

НОРМАЛИЗАЦИЈА НА БАЗИ НА ПОДАТОЦИ

Атрибутот В е функционално зависен од атрибутот А (А повлекува В, $A \rightarrow B$) ако во секој момент секоја вредност на А има придружено една вредност на В. На пример, за релациите

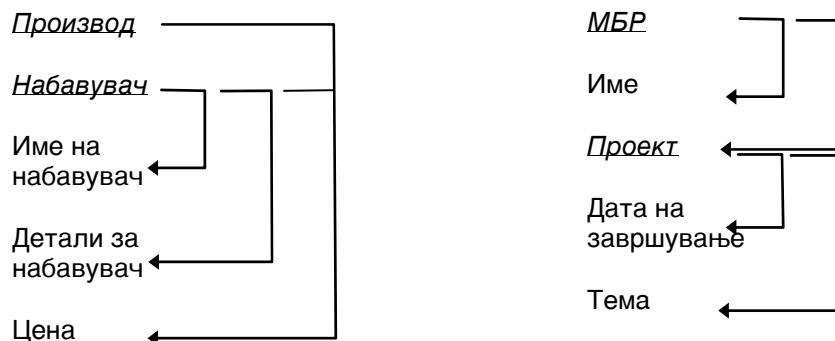
Набавка

<i>Производ</i>	<i>Набавувач</i>	Име на набавувач	Детали за набавувач	Цена
...

Вработен

<i>Мбр</i>	Име	<i>Проект</i>	Дата на завршување	Тема
...

функционалните зависимости би можеле да се претстават вака:



1. Прва нормална форма

Релацијата е во прва нормална форма ако нема повторувачки групи податоци. Ако има повторувачка група податоци, таа се става во посебна табела.

На пример табелата

Вработен

<i>Мбр</i>	Детали за вработен	Дете 1	Дете 2	Дете 3	...
...

Се двои на две табели:

Вработен

<i>Мбр</i>	Детали за вработен
...	...

Деца

<i>Мбр</i>	Дете
...	...

Првата нормална форма кажува дека сите случувања (записи податоци) во релацијата мораат да содржат ист број полиња, т.е. променливо повторливите полиња мора да се исклучат од релациите.

2. Втора нормална форма

Релацијата е во втора нормална форма ако секој атрибут (кој не е клуч) зависи од комплетниот клуч.

Набавка

<i>Производ</i>	<i>Набавувач</i>	Име на набавувач	Детали за набавувач	Цена
...

Поради тоа што *Име на набавувач* не зависи од *Производ*, релацијата не е во 2-ра нормална форма и треба да се подели на:

Набавувач

<i>Набавувач</i>	Име на набавувач	Детали за набавувач
...

Набавка

<i>Производ</i>	<i>Набавувач</i>	Цена
...

Втората нормална форма (као и третата) се однесуваат на зависностите меѓу атрибутите што се кучеви и оние што не се. Атрибутите што не се клучеви мораат да зависат од клучот.

Забележи дека втората нормална форма е релевантна само во случаите кога клучот во релацијата се состои од повеќе атрибути.

3. Трета нормална форма

Релацијата е во трета нормална форма ако во неа нема транзитивни зависности ($A \rightarrow B, B \rightarrow C$).

Третата нормална форма е нарушена во релацијата:

Вработен

<u>Мбр</u>	Оддел	Адреса на оддел
...

бидејќи имаме транзитивна зависност $Мбр \rightarrow$ Оддел \rightarrow Адреса на оддел. Проблемите со ваквата релација се:

- Адресата на одделот се повторува за секој негов вработен,
- Ако адресата на одделот се смени тоа мора да се направи во повеќе записи,
- Ако Одделот нема свои вработени, ќе ја немаме неговата адреса (празна табела).

Табелата треба да се подели на две табели:

Вработен

<u>Мбр</u>	Оддел
...	...

Оддел

<u>Оддел</u>	Адреса на оддел
...	...

Малку посилно барање е таканаречената Boyce/Codd-овата нормална форма (BCNF) којашто може да се дефинира вака:

Секој не-клучен атрибут зависи од клучот, од целиот клуч и само од клучот. На пример, во релациите:

Вработен

<u>Мбр</u>	Детали за вработен	<u>Проект</u>	Дата на завршување	Тема
...

Проект зависи од Мбр, а Дата на завршување и Тема зависат од Проект и транзитивно од Мбр. Релацијата треба да се подели на две релации:

Вработен

<u>Мбр</u>	Детали за вработен	<u>Проект</u>
...

Проект

<u>Проект</u>	Дата на завршување	Тема
...

Следната табела е во 3-та нормална форма, но не е во BCNF:

КиноПрет

<u>Филм</u>	Кино	<u>Општина</u>
...

