Tutorial 01
ER Modelling
Big Data Engineering

Prof. Dr. Jens Dittrich

bigdata.uni-saarland.de

April, 23 2025

Repetition - Question 1

Question

Why does it make sense to create a data model and implement it in a database management system *before* the concrete realisation of the information system?

Repetition - Question 1

Question

Why does it make sense to create a data model and implement it in a database management system *before* the concrete realisation of the information system?

Solution

- Structuring the requirements leads to a systematic approach.
- You can already obtain feedback from the customer after the modelling.
- You can recognise potential issues/misunderstandings relatively early and thus reduce the costs/expenses to fix them.
- You can use the same model/data in multiple applications.
- The maintenance and extension of the system is easier.

Repetition - Question 2

Question

Which of the following statements are correct? Justify your answer.

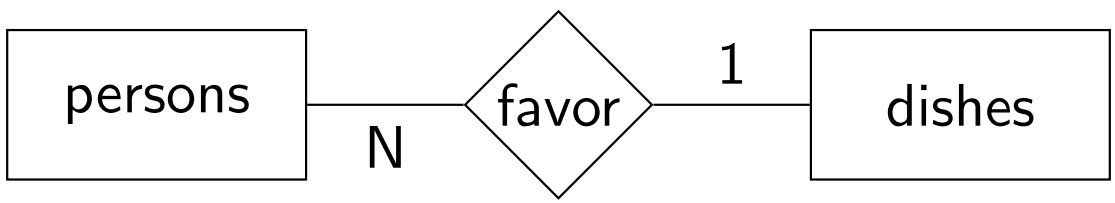
(a) Entities from dishes have a relation with at most one entity from persons.

(b) Entities from dishes can have a relation with multiple persons.

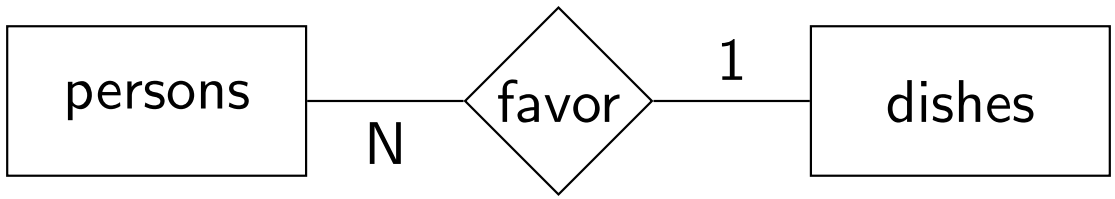
(c) Persons can have a relation with multiple dishes.

(d) Persons can have a relation with at most one dish.

(e) Entities from persons have a relation with exactly one entity from dishes.



Repetition - Question 2



Solution

Statements (b) and (d) are correct:
Dishes are functionally determined by persons, which means, that every person can favor at most one dish, i.e. either none or exactly one dish. However, dishes can have a relation with multiple persons, as they don't functionally determine persons.

Repetition - Question 3

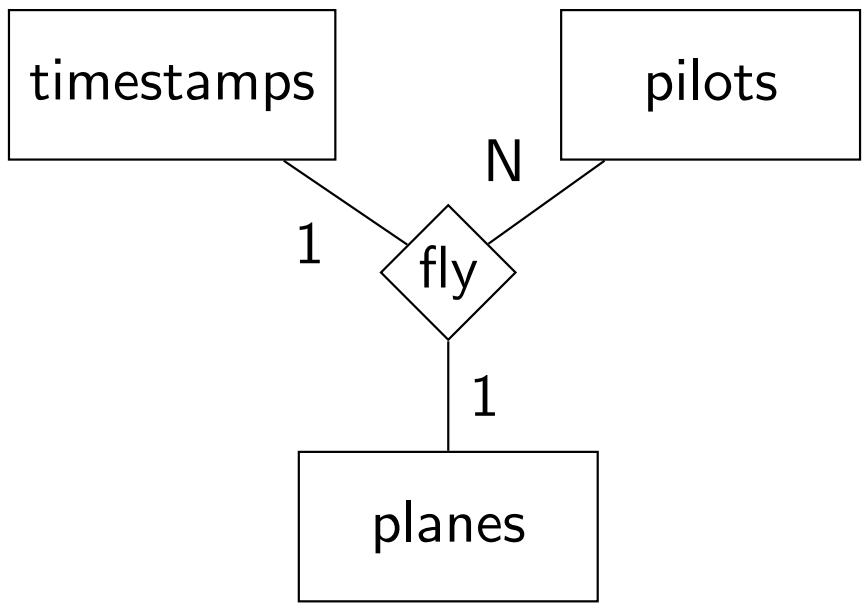
Question

Which of the following statements are correct? Justify your answer.

