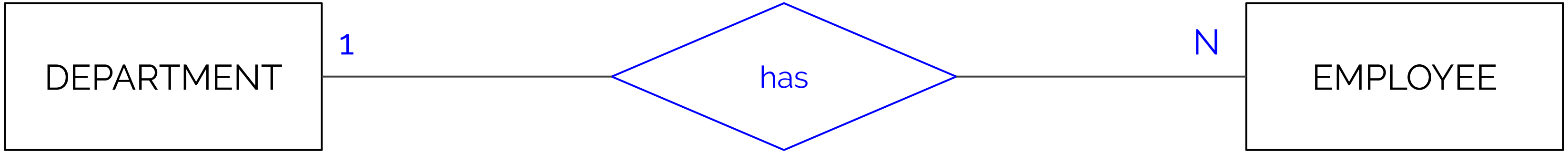




Degree?	Binary
Cardinality?	Many (mandatory) to many (optional)

Chen Notation

One to Many:



DEPARTMENT

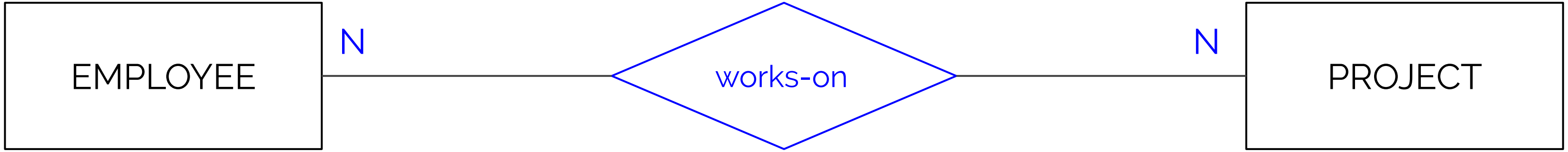
1

has

N

EMPLOYEE

Many to Many:



