



Базы от данни

III. Релационен модел на данни. Нормализация.

Проектиране на схеми на релации

- ▶ От предишни лекции разбрахме, че основните аномалии са:
 - ▶ Излишества – когато се повтаря информацията без да е необходимо
 - ▶ Аномалии при обновяване – когато направим промени в един кортеж, а забравим да направим същите промени в свързаните с него кортежи
 - ▶ Аномалии при изтриване – когато изтрием даден ред и това доведе до загуба на информация
 - ▶ Например ако имаме актьор който участва във филм и ние искаме да изтрием актьора, а като ефект да получим и изтрита информация за филма в който е участвал актьора.

StarMovies(title, year, length, filtype, studioName, starName)

Декомпозиране на релации

- ▶ Начин да избегнем излишества е чрез декомпозиране на релации.
- ▶ Декомпозирането на една релация R означава разделяне на атрибутите на релацията R в две нови релации
- ▶ Например: Нека ни е дадена релацията $R(A_1, A_2, \dots, A_n)$, можем да декомпозираме релацията R до две релации $S(B_1, B_2, \dots, B_m)$ и $T(C_1, C_2, \dots, C_k)$, такива че:
 - ▶ $\{A_1, A_2, \dots, A_n\} = \{B_1, B_2, \dots, B_m\} \cup \{C_1, C_2, \dots, C_k\}$
 - ▶ Кортежите в релацията S са проекция по атрибутите B_1, B_2, \dots, B_m на релацията R , а тези на релацията T по атрибутите C_1, C_2, \dots, C_k на релацията R

Декомпозиране на релации

- ▶ Проекция по атрибутите B_1, B_2, \dots, B_m на релацията R означава, че за всеки кортеж от релацията R вземаме само компонентите (стойностите) на кортежите на релацията в атрибутите B_1, B_2, \dots, B_m . Тези компоненти формират кортежите на новата релация S .
- ▶ Понеже релациите са множества е възможно при проекцията на релацията R до релацията S по атрибутите B_1, B_2, \dots, B_m да получим от два различни кортежа за релацията R , два еднакви кортежа в релацията S . Ако се получи така в релацията S попада само един кортеж (защото е множество)
- ▶ Пример: Нека разгледаме релацията

`StarMovies(title, year, length, filtype, studioName, starName)`

Декомпозиране на релации

- ▶ `StarMovies(title, year, length, filtype, studioName, starName)`
- ▶ Очевидно може да се получат излишества в нея.
- ▶ Ще я декомпозираме до две релации:
 - ▶ `Movies1(title, year, length, filtype, studioName)`
 - ▶ `StarMovieI(title, year, starName)`
- ▶ Ако разгледаме ново-получените релации, лесно се установява, че проблемите с излишествата са разрешени.
- ▶ Ако в първия вариант `StarMovies` са възможни и излишества и аномалии при изтриване и аномалии при обновяване, то при декомпозираните релации това не може да се получи.
- ▶ Дайте пример за аномалии при релацията: `StarMovies(title, year, length, filtype, studioName, starName)`?

Декомпозиране на релации - Аномалии

StarMovies(title, year, length, filtype, studioName, starName)

- ▶ Излишество: Ако в един филм участват десет актьора, за всеки актьор ще повтаря данните за филма
- ▶ Аномалия при обновяване: Както казахме ако в един филм участват десет актьора, за всеки актьор ще повтаря данните за филма. Ако обновим в един от тези кортежи например дължината на филма, а забравим да го направим в останалите девет това е аномалия!
- ▶ Аномалия при изтриване: Аналогично, ако в един филм участва един актьор и ние искаме да изтрием актьора, това ще доведе до загуба за информацията за филма.

Декомпозиране на релации - Аномалии

Да разгледаме декомпозираните релации:

- ▶ Movies1(title, year, length, filtype, studioName)
- ▶ StarMovie I (title, year,starName)

При тях възможни ли са изброените по-долу аномалии?

- ▶ Излишество: Ако в един филм участват десет актьора, за всеки актьор ще повтаря данните за филма (X)
- ▶ Аномалия при обновяване: Ако обновим в един от повтарящите се за филма кортежи, например дължината на филма, а забравим да го направим в останалите девет това е аномалия (X)
- ▶ Аномалия при изтриване: Ако в един филм участва един актьор и ние искаме да изтрием актьора, това ще доведе до загуба за информацията за филма(X)