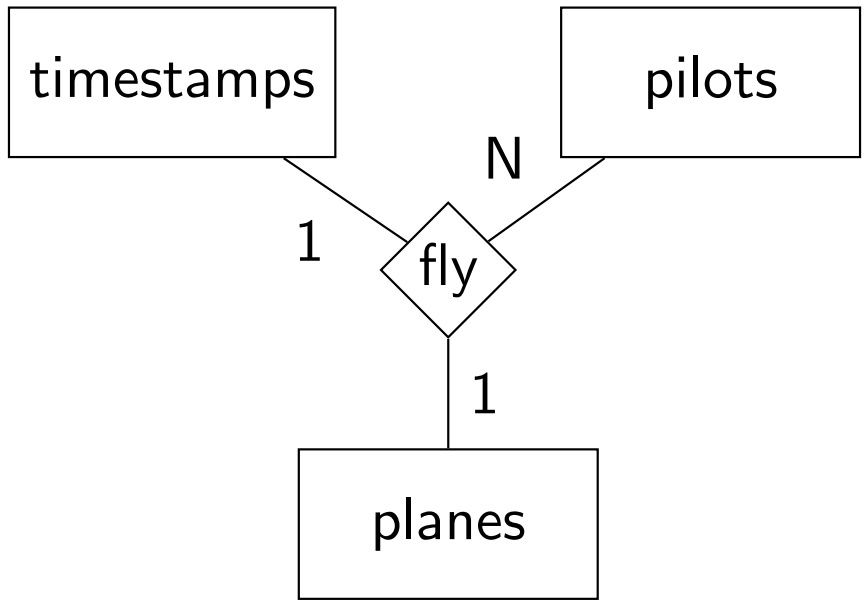
(a) Every pilot can fly at most one plane.

(b) Planes can be flown by multiple pilots at a given timestamp.

(c) Every plane can be flown at most one time.



Repetition - Question 3



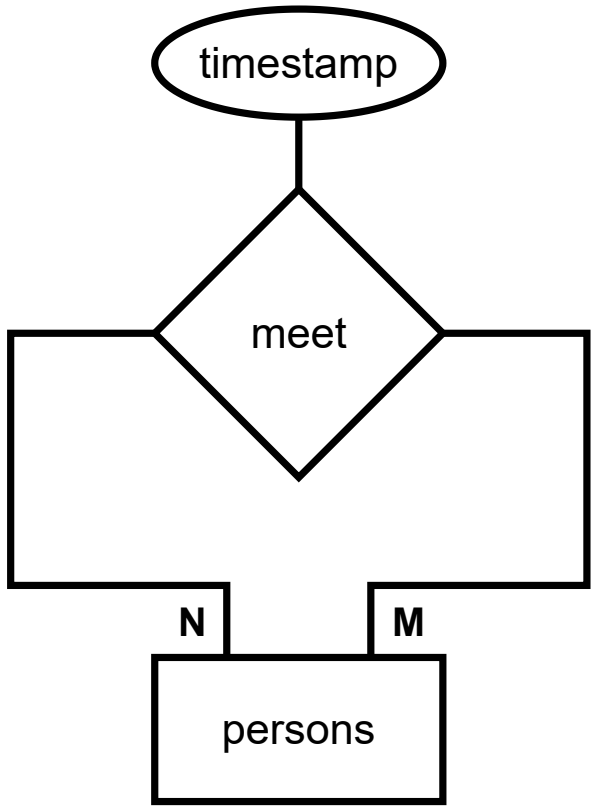
Solution

Only statement (b) is correct:
Since planes and timestamps do not functionally determine pilots, multiple pilots can fly the same plane at the same time.

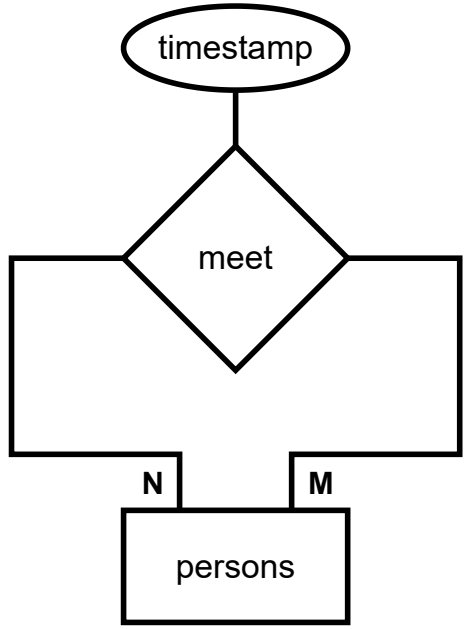
Repetition - Question 4

Question

What is the problem of the following model?



Repetition - Question 4



Solution

The problem is that every person can meet every other person only one time. In order to solve this, you can model the timestamp of the relation 'meet' as a key. As a result, persons can meet each other at different timestamps.
However, note that keys in relations are not supported by every definition of an ER model.

Repetition - Question 5

Question

Which version of the relation [play_in] does make more sense in the context of our IMDb ER model? Justify your answer.