EMPLOYEE

N

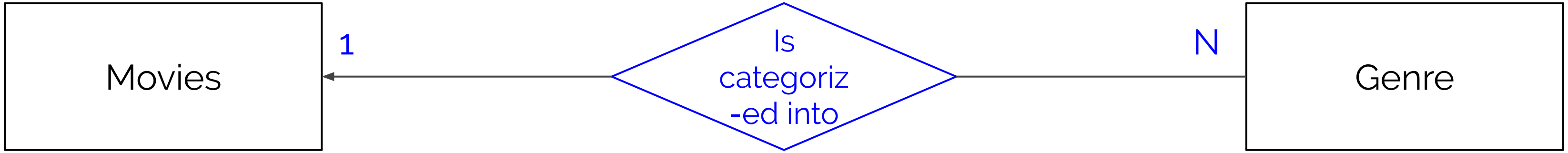
works-on

N

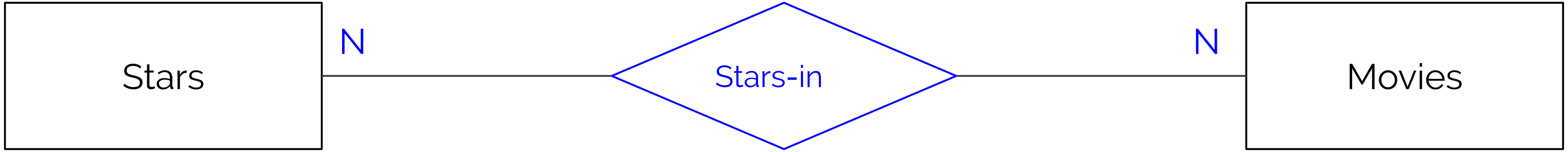
PROJECT

Chen Notation

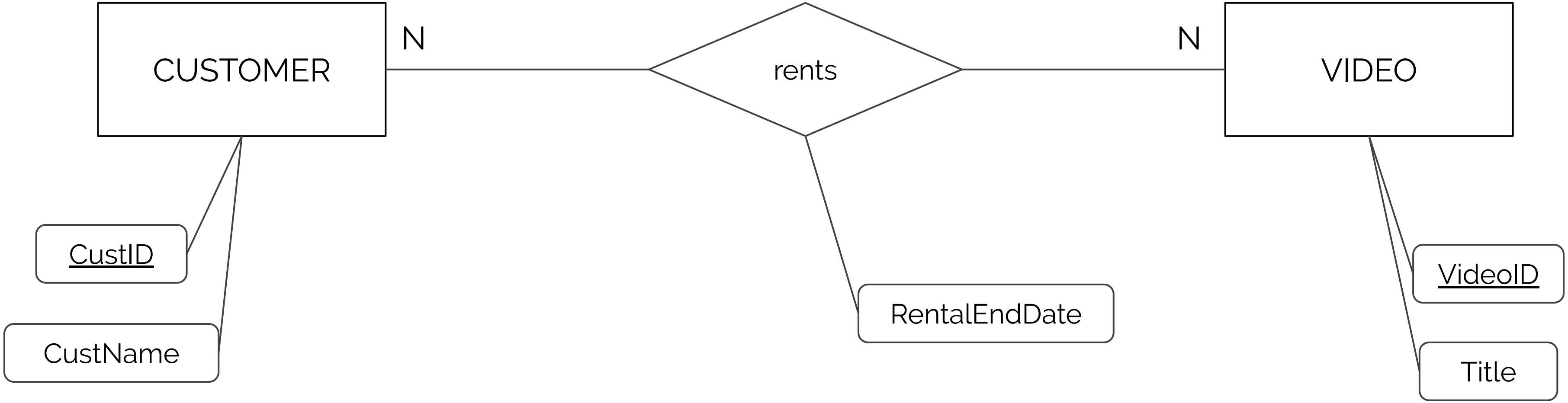
One to Many:



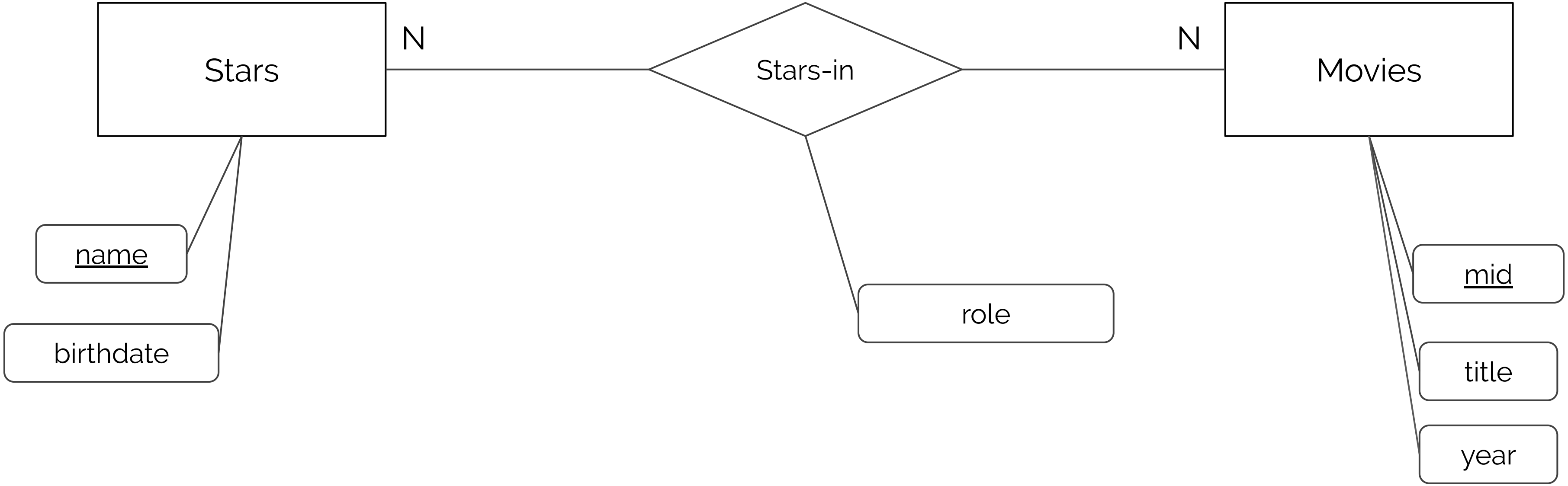
Many to Many:



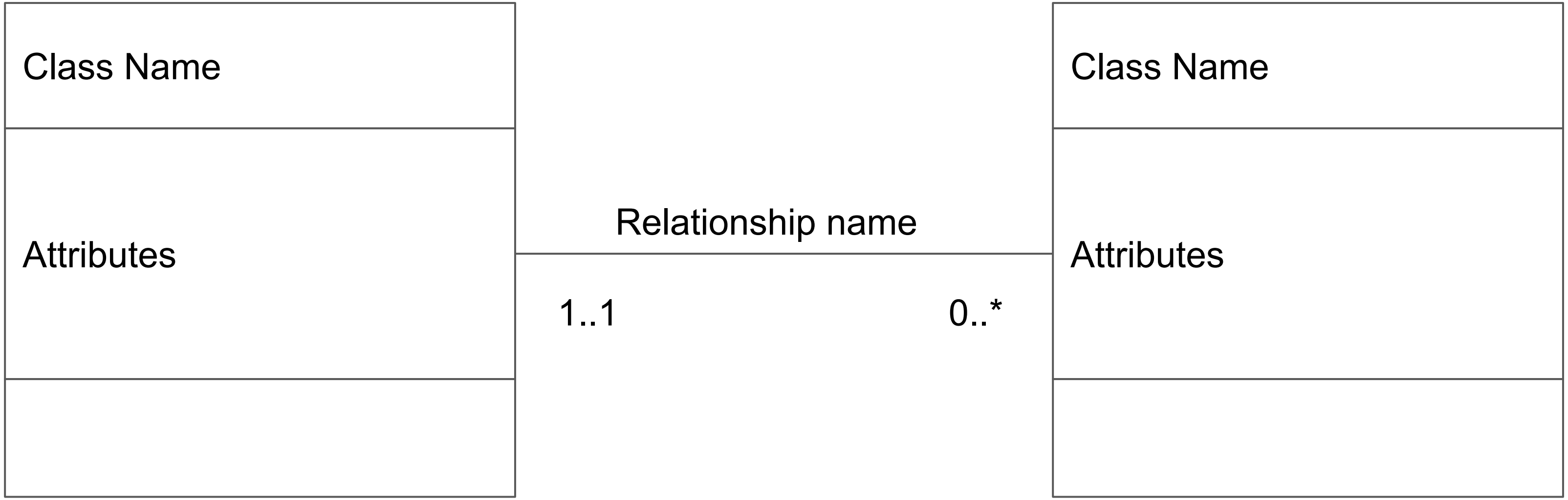
Chen Notation - Attributes



Chen Notation - Attributes



UML Diagram



Class Name

Attributes

Relationship name

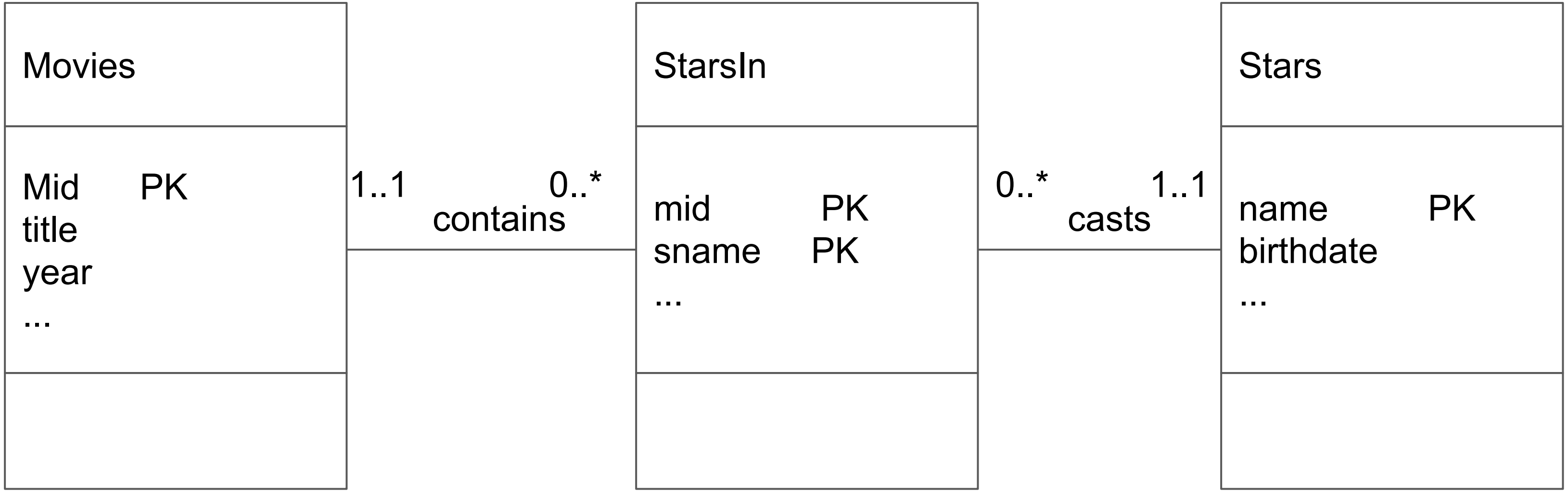
1..1

0..*