Нормална форма на Бойс-Код

- ▶ Целта на декомпозицията е да замени релация с други, които да не позволяват аномалии. Има просто условие, което ако една релация е подчинена на него то в съответната релация не може да възникнат аномалии от изброените по-горе.
- ▶ Това условие се нарича нормална форма на Бойс-Код (НФБК)
- ▶ Казваме, че една релация се намира в НФБК, тогава и само тогава когато ако $A_1, A_2, \dots, A_n \rightarrow V$ е нетривиална ФЗ която е в сила за R , то $\{A_1, A_2, \dots, A_n\}$ е супер-ключ за R
- ▶ Или казано по друг начин лявата страна на всяка нетривиална ФЗ, която е в сила за R трябва да съдържа ключ на релацията R
- ▶ Ако се случи че има ФЗ, които са в сила за релацията но не отговарят на посоченото по-горе правило (НФБК), тогава трябва да декомпозираме за да приведем релацията до НФБК

Нормална форма на Бойс-Код

- ▶ Разширена формулировка на НФБК е казваме, че една релация се намира в НФБК, тогава и само тогава когато ако има нетривиална ФЗ: $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ която е в сила за R , то $\{A_1, A_2, \dots, A_n\}$ е супер-ключ за R
- ▶ Това изискване е еквивалентно на оригиналната формулировка на НФБК, от правилото за разделяне и комбиниране, като се изисква поне едно от B -тата (B_j) да не е сред A_i , т.е. да има поне една нетривиална ФЗ.
- ▶ Пример: Релацията
- ▶ `StarMovies(title, year, length, filtype, studioName, starName)`
- ▶ Не е в НФБК
- ▶ За да проверим това, нека първо да определим кой е ключа на релацията

Нормална форма на Бойс-Код

- ▶ Ключът е {title, year, starname}
- ▶ Тогава всяко множество от атрибути, което съдържа тези три атрибута е супер-ключ
- ▶ За релацията е в сила ФЗ:
 - ▶ title, year - > length, filmtype, studioName
 - ▶ Също така знаем че Title, year не определя функционално starname
 - ▶ Ако релацията беше в НФБК, то в лявата част на ФЗ щеше да бъде супер-ключ
 - ▶ В случая това не е така
- ▶ Пример2: Релацията
 - ▶ Movies1(title, year, length, filtype, studioName)
- ▶ Тя е в НФБК
- ▶ За нея е в сила ФЗ
 - ▶ title, year - > length, filmtype, studioName
- ▶ Лявата част съдържа ключа, следователно релацията е в НФБК

Твърдение

- ▶ Ще докажем че всяка релация, която има само два атрибута се намира в НФБК
- ▶ Нека $R(A,B)$
 - ▶ Ако нямаме нетривиални ФЗ, т.е. всички ФЗ които се в сила за релацията са тривиални, тогава релацията се намира в НФБК, тъй като само не тривиалните ФЗ могат да нарушат правилото на НФБК. Единственият ключ за релацията ще бъде $\{A,B\}$
 - ▶ В случай че $A \rightarrow B$, но не е в сила че $B \rightarrow A$, тогава A е единственият ключ за релацията и всички нетривиални ФЗ, които са в сила за релацията го съдържат в лявата си част, т.е. релацията е в НФБК
 - ▶ Аналогично ако $B \rightarrow A$, но не е в сила че $A \rightarrow B$
 - ▶ В случай че $A \rightarrow B$ и $B \rightarrow A$, тогава и атрибута A и атрибута B могат да бъдат ключ за релацията. Всички ФЗ съдържат ключ в лявата си част и правилото, следователно релацията е в НФБК

Твърдение - пояснение

- ▶ Ако има повече от един ключ в релацията, НФБК изисква някакъв ключ да се съдържа в лявата част на правилото на всички ФЗ, които са в сила за релацията (не първични ключ)
- ▶ Пример за такива атрибути в релация са – ID, EGN, SSN
- ▶ Всяка релационна схема може да бъде сведена до НФБК, чрез подходящи декомпозиции, така че:
 - ▶ Декомпозираните релации да бъдат също в НФБК
 - ▶ Данните в оригиналната релация са представени напълно достоверно в декомпозираните релации, т.е. да можем да „възстановим“ данните в оригиналната релация от данните в декомпозираните релации
- ▶ От правилото за релациите с два атрибути, може едва ли не да заключим, че е напълно достатъчно да сведем всяка релация до множество от релации с по два атрибута. Това не е достатъчно, защото второто условие за възстановяване няма да бъде изпълнено.