(a) [play_in] : {[movie_id:(movies), actor_id:(actors), role:str]}

(b) [play_in] : {[movie_id:(movies), actor_id:(actors), role:str]}

Edit

Dr. Seltsam oder: Wie ich lernte, die Bombe zu lieben (1964)

Full Cast & Crew

See agent

Directed by

Stanley Kubrick

Writing Credits

Stanley Kubrick
Terry Southern
Peter George

Peter George

Cast (in credits order) verified as complete

Peter Sellers	...	Group Capt. Lionel Mandrake / President Merkin Muffley / Dr. Strangelove
George C. Scott	...	Gen. 'Buck' Turgidson
Sterling Hayden	...	Brig. Gen. Jack D. Ripper
Keenan Wynn	...	Col. 'Bat' Guano
Slim Pickens	...	Maj. 'King' Kong

Solution

ATTENTION!

Peter Sellers had in "Dr. Strangelove" **three** different roles! That's why we need the alternate model:

[play_in] : {[movie_id:(movies), actor_id:(actors), role:str]}

That's because, theoretically, this could happen in every movie.

Dr. Seltsam oder: Wie ich lernte, die Bombe zu lieben

Details

Full Cast and Crew

User Lists

Related lists from IMDb users

1960s
a list of 29 titles
created 29 Oct 2017

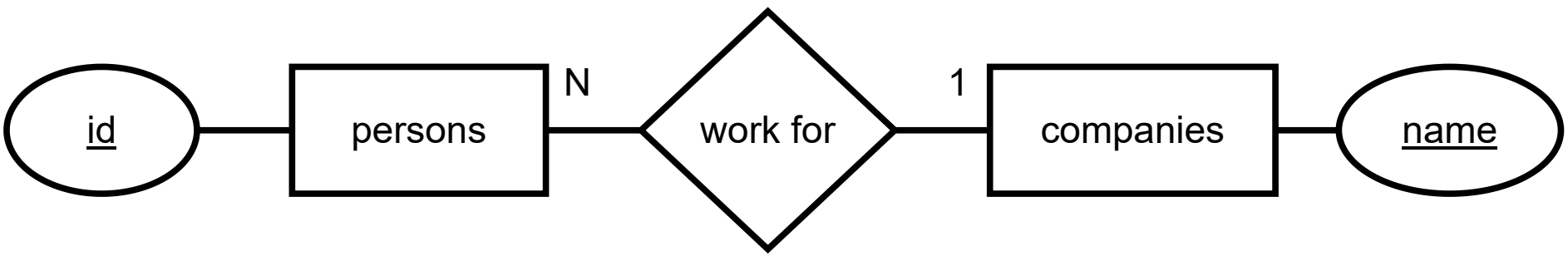
must see
a list of 37 titles
created 1 day ago

The 25 Best War Movies of All Time
a list of 25 titles

Repetition - Question 6

Question

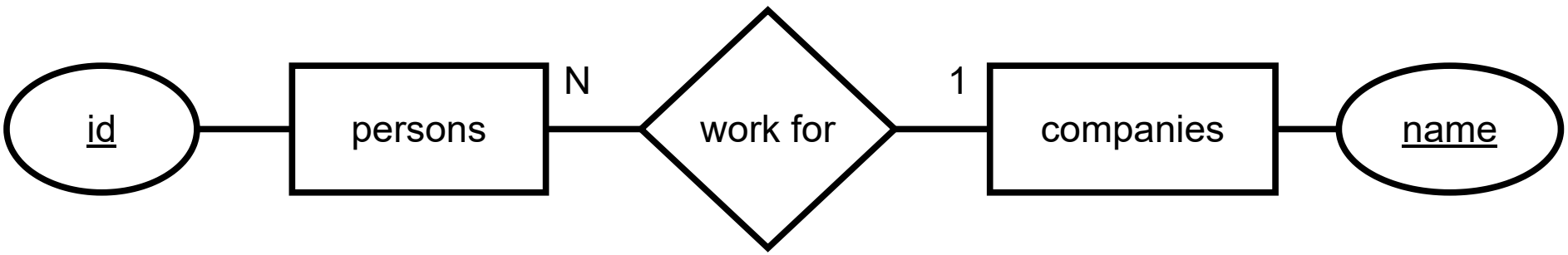
Translate the following ER model into a minimised relational schema.



Repetition - Question 6

Question

Translate the following ER model into a minimised relational schema.



Solution

$$[\text{persons}] : \{[\underline{\text{id:int}}, \text{company\_name:(companies)}]\}$$
$$[\text{companies}] : \{[\underline{\text{name:string}}]\}$$

Repetition - Question 7

Question

With how many cars can a person be in a relation according to the relational schema below?

$$[\text{sit\_in}] : \{[\underline{\text{person\_id:(persons), timestamp:time}}, \text{car\_id:(cars)}]\}$$

Repetition - Question 7

Question

With how many cars can a person be in a relation according to the relational schema below?

$$[\text{sit\_in}] : \{[\underline{\text{person\_id:(persons), timestamp:time}}, \text{car\_id:(cars)}]\}$$