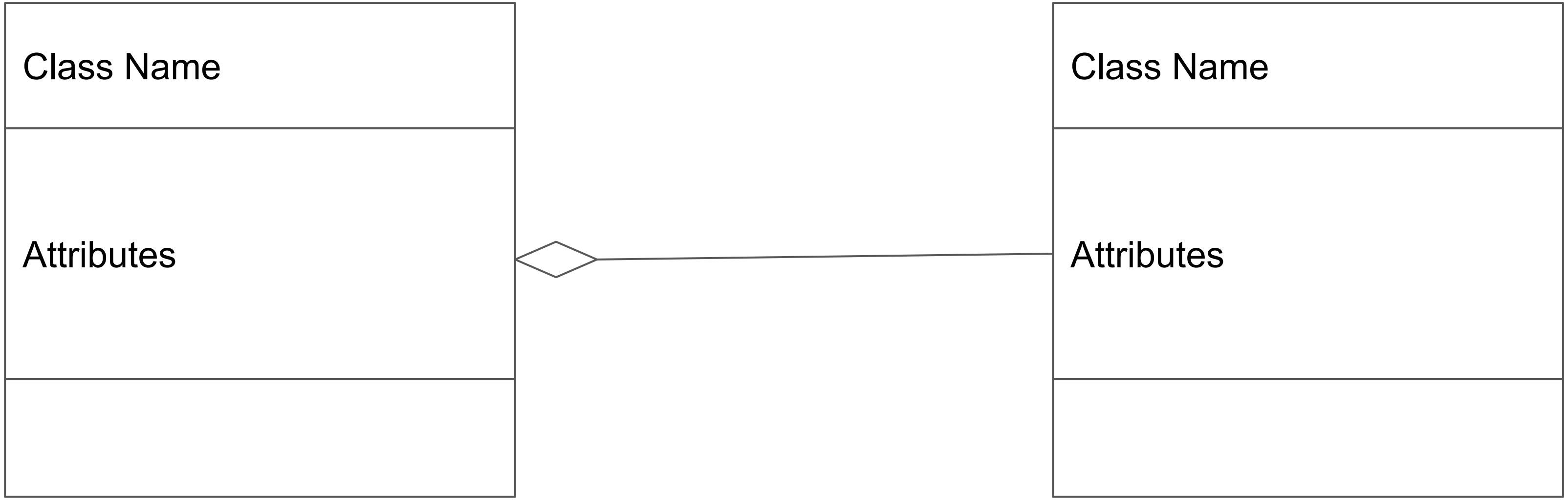
Class Name

Attributes

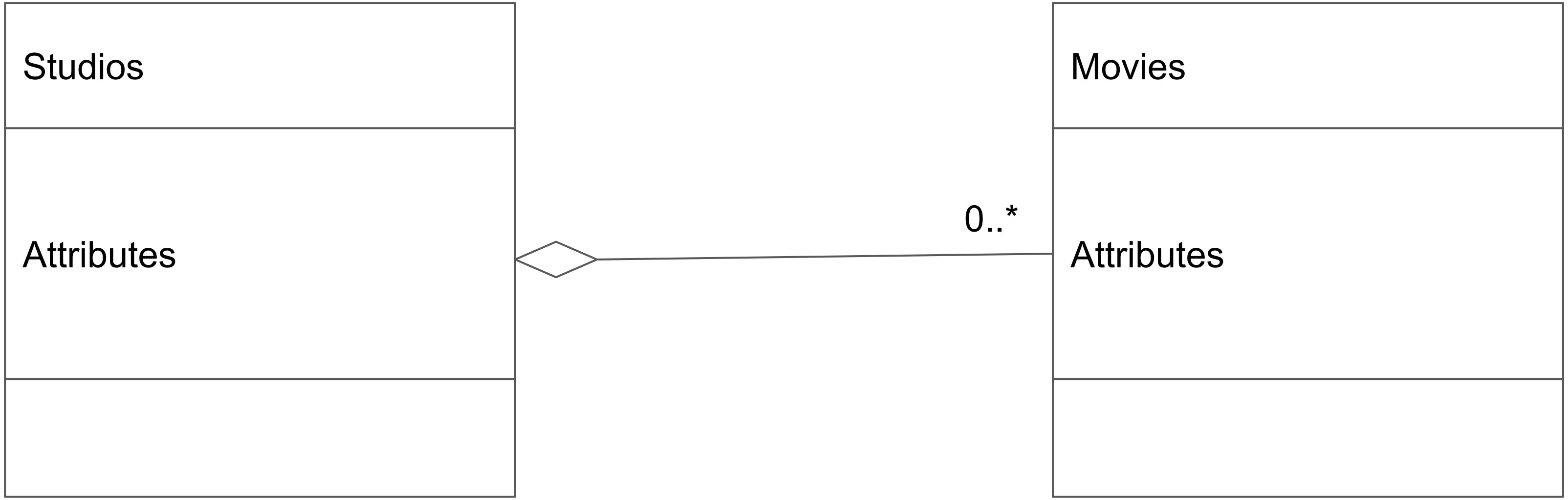
UML Diagram


Movies
Mid PK
title
year
...
1..1
contains
0..*
StarsIn
mid PK
sname PK
...
0..*
casts
1..1
Stars
name PK
birthdate
...