Алгоритъм за свеждане към НФБК

- ▶ Първо се търсят нетривиални ФЗ: $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ такива че да нарушават НФБК, т.е. лявата част да не съдържа ключа
- ▶ По правилото за комбиниране добавяме в дясната част всички атрибути, които са функционално определени от A_1, A_2, \dots, A_n
- ▶ Правим декомпозиция на релацията на две релации.
 - ▶ Първата релация съдържа атрибутите от лявата част на функционалната зависимост и тези атрибути,, които не са функционално определени от тях
 - ▶ Втората релация съдържа атрибутите от лявата част на ФЗ и атрибутите от дясната част

Алгоритъм за свеждане към НФБК - Пример

- ▶ StarMovies(title, year, length, filtype, studioName, starName)
- ▶ ФЗ, която нарушава правилото е

title, year - > length, filmtype, studioName

- ▶ По горе-споменатият алгоритъм разделяме релацията на две релации
 1. Movies1(title, year, length, filtype, studioName)
 2. StarMovies(title, year, starname)

Възстановяване на информацията след декомпозиция

- ▶ Нека разгледаме релацията $R(A, B, C)$ и ФЗ: $B \rightarrow C$, която допускаме че нарушава НФБК. Има два варианта:
 1. Да съществува транзитивна зависимост $A \rightarrow B$, в този случай $\{A\}$ е единственият ключ, а лявата страна на ФЗ $B \rightarrow C$ не е супер ключ.
 2. ФЗ $B \rightarrow C$ да е единствената нетривиална ФЗ. В този случай единственият ключ е $\{A, B\}$ и отново лявата страна на ФЗ $B \rightarrow C$ не е супер-ключ.
- ▶ Декомпозицията базирана на ФЗ $B \rightarrow C$, разделя атрибутите на две схеми $\{A, B\}$ и $\{B, C\}$
- ▶ Нека t е кортеж от R , така че $t = (a, b, c)$. Като приложим декомпозицията t ще се проектира като (a, b) за схемата $\{A, B\}$ и (b, c) за схемата $\{B, C\}$
- ▶ Възможно е да съединим кортежа от $\{A, B\}$ с кортежа от $\{B, C\}$ по съвпадащите стойности на атрибута B . При съединението на двата кортежа се получава първоначалния кортеж $= (a, b, c)$.

Възстановяване на информацията след декомпозиция

- ▶ Не можем да сме сигурни обаче, че R може да бъде наистина възстановена след декомпозицията.
- ▶ Например нека да вземем кортежите $t=(a, b, c)$ и $v=(d, b, e)$
- ▶ Ако проектираме кортежите по двете схеми получаваме за кортежа t - (a, b) (b, c) , съответно за схемите $\{A, B\}$ и $\{B, C\}$ и за кортежа v (d, b) (b, e) съответно за схемите $\{A, B\}$ и $\{B, C\}$
- ▶ Ако се опитаме да възстановим първоначалната релация обаче и съединим проектираните кортежи по съвпадащите стойности на атрибута B получаваме кортежа (a, b, e) , който реално не е кортеж на релацията R
- ▶ Докато е в сила обаче ФЗ: $B \rightarrow C$, такъв кортеж не можем да получим. Ако кортежите t и v съвпадат в атрибута B , те трябва да съвпадат и в C . Следователно самата релация R не може да съществуват такива кортежи като t и v . Докато ФЗ е в сила съединението на двата проектирани кортежа не може да доведе до лъжлив кортеж.

Възстановяване на информацията след декомпозиция

- Дори нещо повече, всеки кортеж получен чрез съединение със сигурност е от релацията R . Изводът е че за всяка ФЗ, която нарушава НФБК, ако декомпозираме релацията по метода за декомпозиция до НФБК, тогава оригиналната релация може да бъде точно възстановена чрез съединение на проектираните кортежи от новите релации по всички възможни начини. Ако направим декомпозиция, без да спазваме метода за декомпозиция, получените проекции на кортежи няма да могат да бъдат възстановени до оригиналната релация.
- Пример $R(A, B, C)$ и ФЗ $B \rightarrow C$ не е в сила за R . Тогава

A	B	C
1	2	3
4	2	5

A	B
1	2
4	2

B	C
2	3
2	5

(1, 2, 3) (1, 2, 5) (4, 2, 3) (4, 2, 5)

Нормализация

- ▶ Нормализация наричаме процесът на дизайн на базата от данни, при които се отстраняват аномалии. При нормализацията
 - ▶ Се изключват повтаряща се информация в таблиците.
 - ▶ Декомпозират се таблиците до няколко таблици
 - ▶ Минимизират се аномалиите при съхраняване, изтриване и промяна на данните.
- ▶ Има четири нормални форми и НФБК