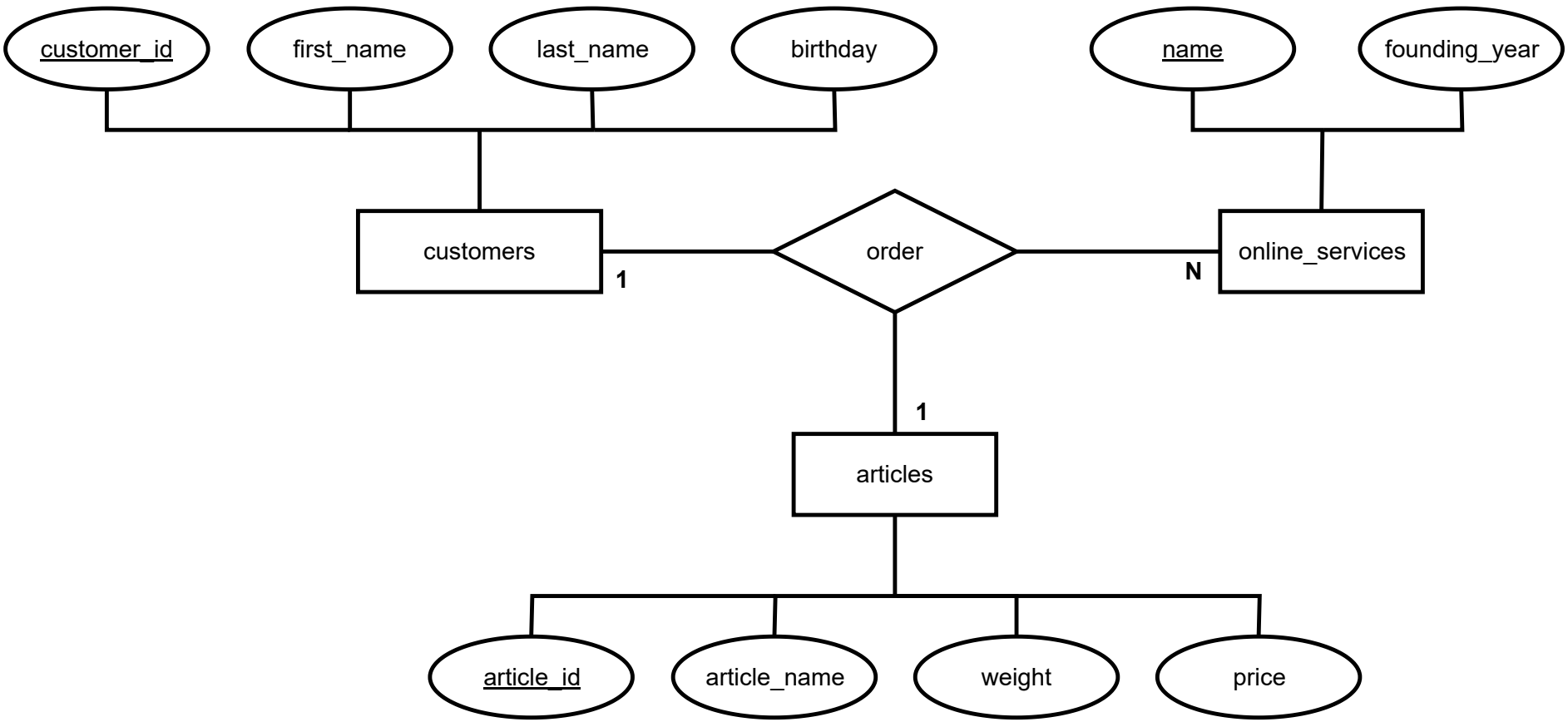
Solution

A person can be in a relation with any number of cars. Only at a given timestamp, a person can sit in at most one car.

Exercise 1.1

Question

Have a look at the following ER model that models the ordering of articles at online services. Which combinations of entities functionally determine which other entity?