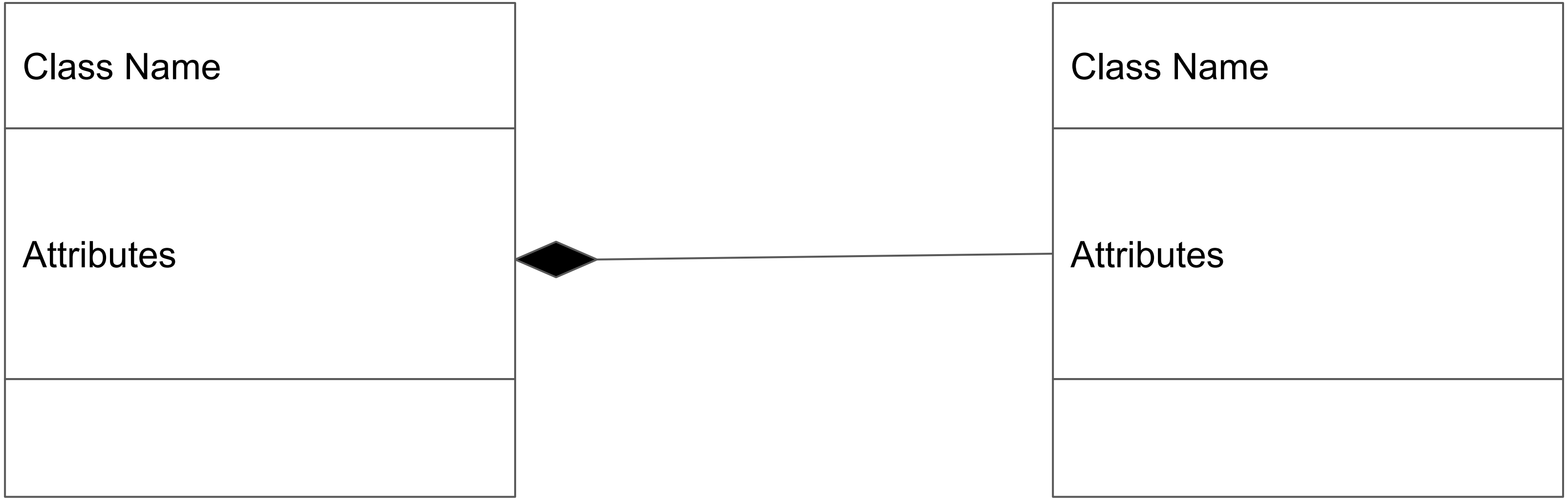
Aggregation



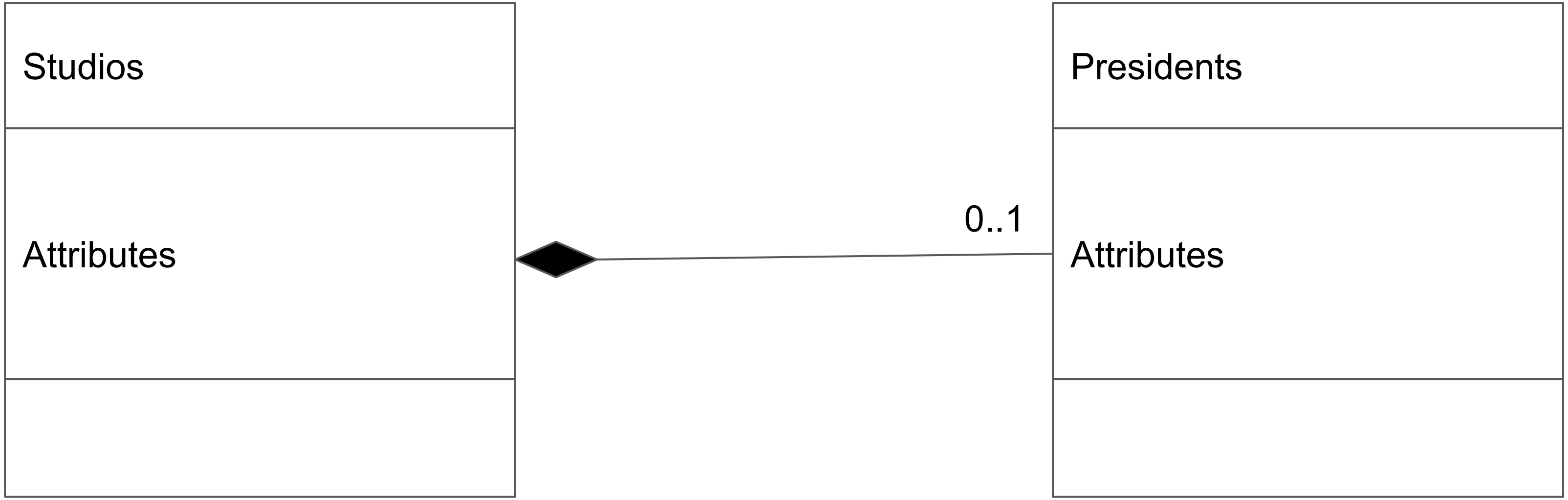
Aggregation