Тука ги имаме зависностите (Филм, Општина) \rightarrow Кино (под услов да не се прикажува истиот филм во две кина од иста општина) и Кино \rightarrow Општина. Оваа релација не ја задоволува BCNF нормалната форма бидејќи во зависноста Кино \rightarrow Општина, атрибутот Кино не е клуч. Од друга страна, во релацијата немаме транзитивна зависност и таа е во трета нормална форма. Ако би сакале да ја задоволиме BCNF, релацијата треба да се подели во две:

Кино

<u>Кино</u>	Општина
...	...

Претстава

<u>Кино</u>	<u>Филм</u>
...	...

Ваквото разбивање не секогаш ги сочувува информациите содржани во почетната релација. Додатно, BCNF се смета за премногу "строга" и вообичаено таа не се форсира и се заменува со 3-тата нормална форма.

Третата нормална форма го комплетира односот меѓу атрибутите што се клуч и оние што не се, со барањето *секој атрибут што не е дел од клучот да зависи од клучот, од целиот клуч и од ништо друго.*

4. Четврта нормална форма

Мултивредноста зависност на атрибут В од атрибут А во релација од најмалку 3 атрибути е кога една вредност на А повлекува повеќе вредности на В и истовремено В не зависи од друг атрибут. Тогаш пишуваме $A \rightarrow\rightarrow B$.

Релацијата е во 4-та нормална форма ако при мултивредноста зависност само меѓу еден атрибут (обично клучот) и повеќе други атрибути, тогаш тие атрибути имаат директна зависност меѓусебе. Со други зборови, ако има мултивредносна зависност во некоја релација $A \rightarrow\rightarrow B$ и $A \rightarrow\rightarrow C$, тогаш за релацијата да биде во 4-та нормална форма мора да има и зависност меѓу В и С.

Четвртата нормална форма кажува дека во релацијата не треба да има 2 или повеќе зависности од тип $\rightarrow\rightarrow$ коишто се независни. На пример, релацијата:

<i>Вработен</i>	<i>Стручност</i>	<i>Јазик</i>
...

не е во 4-та нормална форма бидејќи имаме $\text{Вработен} \rightarrow\rightarrow \text{Стручност}$ и $\text{Вработен} \rightarrow\rightarrow \text{Јазик}$, но истовремено Стручност и Јазик немаат директна функционална зависност. Кога би имале зависност меѓу Стручност и Јазик , релацијата би ја задоволувала 4-тата нормална форма. Се разбира, оваа релација треба да се подели на 2 релации

<i>Вработен</i>	<i>Стручност</i>
...	...

<i>Вработен</i>	<i>Јазик</i>
...	...

Релациите што не се во 4-та нормална форма (понатаму ќе видиме и во 5-та) растат на мултипликативен начин, додека нормализираните релации растат на адитивен начин. На пример, ако имаме m -стручности и k -јазици за n -вработени, тогаш првата табела може да има максимално $n*m*k$ записи (ако секој вработен ја има секоја стручност и го знае секој јазик), додека вторите две вкупно максимално може да имаат $n*m+n*k = n*(m+k)$ записи.

5. Петта нормална форма

Нормализацијата ги дели релациите на делови, со што се добиваат повеќе релации. Притоа е логично да може да се загубат некои информации. Интуитивно, релацијата е во 5-та нормална форма ако таа не може да се реконструира со соединување на нејзините делови. Овој услов се нарекува зависност на соединување.

Релацијата е во 5-та нормална форма ако секоја зависност на соединување ги вклучува во себе клучевите.

Значи релацијата е во 5-та нормална форма кога информациите што таа ги содржи не може да бидат реконструирани од релациите добиени со нејзино разбивање. Случајот кога помалите релации имаат ист клуч не се разгледува и тогаш се смета дека оригиналната релација е во 5-та нормална форма.

На пример, релацијата

<i>Агент</i>	<i>Компанија</i>	<i>Производ</i>
...

каде што агентот ја претставува компанијата што го произведува производот што агентот го продава. Сега ако, на пример, имаме два записа:

<i>Агент</i>	<i>Компанија</i>	<i>Производ</i>
Перо	Форд	автомобил
Перо	Мерцедес	камион

ние би загубиле дел од информациите со разбивање на релацијата. На пример со разбивањето:

<i>Агент</i>	<i>Компанија</i>		<i>Агент</i>	<i>Производ</i>		<i>Компанија</i>	<i>Производ</i>
Перо	Форд		Перо	автомобил		Форд	автомобил
Перо	Мерцедес		Перо	камион		Мерцедес	камион

Изгубена е информацијата дека Перо продава само автомобили за Форд или само камиони за Мерцедес. Значи почетната релацијата е во 5-та нормална форма.

Ако воведеме додатен услов да ако агентот продава некој производ и ја претставува компанијата што го прави тој производ, тогаш тој го продава тој производ за таа компанија, тогаш релацијата не е во 5-та нормална форма бидејќи сите информации може да се реконструираат од

<i>Агент</i>	<i>Компанија</i>		<i>Агент</i>	<i>Производ</i>		<i>Компанија</i>	<i>Производ</i>
...

Како втор пример да ја разгледаме релацијата:

Предмети

<i>Предмет</i>	<i>Професор</i>	<i>Книга</i>
...