Exercise 1.1

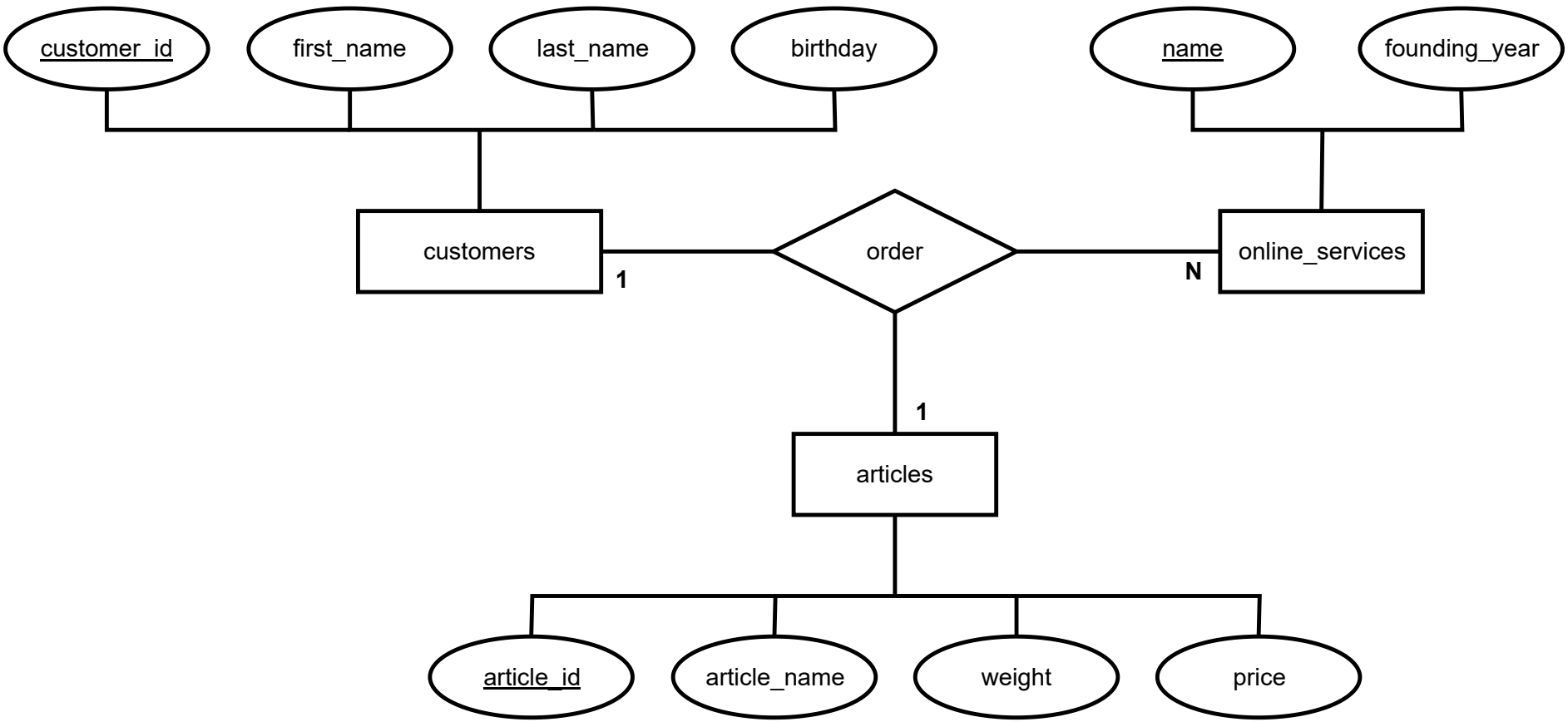
Solution

- Online services and customers determine articles:
  (online_services $\times$ customers $\rightarrow$ articles)
- Online services and articles determine customers:
  (online_services $\times$ articles $\rightarrow$ customers)

Exercise 1.2

Question

Does this modelling strategy make sense? Present an alternative model.



