

Database systems

Normal forms



An example of a bad model

SSN	Surnam	OfficeNo	City	Street	No	ZIP	Region	President_of_Region
1001	Novák	238	Liteň	Hlavní	10	26727	Středočeský	Rath
1001	Novák	238	Bystřice	Benešova	112	25751	Středočeský	Rath
1001	Novák	238	Benešov	Jiráskova	124	25601	Středočeský	Rath
1201	Řezáč	204	Praha	Ječná	21	12000	Praha	Bém
1221	Falc	216	Kladno	Uhelná	234	27201	Středočeský	Rath
1234	Lupínek	308	Kladno	Rudná	23	27202	Středočeský	Rath

Paradox on update:

- When Mr. Novak changes his office number, we will have to apply this change to all rows related to Mr. Novak.
- Similarly, a change of the president of the region „Středočeský kraj“ would initiate modification of multiple rows.

Příklad špatného modelu

SSN	Surnam	OfficeNo	City	Street	No	ZIP	Region	President_of_Region
1001	Novák	238	Liteň	Hlavní	10	26727	Středočeský	Rath
1001	Novák	238	Bystřice	Benešova	112	25751	Středočeský	Rath
1001	Novák	238	Benešov	Jiráskova	124	25601	Středočeský	Rath
1201	Řezáč	204	Praha	Ječná	21	12000	Praha	Bém
1221	Falc	216	Kladno	Uhelná	234	27201	Středočeský	Rath
1234	Lupínek	308	Kladno	Rudná	23	27202	Středočeský	Rath

Paradox on insert:

- When we will hire the first employee with address in the region of „Jihočeský kraj“, we would have to acquire the name of the president of the region „Jihočeský kraj“, if this column is required (mandatory).

Příklad špatného modelu

SSN	Surnam	OfficeNo	City	Street	No	ZIP	Region	President_of_Region
1001	Novák	238	Liteň	Hlavní	10	26727	Středočeský	Rath
1001	Novák	238	Bystřice	Benešova	112	25751	Středočeský	Rath
1001	Novák	238	Benešov	Jiráskova	124	25601	Středočeský	Rath
1201	Řezáč	204	Praha	Ječná	21	12000	Praha	Bém
1221	Falc	216	Kladno	Uhelná	234	27201	Středočeský	Rath
1234	Lupínek	308	Kladno	Rudná	23	27202	Středočeský	Rath

Paradox on delete:

- If Mr, Řezáče (the only one living in Prague) leaves our company, we will loose the information that the president of the region „Praha“ is Mr. Bém.

Konvention

Let $X = \{X_1, X_2, \dots, X_n\}$ be a table with columns X_1, X_2, \dots, X_n . Symbol X^\times denotes the cartesian product

$$X^\times = X_1 \times X_2 \times \dots \times X_n .$$

Symbol x^\times denotes an ordered tuple $x^\times \in X^\times$.

Functional dependency

Let $R(A)$ be a table, where A is a set of all columns of the table R .

Let X, Y be sets of columns $X, Y \subseteq A$.

Y is said to functionally depend on X iff there is at most one value $y^x \in Y^x$ for any value $x^x \in X^x$.

Functional dependence is marked as $X \rightarrow Y$.

$X \not\rightarrow Y$ means that Y is not functionally dependent on X .

Partial functional dependency

Let X and Y be some sets of columns of a table.

Let $X \rightarrow Y$.

$X \rightarrow Y$ is a **partial** functional dependence iff there exists a set of columns $X_1 \subset X$ such that $X_1 \rightarrow Y$.

Transitive functional dependency

Let X and Y be sets of some columns of a table, and C be a column of the same table such that $C \notin X$ and $C \notin Y$.

The column C is said to be transitively dependent iff the following two conditions hold concurrently:

- $X \rightarrow Y \rightarrow C$
- $Y \not\rightarrow X$.

First normal form

A table is in the first normal form iff the following conditions hold concurrently:

1. The columns are atomic, i.e. any cell of a table contains **exactly one** value from the respective domain.
2. All columns are regular. It means that rows do not have any hidden columns such as row IDs, object IDs, or hidden timestamps.
3. Rows are unique.

Příklad relace, která není v 1.NF:

Rodiče		
Otec	Matka	Děti
Jan	Marta	{Eliška, Lucie}
Jan	Marie	{Jiří, Josef}
Josef	Marta	{Leoš}

Second normal form

A table is in the second normal form if the following conditions hold concurrently:

1. The table is in the 1st normal form.
2. None non-key attribute (i.e. attribute that do not participate in any key) is partially functionally dependent on any key of the table.

Table {StudentId, CourseId, StudentName, Semester} is not in 2NF as:

- {IdStudenta, IdPredmetu} is a key
- Non-key column StudentName is functionally dependent on {StudentId}. It means it is partially functionally dependent on a key.
- Non-key column Semester is functionally dependent on {CourseId}. It means it is partially functionally dependent on a key.

Second normal form (2NF)

The table {StudentId, CourseId, StudentName, Semester} is not in 2NF because of:

- {StudentId, CourseId} is a key.
- Non-key column StudentName is functionally dependent on {StudentId}. Hence, it is partially functionally dependent on a key.
- Non-key attribute Semester is functionally dependent on {CourseId}. Hence, it is partially functionally dependent on a key.

Solution:

The relation formed by the table above shall be decomposed into three tables:

- {StudentId, CourseId}
- {StudentId, StudentName}
- {CourseId, Semester}

Third normal form (3NF)

A table is in the third normal form iff the following holds concurrently:

1. The table is in 2NF.
2. Non non-key attribute is transitively dependent on any key of the table.

E.g. the table {Citizen, City, CityPopulation} is not in 3NF, as:

- It is in 2NF, but
- Citizen is a key
- City is functionally dependent on Citizen but not vice-versa.
- CityPopulation is functionally dependent on City.
- Hence, CityPopulation is transitively functionally dependent of Citizen.

Třetí normální forma

E.g. the table {Citizen, City, CityPopulation} is not in 3NF, as:

- It is in 2NF, but
- Citizen is a key
- City is functionally dependent on Citizen but not vice-versa.
- CityPopulation is functionally dependent on City.
- Hence, CityPopulation is transitively functionally dependent of Citizen.

Solution:

The table will be split into:

- {Citizen, City}
- {City, CityPopulation}