Нормални форми

- ▶ **Първа нормална форма (1НФ)** – изисква всеки компонент в кортежите да е атомарна стройност.
- ▶ **Втора нормална форма (2НФ)** – изисква релацията да е в 1НФ и всеки атрибут да е функционално зависим от атрибутите, съставляващи първичния ключ.
- ▶ **Трета нормална форма (3НФ)** – изисква релацията да е в 2НФ и ако $A_1, A_2, \dots, A_n \rightarrow V$ е нетривиална ФЗ която е в сила за R , то или $\{A_1, A_2, \dots, A_n\}$ да е супер-ключ за R или V да е член на някой ключ.
- ▶ **Нормална форма на Бойс-Код (НФБК)** – изисква релацията да е в 3НФ и само ако $A_1, A_2, \dots, A_n \rightarrow V$ е нетривиална ФЗ, която е в сила за R , то $\{A_1, A_2, \dots, A_n\}$ да е супер-ключ за R .
- ▶ **Четвърта нормална форма (4НФ)** изисква релацията да е в НФБК и ако $A_1, A_2, \dots, A_n \twoheadrightarrow V$ е нетривиална МФЗ, която е в сила за R , то $\{A_1, A_2, \dots, A_n\}$ да е супер-ключ за R .
- ▶ **Пета нормална форма (5НФ)** изисква релацията да е в 4НФ и след декомпозиция за получените таблици да е в сила, че ако се съединят няма да има загуба на кортежи или да се генерират нови несъществуващи кортежи.

Задача

DreamHome Lease

DreamHome Lease

DreamHome Lease

DreamHome Lease

<p>Client Number <u>CR76</u> (Enter if known)</p> <p>Full Name <u>John Kay</u> (Please print)</p>	<p>Property Number <u>PG4</u></p> <p>Property Address <u>6 Lawrence St, Glasgow</u></p>
<p>Monthly Rent <u>350</u></p> <p>Rent Start <u>01/07/12</u></p> <p>Rent Finish <u>31/08/13</u></p>	<p>Owner Number <u>C040</u> (Enter if known)</p> <p>Full Name <u>Tina Murphy</u> (Please print)</p>

Книга: Database systems.A practical approach to design, implementation and management,
Автори: Thomas Connolly, Carolyn Begg

Ненормализиран модел

ClientRental

clientNo	cName	propertyNo	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	John Kay	PG4	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
		PG16	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	50	CO93	Tony Shaw
CR56	Aline Stewart	PG4	6 Lawrence St, Glasgow	1-Sep-11	10-June-12	350	CO40	Tina Murphy
		PG36	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
		PG16	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

Figure 14.10 ClientRental unnormalized table.

1НФ

Първа нормална форма (1НФ) – изисква всеки компонент в кортежите да е атомарна стойност.

ClientRental

clientNo	propertyNo	cName	pAddress	rentStart	rentFinish	rent	ownerNo	oName
CR76	PG4	John Kay	6 Lawrence St, Glasgow	1-Jul-12	31-Aug-13	350	CO40	Tina Murphy
CR76	PG16	John Kay	5 Novar Dr, Glasgow	1-Sep-13	1-Sep-14	450	CO93	Tony Shaw
CR56	PG4	Aline Stewart	6 Lawrence St, Glasgow	1-Sep-11	10-Jun-12	350	CO40	Tina Murphy
CR56	PG36	Aline Stewart	2 Manor Rd, Glasgow	10-Oct-12	1-Dec-13	375	CO93	Tony Shaw
CR56	PG16	Aline Stewart	5 Novar Dr, Glasgow	1-Nov-14	10-Aug-15	450	CO93	Tony Shaw

Figure 14.11 First Normal Form ClientRental relation.