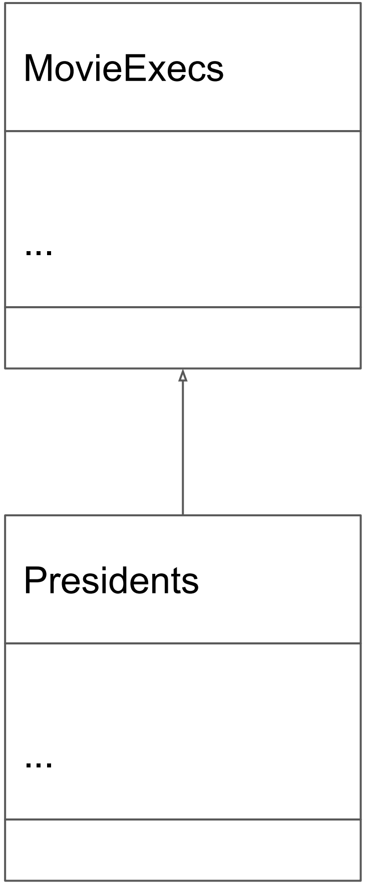
Composition



Composition



Is a



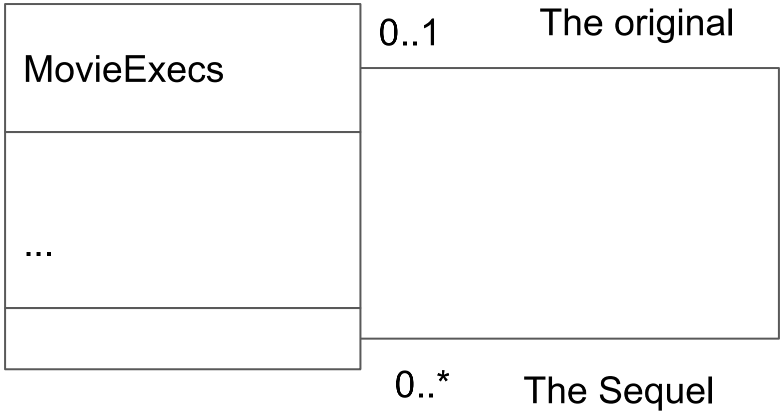
MovieExecs

...