Exercise 1.2

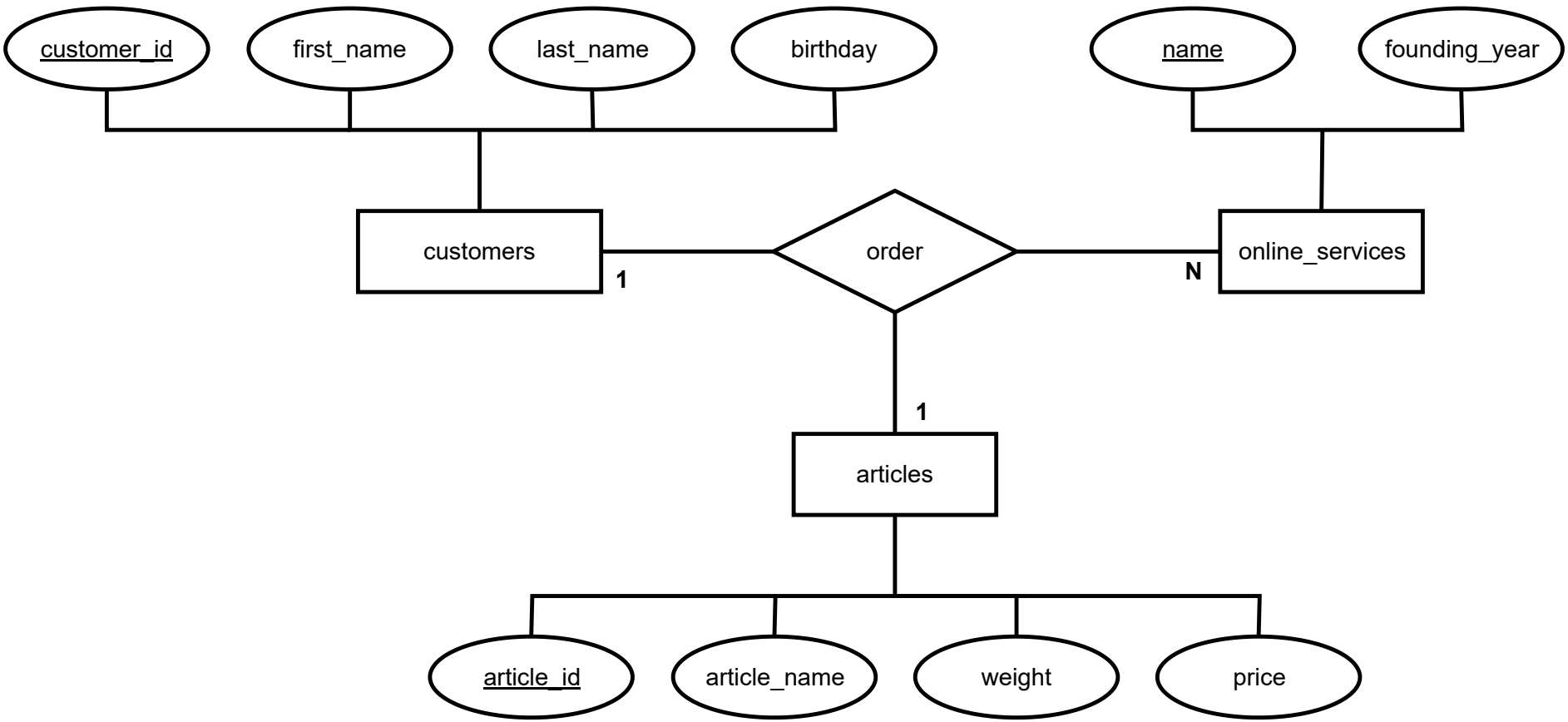
Solution

Because of the cardinalities, one customer can order only one article at a given online service. Further, given an online service, an article can be bought by one customer only. A better idea would be to change to cardinalities at customers and articles to M and K, respectively. This way, multiple customers could order an article at an online service and one customer could order multiple articles.

Exercise 1.3

Question

Which key combinations are possible when translating the following ER model into the relational model? Provide the corresponding relation [order] for each possibility.



Exercise 1.3

Solution

You can use either of the two modellings presented below.

$$[\text{order}] : \{[\text{customer:(customers)},\ \underline{\text{article:(articles)},}\ \underline{\text{online\_service:(online\_services)}}]\}$$

$$[\text{order}] : \{[\underline{\text{customer:(customers)}},\ \text{article:(articles)},\ \underline{\text{online\_service:(online\_services))}}]\}$$

Exercise 1.4

Question

Which problem arises from the translation into the relational model? What information will be lost?

Exercise 1.4

Solution

Depending on the implementation, one of the following information is lost:

- Given an article and an online service, only one customer should be in the relation.
- Given a customer and an online service, only one article should be in the relation.

For example, the tuples (12, 5, 3) and (12, 2, 3) would be allowed in the first schema, but not in the second. On the other side, the tuples (2, 5, 3) and (3, 5, 3) are allowed in the second schema, but not in the first one.

Exercise 1.5

Question

How could you solve the problem from 1.4 without changing the relational model?

Exercise 1.5

Solution

We cannot express this constraint in the relational model.
However, since in most cases the relational model is only an intermediate step, we can note down this integrity constraint and enforce it during the table definition (e.g. in SQL through UNIQUE).