ФЗ за модела

ClientRental

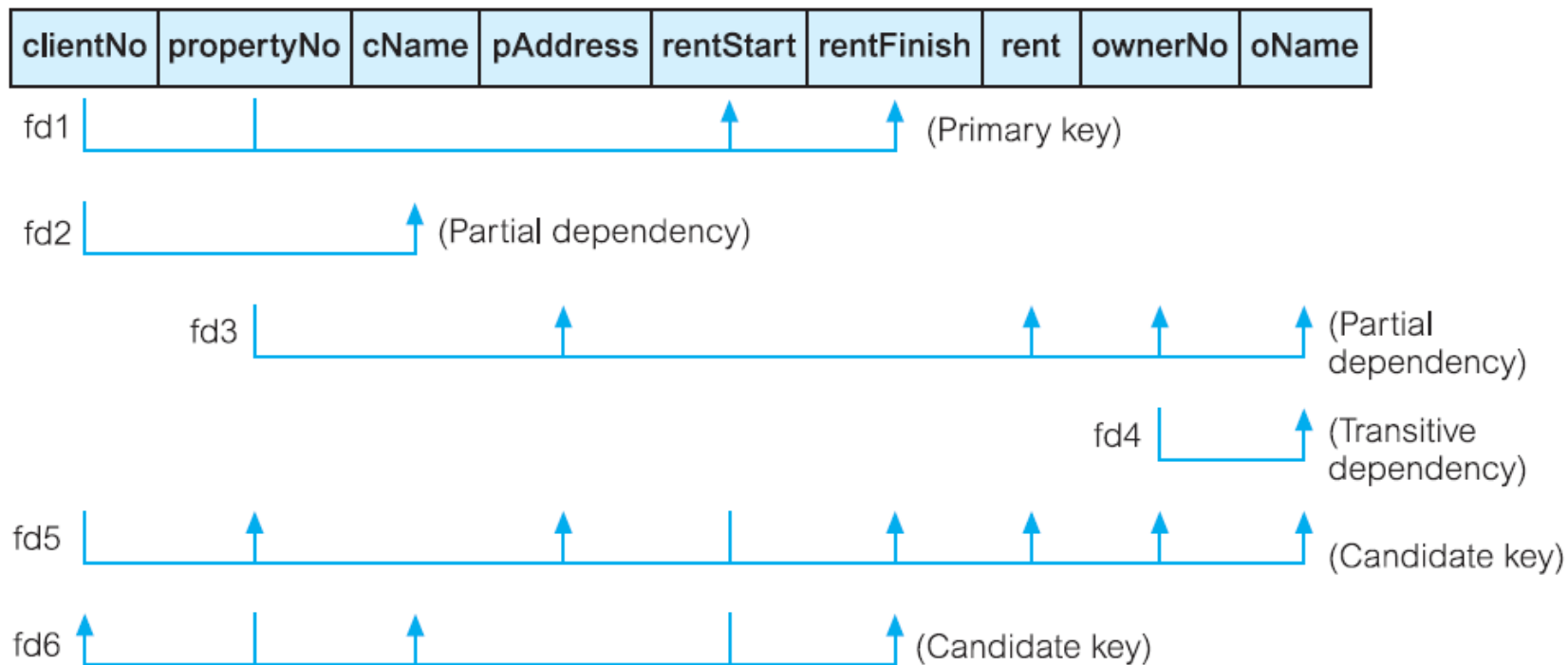


Figure 14.12 Functional dependencies of the ClientRental relation.

Ф3 за модела

- fd1 clientNo, propertyNo → rentStart, rentFinish
- fd2 clientNo → cName
- fd3 propertyNo → pAddress, rent, ownerNo, oName
- fd4 ownerNo → oName
- fd5 clientNo, rentStart → propertyNo, pAddress, rentFinish, rent, ownerNo, oName
- fd6 propertyNo, rentStart → clientNo, cName, rentFinish

2НФ

Втора нормална форма (2НФ) – изисква релацията да е в 1НФ и всеки атрибут да е функционално зависим от атрибутите, съставляващи първичния ключ.

Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

Rental

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-12	31-Aug-13
CR76	PG16	1-Sep-13	1-Sep-14
CR56	PG4	1-Sep-11	10-Jun-12
CR56	PG36	10-Oct-12	1-Dec-13
CR56	PG16	1-Nov-14	10-Aug-15

PropertyOwner

propertyNo	pAddress	rent	ownerNo	oName
PG4	6 Lawrence St, Glasgow	350	CO40	Tina Murphy
PG16	5 Novar Dr, Glasgow	450	CO93	Tony Shaw
PG36	2 Manor Rd, Glasgow	375	CO93	Tony Shaw

14 Second normal form relations derived from the ClientRental relation.

fd2 clientNo → cName (Primary key)

(Primary key)

fd1 clientNo, propertyNo → rentStart, rentFinish (Primary key)

(Primary key)

(Candidate key)

(Candidate key)

fd3 propertyNo → pAddress, rent, ownerNo, oName (Primary key)

(Primary key)

(Transitive dependency)

3НФ

Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

Rental

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-12	31-Aug-13
CR76	PG16	1-Sep-13	1-Sep-14
CR56	PG4	1-Sep-11	10-Jun-12
CR56	PG36	10-Oct-12	1-Dec-13
CR56	PG16	1-Nov-14	10-Aug-15

Трета нормална форма (3НФ)

– изисква релацията да е в 2НФ и ако $A_1, A_2, \dots, A_n \rightarrow B$ е нетривиална ФЗ която е в сила за R, то или $\{A_1, A_2, \dots, A_n\}$ да е супер-ключ за R или B да е член на някой ключ

PropertyForRent

propertyNo	pAddress	rent	ownerNo
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner

ownerNo	oName
CO40	Tina Murphy
CO93	Tony Shaw

14.15 Third normal form relations derived from the PropertyOwner relation.



НФБК

Client

clientNo	cName
CR76	John Kay
CR56	Aline Stewart

Rental

clientNo	propertyNo	rentStart	rentFinish
CR76	PG4	1-Jul-12	31-Aug-13
CR76	PG16	1-Sep-13	1-Sep-14
CR56	PG4	1-Sep-11	10-Jun-12
CR56	PG36	10-Oct-12	1-Dec-13
CR56	PG16	1-Nov-14	10-Aug-15

