

Database systems, fall 2006

Rasmus Pagh

Exercises for the lecture on October 24

This week's exercises are theoretical, and consist of problems concerning normalization.

1. Problem 2 from exam in Introduction to Databases, June 2006

Consider a relation with the schema `Sales(reseller, manufacturer, product, amount)`. The following is a valid instance of `Sales`:

reseller	manufacturer	product	amount
Silman	SoftFloor AG	Velour	101000
Bjarnes Tæpper	Bøgetæpper	Berber	207000
Top Tæpper	Bøgetæpper	Kashmir	77000
Silman	SoftFloor AG	Berber	72000
Bjarnes Tæpper	Bøgetæpper	Valnød	17000

- a) Which of the following potential FDs do *not* hold, based on the instance above?
1. `amount` \rightarrow `product`
 2. `amount` \rightarrow `product reseller`
 3. `product` \rightarrow `manufacturer`
 4. `manufacturer` \rightarrow `product`
 5. `reseller product` \rightarrow `amount`

Note that the instance above is the “natural join” of the following relation instances:

reseller	manufacturer
Silman	SoftFloor AB
Bjarnes Tæpper	Bøgetæpper
Top Tæpper	Bøgetæpper

reseller	product	amount
Silman	Velour	101000
Bjarnes Tæpper	Berber	207000
Top Tæpper	Kashmir	77000
Silman	Berber	72000
Bjarnes Tæpper	Valnød	17000

- b) State a functional dependency (FD), that ensures that `Sales` can be split as in the above example without loss of information. The FD should in other words ensure that the SQL statement:

```
(SELECT reseller, manufacturer FROM Sales) NATURAL JOIN  
(SELECT reseller, product, amount FROM Sales)
```

always returns a relation that is identical to `Sales`. Further, express in words what the FD says.

- c) Give an instance of `Sales`, where the splitting fails, i.e., where the SQL statement of question b) does *not* return the same instance.

2. RG exercise 19.4.1.

3. **Problem 2 from exam in Databasesystemer, June 2005**

We consider the relation `Doctor`, containing information on medical doctors and the diseases they treat:

`Doctor(id, address, postcode, city, speciality, disease)`

Each doctor has a unique `id`, and furthermore precisely one address. For every post code there is exactly one city. A doctor can have more than one speciality, but in every city there is at most one doctor with a given speciality. A disease belongs to exactly one speciality, but a speciality may cover many diseases.

- (a) State all keys of `Doctor`. Explain any assumptions you make on the data.
- (b) Perform normalization to 3NF, stating the resulting relation schema, and what functional dependencies have been used.

4. Consider the following relation schemas, aimed at registering data on resellers of products in different cities:

`Reseller(id, address, city, product_id)`

A reseller can have several products, but `address` and `city` are functionally dependent on `id`. That is, `id` and `product_id` constitute a candidate key. We assume that each product has a reseller with exclusive rights in each city, i.e.,

`city, product_id` \rightarrow `id`.

In particular, this means that `city` and `product_id` also form a candidate key. Perform normalization to 3NF and BCNF, respectively. What are the advantages and disadvantages of the two normal forms in this case?

5. RG exercise 19.14.