Exercise 2

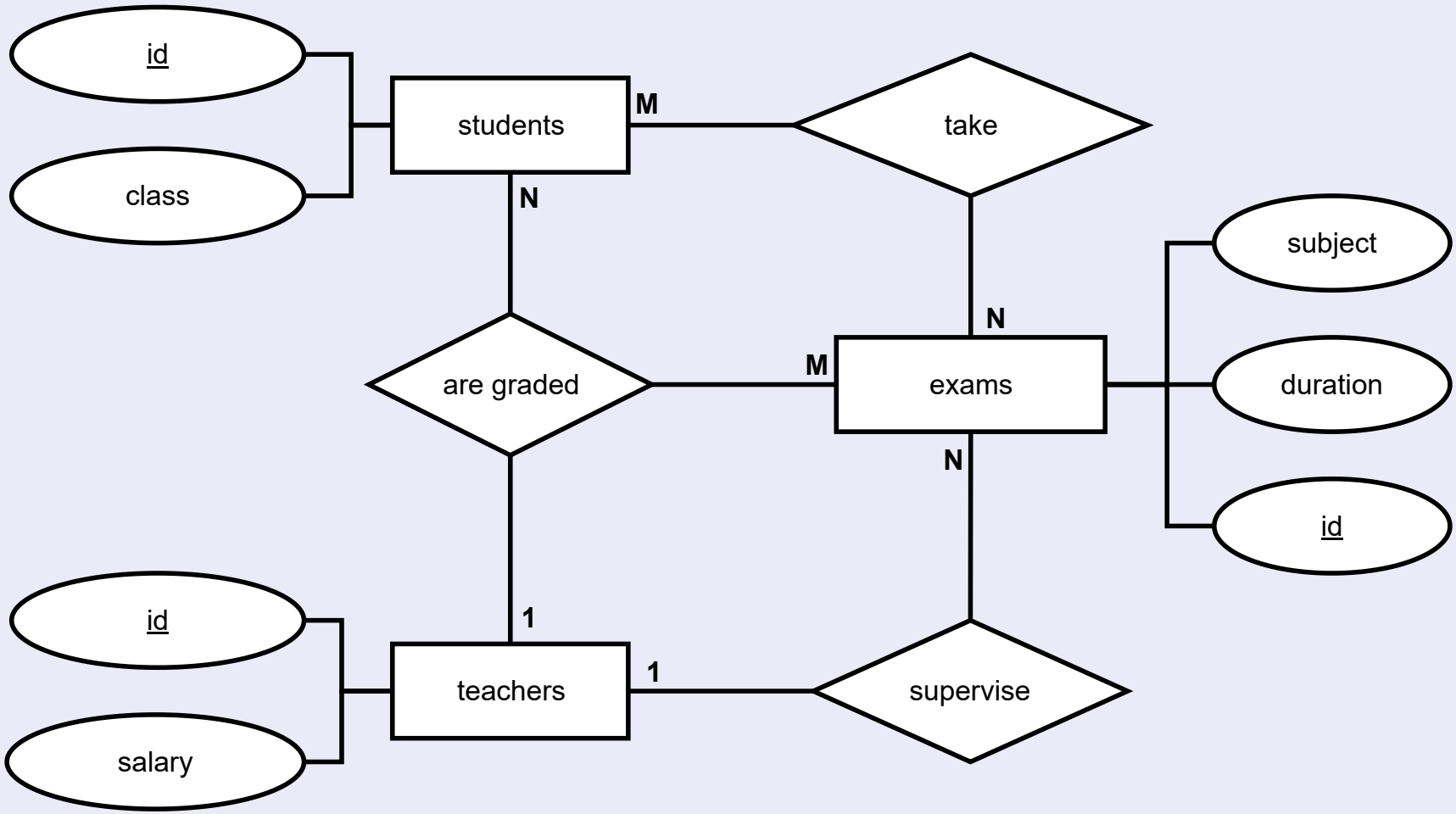
Question

Create an ER model according to the following specification. Choose meaningful cardinalities and use Chen notation. Introduce artificial keys where necessary.

- Exams have a subject and a duration.
- Teachers have a salary, while students are in a class.
- Each exam is supervised by one teacher and taken by multiple students.
- Exams are graded by one teacher for each student.