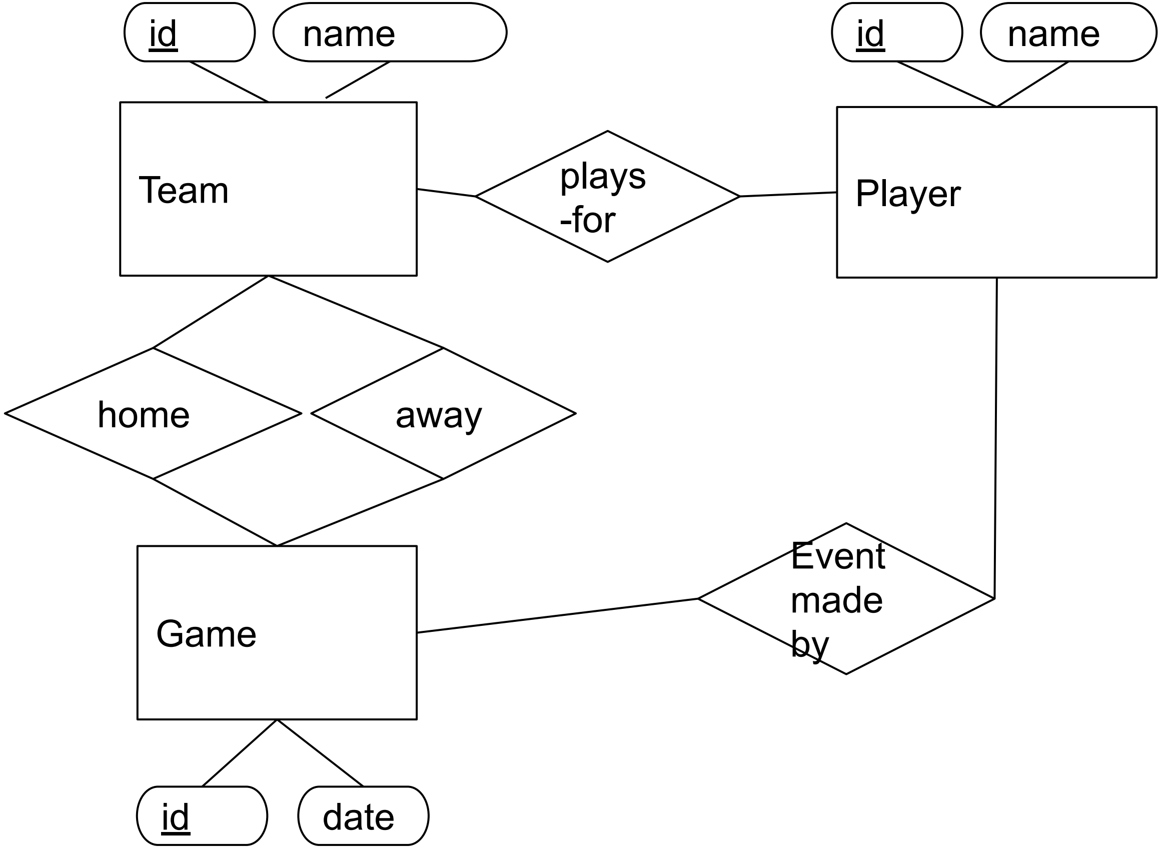
Presidents

...

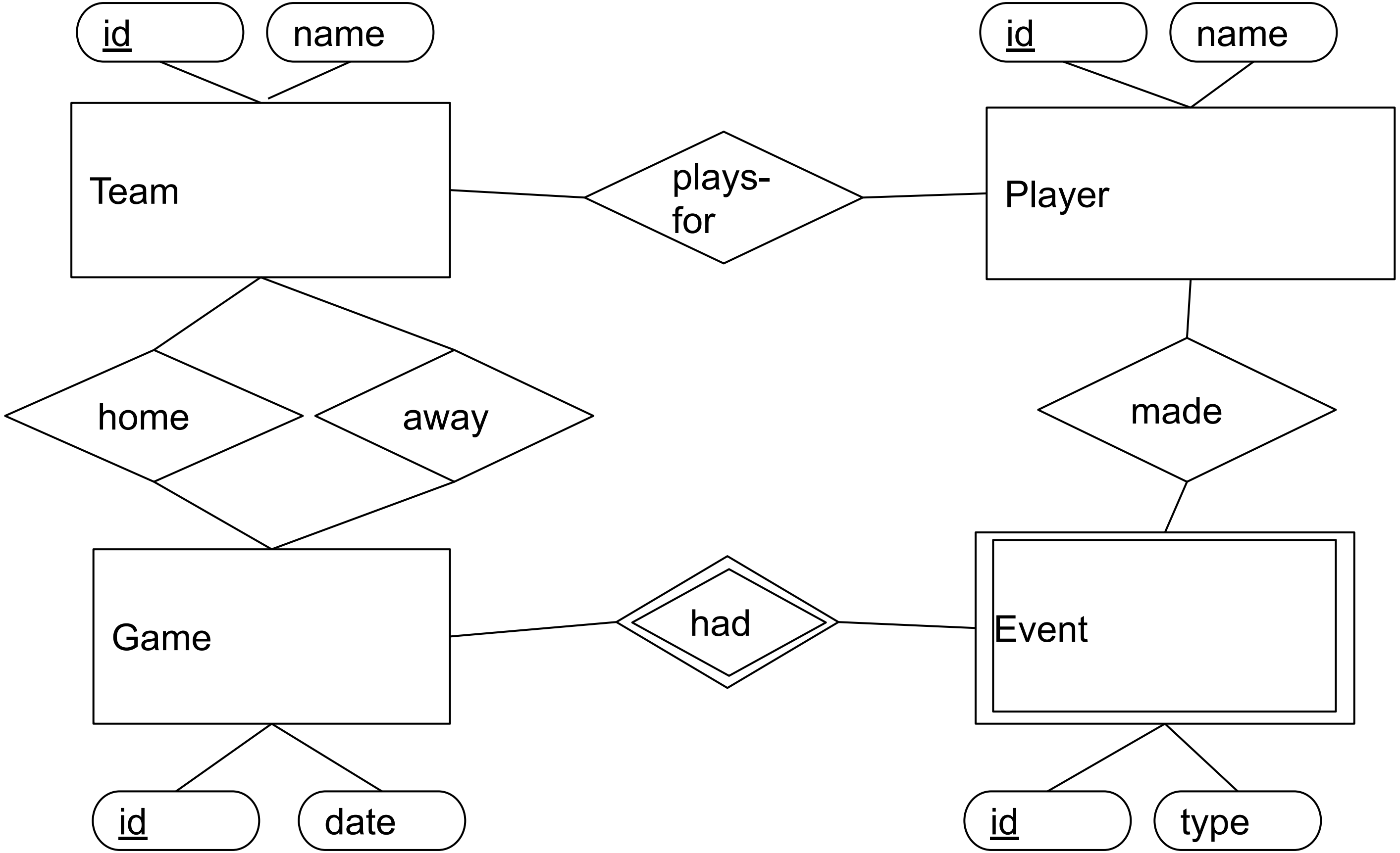
Self-Association (Recursive Relationship)