Нормална форма на Бойс-Код (НФБК) – изисква релацията да е в 3НФ и само ако $A_1, A_2, \dots, A_n \rightarrow B$ е нетривиална ФЗ, която е в сила за R, то $\{A_1, A_2, \dots, A_n\}$ да е супер-ключ за R

PropertyForRent

propertyNo	pAddress	rent	ownerNo
PG4	6 Lawrence St, Glasgow	350	CO40
PG16	5 Novar Dr, Glasgow	450	CO93
PG36	2 Manor Rd, Glasgow	375	CO93

Owner

ownerNo	oName
CO40	Tina Murphy
CO93	Tony Shaw

14.15 Third normal form relations derived from the PropertyOwner relation.



Задача

- ▶ Дадена е релацията **Orders** с информация за продукти цена на продукт, ДДС (VAT), клиенти, количество поръчани продукти и дата на поръчка

pID	pName	cID	cName	oDate	itPrice	Amount	VAT	NTotal	GTotal
-----	-------	-----	-------	-------	---------	--------	-----	--------	--------

- ▶ Известно е че:
 - ▶ Стойността на VAT може да варира за различните продукти (Products)
 - ▶ $GTotal = NTotal + VAT$
 - ▶ Поръчките на клиентите, направени в един и същ ден (дата) се обединяват. Има само една поръчка за клиент за дадена дата.
- ▶ Задача: Нормализирайте релацията до НФБК.

Многозначни ФЗ

- ▶ Нека разгледаме релацията `stars_in(name, street, city, title, year)`
- ▶ Ако предположим, че един актьор може да има повече от един адрес с различен град и улица, тогава информацията за адреса води до излишества в релацията `stars_in`
- ▶ Няма причина да свързваме филма с адреса на актьора. Единственият начин да изразим факта, че адреса на актьора и филма са независими е да използваме многозначни функционални зависимости.
- ▶ Казваме че $A_1, A_2, A_3, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ (многозначно определя) тогава и само тогава когато, многозначната зависимост е в сила за релацията R , то за тези кортежите които съвпадат по атрибутите $A_1, A_2, A_3, \dots, A_n$, множеството от стойности за B_j са независими от множеството от стойности за другите атрибути, които не са нито A -та нито B -та
- ▶ Ако имаме два кортежа U и T на релацията R , които съвпадат по всички A -та, можем да намерим такъв кортеж V в релацията R :
 - ▶ Така че да съвпада с кортежите U и T по A -тата
 - ▶ С кортежа T по B -тата
 - ▶ С кортежа U по всички останали атрибути, които не са нито A -та нито B -та

Многозначни ФЗ

- ▶ Ако имаме два кортежа **U** и **T** на релацията **R**, които съвпадат по всички **A**-та, можем да намерим такъв кортеж **V** в релацията **R**:
 - ▶ Така че да съвпада с кортежите **U** и **T** по **A**-тата
 - ▶ С кортежа **T** по **B**-тата
 - ▶ С кортежа **U** по всички останали атрибути, които не са нито **A**-та нито **B**-та

Кортеж	A	B	Други
t	a1	b1	c1
v	a1	b1	c2
u	a1	b2	c2

Многозначни ФЗ

- ▶ Да разгледаме отново релацията `stars_in(name, street, city, title, year)`
- ▶ Така представената релация не нарушава НФБК.
 - ▶ Определяме ключа на релацията
 - ▶ Търсят нетривиални ФЗ: $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ такива че да нарушават НФБК, т.е. лявата част да не съдържа ключа
- ▶ Нито един от атрибутите не може да бъде функционално определен от всички останали атрибути. Например `name, street, city, title` не определя функционално `year`. Аналогично може да се провери и за останалите атрибути на релацията
- ▶ От тук следва, че всичките пет атрибута са ключ на релацията, т.е. няма нетривиална ФЗ, която да нарушава правилото по-горе.
- ▶ От тук следва че релацията е в НФБК
- ▶ До такива ситуации може да се стигна когато обединяваме две или повече много-много връзки в една релация

Многозначни ФЗ

Кортеж	name	street	city	title	year
t	Fisher	s1	c1	Star	1977
u	Fisher	s2	c2	Empire	1980
v	Fisher	s1	c1	Empire	1980

- ▶ Например ако имаме МФЗ $\text{name} \rightarrow \text{street, city}$ за релацията $\text{stars_in}(\text{name, street, city, title, year})$, МФЗ проверява дали има кортеж V с име Fisher и улица и град, които съвпадат по T за street и city и с кортежа U по останалите атрибути

