Organization of Data bases

semester 4

lector: Lyudmyla Rozova, associate professor, phd

email: lyudmyla.rozova@khpi.edu.ua

Lecture 7. High-order normal forms

1. Multi-valued dependencies
2. Feigin Theorem
3. The fourth normal form (4NF)
4. Algorithm for normalization of relations

1. Multi-valued dependencies

To introduce **4NF**, it is necessary to introduce the concept of ***multivalued dependence*** (***MD***).

R is a *relation.*

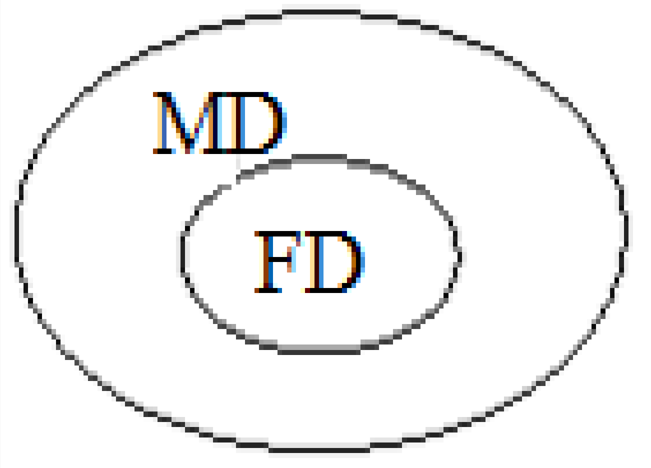
X, Y, Z some of its attributes (random subsets of the set of attributes of the relation **R**). They are disjoint sets of attributes.

Then the attributes (sets of attributes) **Y** and **Z** ***multi-valuedly*** depend on **X** (**$X \rightarrow\rightarrow Y|Z$**), if and only if from the fact that the relation ***R*** contains corteges **$r_1=(x,y,z_1)$** and **$r_2=(x,y_1,z)$** it follows that the relation **R** also contains the cortege **$r_3=(x,y,z)$**.

We can say: "*X multi-valued defines Y*" or "*Y multi-valued depends on X*".

In a simplified way, it can be said as follows:
X multi-valued defines ***Y*** if for each value of ***X*** there is no unique value ***Y*** corresponding to it (**it is not FD**), but each value of ***X*** determines the set of ***Y*** corresponding to it.
A multivalued dependence **X→→Y|Z** is called a **nontrivial multivalued dependence** if there are no functional dependencies **X→Y** and **X→Z**.

In fact, the **FD** is a special case of the **MD**:



Example 1.

An example to understand the essence of the MD:

Let there be a relation

STC (*Subject, Teacher, Classroom*).

Subject	Teacher	Classroom
Programming	Jonson	140
Programming	Jonson	218
Programming	Adison	140
Programming	Adison	218
Database	Richi	315
Database	Richi	411

The following restrictions apply:

a) any number of teachers and any number of classrooms can correspond to each subject;

b) teachers and classrooms are independent of each other (the same set of classrooms is used for one subject, regardless of who teaches a given subject);

c) a specific teacher and a specific room can be associated with any number of subjects.

In this case, the primary key is the set of ***all attributes***.

STC (<u>*Subject, Teacher, Classroom*</u>).

Let's check in which NF this relation is located.

1)**in 1 NF** - because all attributes are indivisible (we will assume that a teacher's code is used instead of the teacher's full name, but for convenience, we will use the full name in the example).

2**) in 2 NF** - because there are no non-key attributes at all.

3)**in 3 NF** - for the same reason.

4) NFBC - not suitable for this relation, because it does not have multiple overlapping potential keys.

Also, we can conclude that the completely key relation ***STC*** (if it is in 1 NF) is in 3 NF and in NFBC.

By constraint ***b),*** if there are two corteges:
{***Subject1, Teacher1, Classroom1***} and
{ ***Subject 1, Teacher2, Classroom 2***},

then there are also corteges
{ ***Subject1, Teacher1, Classroom2***} and
{ ***Subject1, Teacher2, Classroom 1***}.

For example, if corteges exist

Programming	**Jonson**	**140**
Programming	Adison	218

then there are also corteges

Programming	**Jonson**	**218**
Programming	Adison	140

In this case, the relation is ***clearly redundant*** and can lead to ***update anomalies***.

For example, to add information that the ***Programming*** will be taught by a new teacher , it is necessary to create as many new corteges as there are ***Classrooms,*** suitable for this subject.

And at the same time, you can mistakenly enter corteges not for all ***Classrooms***, or, with extra classrooms.

These problems are usually caused by the independence of the attributes.

In our example, the attributes ***Teacher*** and ***Classroom***.

How can we avoid redundancy and update anomalies?

We can replace the relation ***STC*** with two of its projections:

ST (Subject, Teacher) and ***SC (Discipline, Classroom)***

Both projections are completely key, therefore, they are in the ***NFBC***.

In addition, it is a lossless decomposition. If we join them by the attribute ***Subject,*** we will get the initial relation STC.