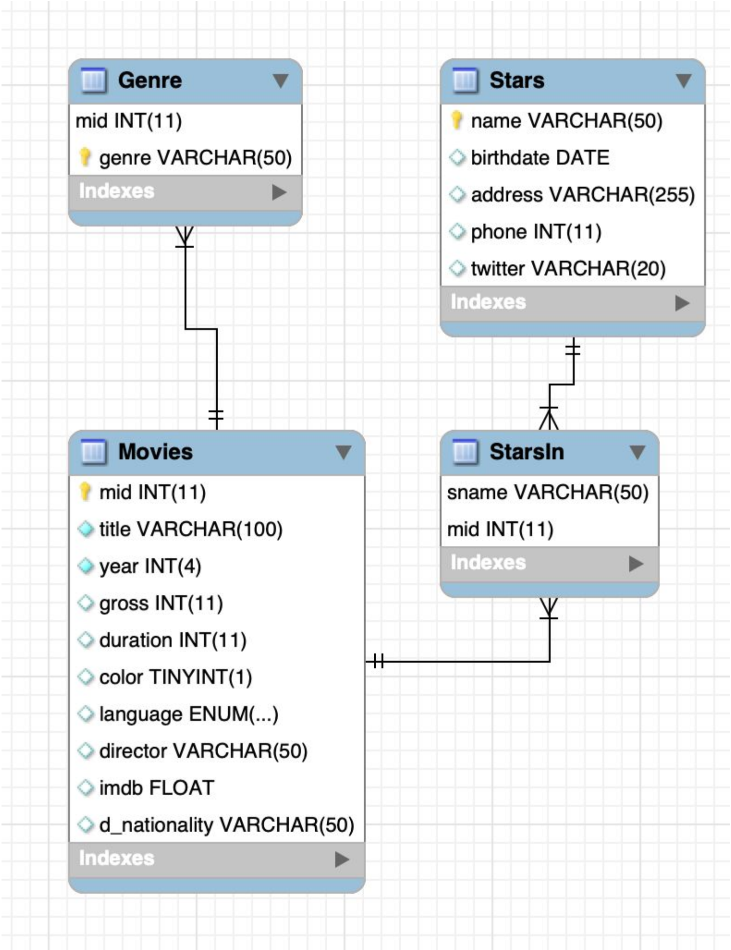
Examples


id
name
Team
plays
-for
id
name
Player
home
away
Game
Event
made
by
id
date

Examples


id
name
id
name
Team
plays-for
Player
home
away
made
Game
had
Event
id
date
id
type

Examples



Genre

mid INT(11)

genre VARCHAR(50)

Indexes

Stars

name VARCHAR(50)

birthdate DATE

address VARCHAR(255)

phone INT(11)

twitter VARCHAR(20)

Indexes

Movies

mid INT(11)

title VARCHAR(100)

year INT(4)

gross INT(11)

duration INT(11)

color TINYINT(1)

language ENUM(...)

director VARCHAR(50)

imdb FLOAT

d_nationality VARCHAR(50)

Indexes

StarsIn

sname VARCHAR(50)

mid INT(11)

Indexes

Exercises