Правила за МФЗ

- ▶ **Тривиални МФЗ.** Ако МФЗ $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ е в сила то $A_1, A_2, \dots, A_n \rightarrow C_1, C_2, \dots, C_k$ също е в сила, където C -тата са B -та и добавени A -та. Аналогично може да премахнем от B -тата A -тата и да получим $A_2, \dots, A_n \rightarrow D_1, D_2, \dots, D_r$, където D -тата са B -та, които не са A -та.
- ▶ **Правило за транзитивност.** Ако МФЗ $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ и $B_1, B_2, \dots, B_m \rightarrow C_1, C_2, \dots, C_k$ то $A_2, \dots, A_n \rightarrow C_1, C_2, \dots, C_k$. Ако има тривиални МФЗ, те могат да бъдат премахнати от правилото.
- ▶ **Нови правила.** Всяка ФЗ е и МФЗ. Ако $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ то $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$
 - ▶ Доказателство: Нека R е релация, за която е в сила ФЗ : $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$. Да предположим че T и U са кортежи на релацията, които съвпадат по A -тата, за да покажем че $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ е в сила трябва да покажем че в релацията съществува кортеж V , които съвпада с T и U по A -тата, с кортежа T по B -тата и с кортежа U по всички останали атрибути. Кортежа U отговаря на тези условия. От ФЗ следва че U съвпада с T по B -тата и със себе си по останалите атрибути.

Правила за МФЗ

- ▶ **Правилото за разделяне и комбиниране не е в сила за МФЗ**
 - ▶ $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ не е равносилно на $A_1, A_2, \dots, A_n \rightarrow B_i$
- ▶ **Правило за допълнение** – ако $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ то $A_1, A_2, \dots, A_n \rightarrow C_1, C_2, \dots, C_m$, където C -тата са останалите атрибути на релацията, различни от A и B
 - ▶ Например: $\text{name} \rightarrow \text{street, city}$ и $\text{name} \rightarrow \text{title, year}$, имаме МФЗ които определят, че всеки актьор има множество от филми, в които участва и те са независими от адреса на актьора
- ▶ Излишествата, които открихме за релацията $\text{stars_in}(\text{name, street, city, title, year})$ могат да бъдат елеминирани с алгоритъм подобен на този за декомпозиране на релацията в НФБК
- ▶ При многозначните ФЗ, този алгоритъм се казва декомпозиране на релацията в четвърта нормална форма.

4НФ

- ▶ В 4НФ, нетривиалните МФЗ които нарушават 4НФ се елиминират, подобно на нетривиалните ФЗ при НФБК. В резултат на това декомпозираната релация няма да има излишества
- ▶ Една МФЗ $A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m$ е нетривиална ако нито едно от B -тата не се среща в A -тата и не всички атрибути на релацията R са сред A -тата и B -тата.
- ▶ Казваме че, релацията R е в 4НФ, ако $A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m$ е нетривиална МФЗ, то $\{A_1, A_2, \dots, A_n\}$ е суперключ за релацията R .
- ▶ Пример: Релацията `stars_in(name, street, city, title, year)` не е в 4 НФ. Например `name`->> `street, city` е нетривиална МФЗ, въпреки това `name` не е суперключ за релацията. Ключът на релацията са всички атрибути взети заедно.
- ▶ Всяка релация, която се намира в 4НФ е и в НФБК.