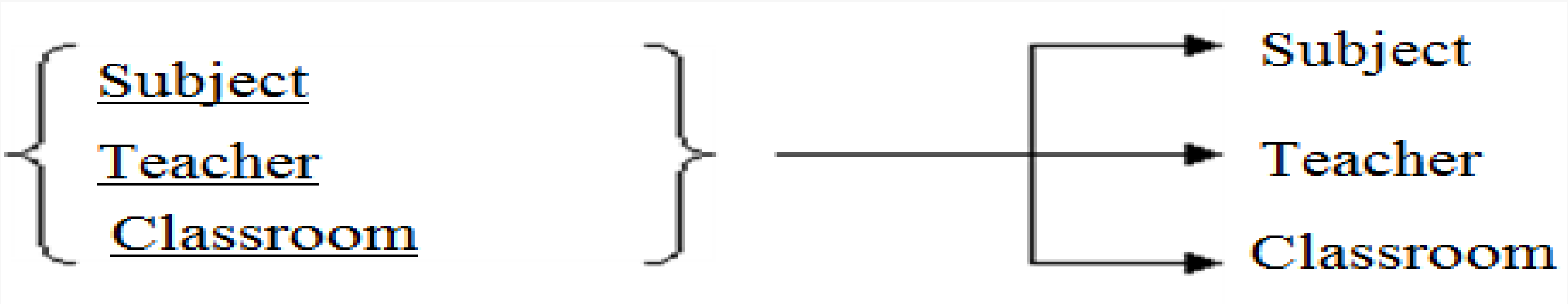
We can say that all this explanations were an intuitive level. We did not use any strict rules.

Only in 1971 these considerations were formalized by ***Feigin*** using the concept of <u>*multi-valued dependencies*</u>.

Decomposition of the *STC* relation

Let us formalize the process of decomposition of the ***STC*** relation.

This cannot be done, based on the ***FD***, because the ***FD*** diagram looks like this:



And all FDs are trivial here.

Decomposition of the ***STC*** relation

But the performed decomposition can be done on the basis of the ***MD***. And this relation has two of them:

Subject →→ Teachers

Subject →→ Classrooms

Subject →→ Teacher, means:

Each ***Subject*** defines a set of corresponding ***Teachers.***

So, for each ***Subject*** there is no single corresponding ***Teacher*** (FD Subject →Teacher is not correct)

Subject →→ Classroom means:

Each ***Subject*** defines a set of corresponding ***Classrooms***.

So, for each ***Subject*** there is no single corresponding ***Classroom*** (the FD Subject → Classroom is not correct).

Decomposition can be made on these ***MD***s.

2.Feigin Theorem

It can be noted that for the ***STC*** relation a multivalued dependence ***Subject*** $\rightarrow\rightarrow$ ***Teacher*** is performed if and only if the ***MD*** **Subject** $\rightarrow\rightarrow$ **Classroom** is performed.
For such relations, this is always true. So in a generalized form, the rule of multi-valued dependencies can be formulated:

For R (X, Y, Z) X $\rightarrow\rightarrow$ Y if and only if X $\rightarrow\rightarrow$ Z.

Theorem (Feigin): Let ***X, Y, Z*** be disjoint sets of attributes of the relation ***R (X, Y, Z)***. Decomposition of the relation ***R*** on the projection $\mathbf{R_1}$ ***= R [X, Y]*** and $\mathbf{R_2}$= ***R [X, Z]*** will be a lossless decomposition if and only if there is a multi-valued dependencies **X$\rightarrow\rightarrow$Y|Z**.

3. The fourth normal form (4NF)

Relation ***R*** is in the **fourth normal form (4NF)** if and only if the relation is in ***NFBC*** and does not contain ***nontrivial multi-valued dependencies.***

<u>Another formulation of 4NF:</u>

The ***R*** relation is in **4 NF** if it is in the ***NFBC*** and all ***MD***s of the relation ***R*** are in fact ***FD***s from potential keys.

The presented above ***StudentTeacherClassroom*** relation is not in 4 NF because (according to the first formulation):

1) There is an ***MD Subject→→Teacher***, but the *Classroom* attribute does not functionally depend on the *Subject* attribute.

2) There is an ***MD Subject→→ Classroom***, but the *Teacher* attribute does not functionally depend on the *Subject* attribute.

But both projections ***ST*** and ***SC*** are in ***4 NF***, because each of them has one ***MD***, and there are no other attributes (not included in these MDs).

Feigin also proved that ***4NF*** can always be obtained. Because any relation, containing an ***MD***, can be decomposed without loss into a set of relations in ***4 NF***.

4.Algorithm for normalization of relations.

Let us remember once again the algorithm for normalization of relations, *step by step*.

Step 1. (**to 1NF**). At the first step, one or more relations are set. They reflect the concepts of the subject area.

By the model of the subject area (not by the view of the obtained relations!), the existing functional dependencies are written out.

All relations are automatically in 1NF.

4.Algorithm for normalization of relations

Step 2. (**to 2NF**). If in some relations the dependence of attributes on a part of a complex key is found, then we decompose these relations into several relations as follows:

- Those attributes that depend on a part of a complex key are taken out into a separate relation together with this part of the key.

However, the part of the complex key that the non-key attributes depend on stays in the original relation also.

4.Algorithm for normalization of relations

Step 3. (*to 3NF*).

If in some relations the dependence of some non-key attributes on other non-key attributes is found, then the decomposition of these relations is carried out as follows:

- those non-key attributes that depend on other non-key attributes are taken into a separate relation together with the determinants of the functional dependencies of the non-key attributes.
- In addition, the determinants of these dependencies remain in the original relation.

The connection between relations obtained as a result of decomposition is carried out by means of foreign keys.

4.Algorithm for normalization of relations

Step 4. (***to NFBC***).

If there is a relation, containing several potential keys: then it is necessary to check whether there are functional dependencies, the determinants of which are not potential keys.

If there are such functional dependencies, then

- the decomposition of relations is carried out.
- those attributes that depend on determinants that are not potential keys are taken out into a separate relation together with the determinants.

Step 5. (***to 4NF***).

If nontrivial multivalued dependencies are found in the relation, then decomposition is performed to exclude such dependencies.