Organization of Data bases

semester 4

lector: Lyudmyla Rozova, associate professor, phd

email: lyudmyla.rozova@khpi.edu.ua

Lecture 6

1. High-order normal forms (NF)
2. Boyes-Codd normal form

1^{st}, 2^{nd}, 3^{rd} Normal Forms

First normal form

Relation is in **1^{st} NF** if and only if all domains used in it contain only **scalar values**.

In other words: The relation is in **1^{st} NF** if and only if the values of all fields are indivisible. All relations are automatically in **1^{st} NF**.

Second normal form

Relation is in 2^{nd} NF if and only if it is in 1^{st} NF and each non-key attribute is **irreducibly dependent on a primary key** (there are no non-key attributes that depend on a part of a complex key).

Third normal form

Relation is in **3^{rd}** NF if and only if it is in **2^{nd} NF** and each non-key attribute is **nontransitively dependent on the primary key**.

In other words: relation **R** is in third normal form (3NF) if and only if the relation is in 2NF and all non-key attributes are ***mutually independent*** (non-transitively dependent on the primary key)

High-order normal forms (NF)

The first three normal forms (1NF, 2NF, 3NF), in most cases, are sufficient to develop workable databases.

However, there are situations when the described algorithm of normalization to 3NF does not allow to completely eliminate all anomalies.

Therefore, further normal forms of higher orders will be considered:

- the Boyes - Codd normal form (BCNF),
- the fourth normal form (4NF),
- the fifth normal form (5NF).

Boyes-Codd normal form

When defining 3 NF, the assumption was made that the relation has ***only one potential key***, which is the ***primary*** one.

But the definition of 3 NF is not entirely appropriate for a relation with the following properties:

1. the relation has two or more potential keys;
2. potential keys are complex;
3. potential keys overlap, so, they have at least one common attribute.

For relations with these properties, the previously given definition of 3 NF is not sufficient, so Boyes and Codd introduced the definition of **Boyes-Codd NF.**

Boyes-Codd normal form

Def of BCNF : A relation is in BCNF if and only if each (nontrivial and left irreducible) FD has a potential key as a ***determinant.***

Or another definition of BCNF:

The **R** relation is in Boyes-Codd normal form (**BCNF**) if and only if the determinants of all functional dependencies are potential keys.

Note: if the relation is in the BCNF, it is automatically in 3NF.

In practice, relations with a combination of the listed properties are extremely rare, and for relations without these properties, 3NF and BCNF are equivalent.

Example 1

For example, the relation with two overlapping candidate keys.

Academic performance

(***GradebookNo., SubjectCode, SubjectName, Mark***)

Potential keys:

{***GradebookNo., SubjectCode*** } and

{***Gradebook No., SubjectName*** }.

This relation is not in BCNF, because it contains the following functional dependance **(FD)**:

SubjectCode → SubjectName and

SubjectName → SubjectCode

And the determinants of these FD are not potential keys.

This relation is also not in 3 NF, and not in 2 NF, because the attribute ***SubjectName*** depends only on a part of the primary key

(FD {***GradebookNo.***, ***SubjectCode*** } → { ***SubjectName*** } is reducible).

This relation is redundantly.

So the decomposition into two relations is possible:

Subject (***<u>SubjectCode,</u>SubjectName***)

Academic performance1

(***<u>GradebookNo.,</u>*** ***<u>SubjectCode,</u>***, ***Mark***)

Example 2

Consider an example of a relation that is in 3 NF but is not in BCNF.

Let the Faculty database have the relation

Lessons (*GroupCode, Subject, Teacher*):

GroupCode	Subject	Teacher
I-21	Programming	Jonson
I-22	Programming	Adison
I-31	Database	Melen
I-32	Database	Richi
…		

There will be the following restrictions:

a) each subject for a particular group is taught by only one teacher, but the same subject can be taught by different teachers;

b) each teacher teaches only one subject (in reality this is not always of course).

Then there are two potential keys:

{GroupCode, Subject}

and ***{GroupCode, Teacher}.***

Again the situation with two overlapping potential keys.

Let's the first one will be as **primary key**.

This relation is in 3 NF because:

- It is in 2NF.

Let's prove it:

the only FD of a non-key attribute

{GroupCode, Subject }* → *Teacher.

It is irreducible, because the attribute ***GroupCode*** cannot be removed from the determinant (the same subject can be taught by different teachers, so the FD ***Subject*** → ***Teacher*** is not correct);

the attribute ***Subject*** cannot be removed either (the group studies several subjects from different teachers, so the FD ***Group Code* → *Teacher*** is not correct).

- the relation is also in 3 NF.

Let us prove this:

in the only FD for a non-keyed attribute (see 1), this non-key attribute is nontransitively dependent on the primary key.

But this relation is not in BCNF, because there is a FD ***Teachers → Subjects*** (by condition, each teacher can teach only one subject), and the determinant of this FD is not a potential key.

Therefore, update anomalies may occur in relation ***Lessons***:

For example, if it is required to delete information about the *I-31 group*, then the information that the teacher *Melen* teaches the *Database* subject will be lost.

It is necessary to decompose this relation into projections that will be in the BCNF:

GP (***Group, Teacher***) and **PD (*Teacher, Discipline*)**

But these projections are not independent, because if we try to insert a new record into one of them, a contradiction may occur.

For example, we insert into the ***GP*** relation the cortege with some teacher, who teaches some subject, and, for example, in this relation there is already a cortege with another teacher, who teaches the same subject in this group.

A contradiction will appear (two teachers teach the same subject for one group). And this can only be verified in relation ***PD***.

Conclusion:

It is not always possible to achieve two goals at once

- to break it down into relations in the **BCNF** and get the independent relations.

Indeed, the **Lessons** relation is atomic, but atomicity is neither necessary nor sufficient for a good database layout. And if you want BCNF, then in some cases this complicates the database.

All this shows that, in practice, goals are best to avoid relation with overlapping potential keys, then 3 NF is sufficient.

But if it is difficult to do without them, then special update rules, in each case - their own.