Оваа релација е во 4-та нормална форма бидејќи покрај зависностите *Предмет* →→ *Професор* и *Предмет* →→ *Книга*, имаме и зависност *Професор* →→ *Книга*. Од друга страна, оваа релација не ја задоволува 5-тата нормална форма бидејќи сите информации што таа ги содржи може да се реконструираат од трите релации

ПредПроф

<i>Предмет</i>	<i>Професор</i>
...	...

ПредКнига

<i>Предмет</i>	<i>Книга</i>
...	...

ПрофКнига

<i>Професор</i>	<i>Книга</i>
...	...

4-тата и 5-тата нормална експлицитно се бават со намалувањето на редуванцата во податоците. Во горниот пример е битно да се забележи дека бројот на записи во трите релации расте адитивно, додека во почетната релација бројот на записи може да расте мултипликативно. Ова особено доаѓа до израз кога расте бројот на записите што треба да се внесат.

Петтата нормална форма во основа ги опфаќа сите останати случаи на одбегнување на мултипликативен раст на записите во релациите (со нивно разбивање на помали релации) кои што 4-тата нормална форма не ги покрива.

Релации што не ги задоволуваат 4-тата и 5-тата нормална форма се ретки во практиката бидејќи искусните дизајнери на базите на податоци обично мултивредносните зависности меѓу атрибутите ги ставаат веднаш во посебни релации интентивно чувствувајќи дека така ја намалуваат редуванцата (повторувањата) во податоците. Исто така, ако некој стриктно ја интерпретира првата нормална форма ставајќи ги веднаш мултивредносните зависности во посебни релации, повторно може да одбегне појавување на релации што не се во 4-та или 5-та нормална форма.

Нормализацијата во базите на податоци сигурно не ги отстранува сите редуванци низ податоците. На пример, во случајот на агент, компанија, производ без додатни услови, ако Перо продава автомобили и за Шкода, податокот автомобил секако мора да биде повторен во табелата.

На крај, треба да се каже дека процесот на нормализација не е секогаш најдобра опција (зголемува број на релации) ако се земат предвид сите аспекти на чување и обработка на податоците. Посебно треба да се има предвид пребарувањето на податоците коешто често пати може да биде подолго кај нормализираните бази на податоци.

ДВЕ ПРАВИЛА ЗА РЕШАВАЊЕ НА АСОЦИЈАЦИИТЕ МЕЃУ ЕНТИТЕТИ

Ентитетите и асоцијациите меѓу нив се трансформираат во релации речиси праволинейски. Две правила за кои треба да се води сметка се односите меѓу ентитетите од тип $1 \rightarrow \rightarrow M$ и $M \leftrightarrow \leftrightarrow M$.

- $1 \rightarrow M$ (еден према многу)** Односот на ентитетите од тип $1 \rightarrow \rightarrow M$ се решава така што клучот на релацијата што одговара на ентитетот 1 се повтори во релацијата што одговара на ентитетот M.

На пример, Вработен $\rightarrow \rightarrow$ Дете, Опшина $\rightarrow \rightarrow$ Граѓанин или Факултет $\rightarrow \rightarrow$ Студент.

Факултет

<u>ШифраФ</u>	Информации за факултет
...	...

Студент

<u>Индекс</u>	<u>ШифраФ</u>	Информации за студент
...

Понекогаш на ист начин се решава и односот $M \leftrightarrow \leftrightarrow M$ кога атрибутите на едниот ентитет не се битни. Тогаш тој ентитет може да се гледа како атрибут на другиот ентитет. Таков пример е односот меѓу ентитетите Вработен $\leftrightarrow \leftrightarrow$ Јазик. За ентитетот Јазик што работникот го зборува, немаме атрибути што би биле интересни да се чуваат. Така Јазик-от станува само атрибут на ентитетот Вработен, но сепак мора да оди во посебна релација за да бидат задоволени нормалните форми.

Вработен

<u>МБР</u>	Информации за вработен
...	...

ВработенЈазик

<u>МБР</u>	<u>Јазик</u>	Информации зависни и од вр. и од j.
...

- $M \leftrightarrow \leftrightarrow M$ (многу према многу)** Односот на ентитетите од тип $M \leftrightarrow \leftrightarrow M$, пришто атрибутите и од двата ентитета се од интерес се решава така што се формираат релациите за двата ентитета и *трета релација* за нивниот однос што ги содржи клучевите од двата ентитета (тоа е клуч на релацијата) и атрибути заависни од овие два клуча.

На пример, Добавувач $\leftrightarrow \leftrightarrow$ Дел, Вработен $\leftrightarrow \leftrightarrow$ Проект или Студент $\leftrightarrow \leftrightarrow$ Предмет.

Предмет

<u>Шифра</u>	Информации за предмет
...	...

Студент

<u>Индекс</u>	Информации за студент
...	...

Испит

<u>Шифра</u>	<u>Индекс</u>	Дата	Оценка
...