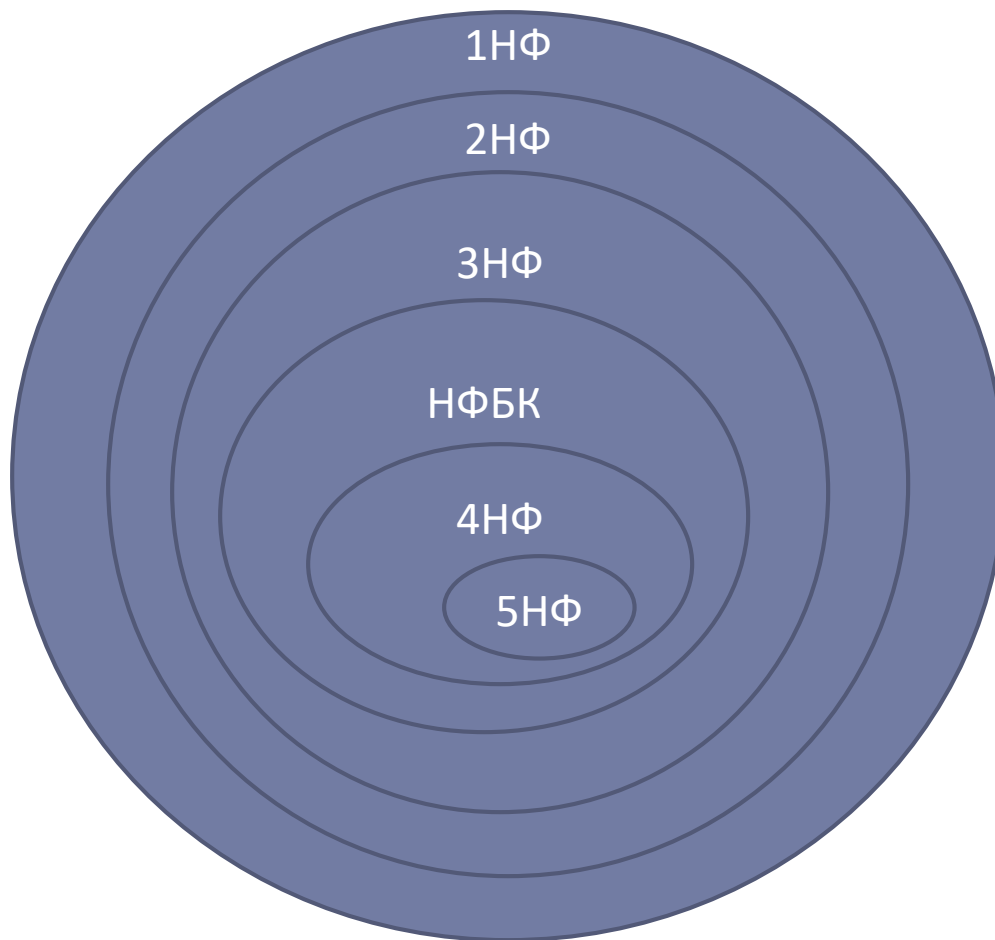
Декомпозиция в 4НФ

- ▶ Алгоритъмът за декомпозиране в 4НФ е почти аналогичен на този за НФБК
 - ▶ Ключът на релацията е определен при нормализиране във 2НФ
 - ▶ Търсим МФЗ, за които е изпълнено, че лявата страна не е суперключ, например $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$
 - ▶ Разделяме схемите на две схеми в първата схема попадат А-тата и В-тата, а във втората схема попадат А-тата и всички останали атрибути, които не са В-та.
- ▶ Например релацията `stars_in(name, street, city, title, year)` не е в 4 НФ. `name → street, city` е нетривиална МФЗ, която нарушава условието в лявата страна да е суперключ.
- ▶ Разделяме схемата на две схеми в едната схема попадат `name, street, city`, а в другата схема `name, title, year`. Получените две схеми са в 4НФ, защото МФЗ: `name → street, city` е тривиална за първата схема, а МФЗ `name → title, year` е тривиална за втората схема. Ако една от релациите не беше в 4НФ, трябваше да приложим алгоритъма още веднъж.

Декомпозиция в 4НФ

- ▶ Както и при НФБК, декомпозицията в 4НФ ни води до получаване на релации с по-малък брой атрибути. Следователно в даден момент със сигурност ще стигнем до схема, която няма да има нужда да се декомпозира повече и ще се намира в 4НФ.
- ▶ Верността на декомпозицията е в сила и при деккомпозиция в 4НФ.
- ▶ Когато декомпозираме според МФЗ, тази зависимост е достатъчна да ни гарантира, че може да възвърнем оригиналната релация от декомпозираните релации без да получаваме лъжливи кортежи.

Връзка между нормалните форми



Връзка между нормалните форми

	3НФ	НФБК	4НФ
Излишества ФЗ	В повечето случаи	Да	Да
Излишества МФЗ	Не	Не	Да
Запазва ФЗ	Да	Може би	Може би
Запазва МФЗ	Може би	Може би	Може би

- НФБК и съответно 4НФ елиминира излишествата при ФЗ
- Само 4НФ елиминира излишествата при МФЗ
- Често 3НФ е достатъчни да елиминира излишествата при ФЗ, но не винаги е така.
- Декомпозицията в 3НФ, винаги може да бъде избрана така, че ФЗ да се запазват.

Изводи за нормализацията

- ▶ Нормализацията е процесът на привеждане на релационната схема до състояние в което нямаме излишества.
- ▶ Има четири нормални форми, които ни гарантират че данните в таблицата са нормализирани
- ▶ Задължително условие е релационната схема да е в поне 3НФ. Това ни гарантира че дизайна на базата от данни е добър.
- ▶ Следването на подхода от E/R модел към релационен е сигурен начин да сме сигурни че релационната схема е в 2НФ, като в повечето случаи е изпълнено че е в 3 НФ.
- ▶ За останалите нормални форми трябва да се направи анализа на ФЗ и МФЗ и да се прибегне към необходимите декомпозиции.