Exercise 2

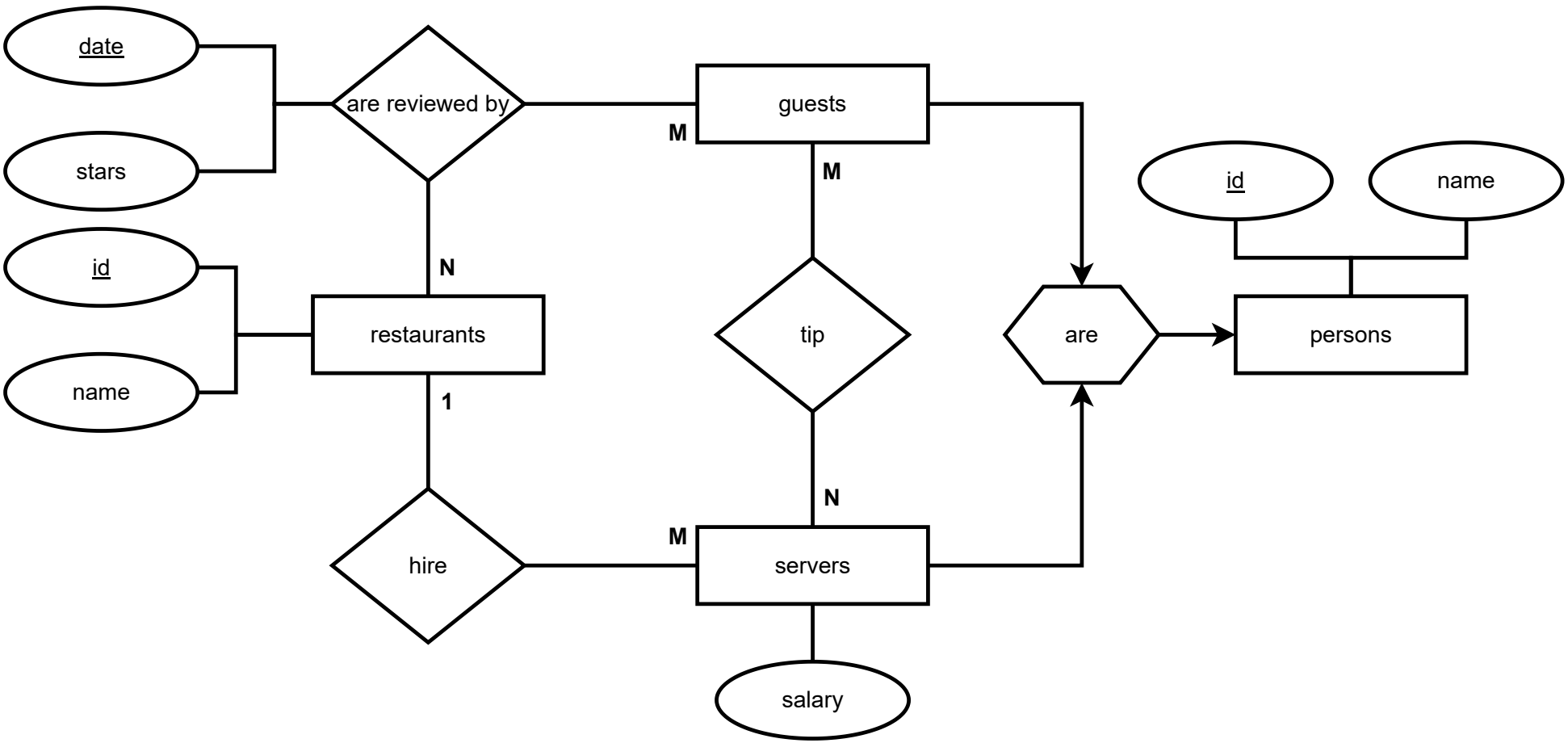
Solution



Exercise 3

Question

Translate the following ER model into the relational model. Simplify the relational schemas as far as possible.



Exercise 3

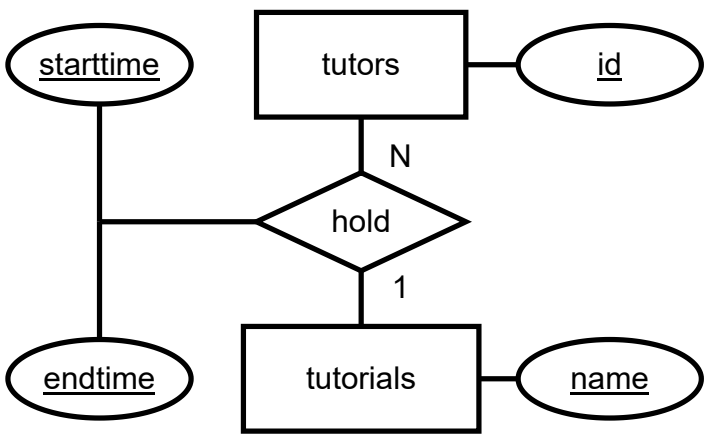
Solution

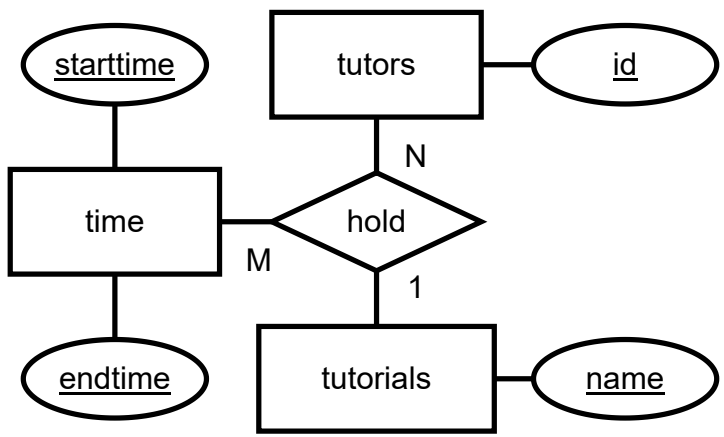
$$
\begin{aligned}
[\text{restaurants}] &: \{[\underline{\text{id:int}}, \text{name:string}]\} \\
[\text{persons}] &: \{[\underline{\text{id:int}}, \text{name:string}]\} \\
[\text{guests}] &: \{[\underline{\text{person\_id:(persons)}}]\} \\
[\text{servers}] &: \{[\underline{\text{person\_id:(persons)}}, \text{salary:int}, \\
&\qquad \text{restaurant:(restaurants)}]\} \\
[\text{are\_reviewed\_by}] &: \{[\underline{\text{guest:(guests)}}, \underline{\text{restaurant:(restaurants)}}, \\
&\qquad \underline{\text{date:time}}, \text{stars:int}]\} \\
[\text{tip}] &: \{[\underline{\text{guest:(guests)}}, \underline{\text{server:(servers)}}]\}
\end{aligned}
$$

Exercise 4

Question

Consider the following two ways of modelling time periods. What problems arise from these models?





Exercise 4

Solution

It is in both cases possible that the start and end time of two time periods overlap. This would mean that a tutor could hold multiple tutorials that overlap in time. This does not violate the key constraints, but it is not possible in reality.

Correction Mini-Test 01 - Common Mistakes

Exercise 1:

- Wrong or no usage of keys (e.g. assuming the name of teachers to be unique or forgetting to add artificial keys).
- Wrong functionalities (e.g. wrong direction of 1-N or multiple use of a single letter).
- Duplicated relationship types (e.g. 'work for' and 'employ').

Exercise 2:

- Functional determinacy based on only two entity types instead of three.
- Vague justifications (e.g. based on intuition instead of given functionalities).

Exercise 3:

- Incorrect symbol or missing arrows for inheritance.
- Missing attributes in introduced inheritance.
- Functionally determine planes by pilots and passengers (does not work without introducing e.g. a date attribute).
- Removing pilots from the ternary relationship 'fly' and introduce binary relationship 'steer'.