

Бази от данни

Нормализация (първа част)

доц. д-р Димитър Димитров

Въведение

- Една и съща информация може да се представи по различен начин
- Как да проектираме релационни схеми правилно?
 - Какво означава “правилно”?

Дефиниции

- *Нормализация* – процесът по структуриране на релационна БД в съответствие с поредица от т.нар. нормални форми
- *Нормалните форми* са условия върху релациите, при изпълнение на които в тези релации със сигурност няма аномалии от определен вид
- Релациите се разбиват (*декомпозират*) на по-малки релации, за които са в сила условията за дадена нормална форма

Аномалии

- Основни аномалии:
 - Излишества
 - Информация се повтаря, без да е необходимо
 - Аномалии при обновяване
 - Ако има излишества, неуспешно обновяване на всички срещания на дадена информация води до неконсистентност
 - Аномалии при изтриване
 - Ако при изтриване на даден кортеж губим и друга информация
 - И обратното – аномалия при добавяне
- Пример – на следващия слайд

Аномалии – пример

<u><i>title</i></u>	<u><i>year</i></u>	<i>length</i>	<i>filmType</i>	<i>studioName</i>	<u><i>starName</i></u>
Star Wars	1977	124	color	Fox	C. Fisher
Star Wars	1977	124	color	Fox	M. Hamil
Star Wars	1977	124	color	Fox	H. Ford
Mighty Ducks	1991	104	color	Disney	E. Estevez
Wayne's World	1992	95	color	color	M. Meyers

Още една полза от нормализацията

- Софтуерните приложения се променят
 - Дори още по време на разработката им – обикновено не работим по каскадната методология (Waterfall)
 - Пример с Movies – да добавим и възможност за ревюта
- Разширяването на структурата на нормализирана БД не би изисквало големи промени в съществуващите схеми
- В резултат приложенията, които ползват БД, са засегнати в по-малка степен

Преговор

- Релационна алгебра
 - Проекция
 - Нова релация, в която участват само част от атрибутите на дадена релация
 - Релациите са множества => повторенията се отстраняват
 - Естествено съединение (\bowtie)

Декомпозиция на релация

- Дадена е релация $R(A_1, \dots, A_n)$
- Декомпозиция на R – заместването ѝ с релации $R_1(B_1, \dots, B_m)$, $R_2(C_1, \dots, C_k)$, ..., за които:
 - $\{A_1, \dots, A_n\} = \{B_1, \dots, B_m\} \cup \{C_1, \dots, C_k\} \cup \dots$
 - R_i е проекция на R по съответните атрибути

Декомпозиция – пример

- Movies → Movies1 и Movies2

Movies	<i>title</i>	<i>year</i>	<i>length</i>	<i>filmType</i>	<i>studioName</i>	<i>starName</i>
	Star Wars	1977	124	color	Fox	C. Fisher
	Star Wars	1977	124	color	Fox	M. Hamil
	Star Wars	1977	124	color	Fox	H. Ford
	Mighty Ducks	1991	104	color	Disney	E. Estevez
	Wayne's World	1992	95	color	color	M. Meyers

Movies1				
<i>title</i>	<i>year</i>	<i>length</i>	<i>filmType</i>	<i>studioName</i>
Star Wars	1977	124	color	Fox
Mighty Ducks	1991	104	color	Disney
Wayne's World	1992	95	color	color


Movies2		
<i>title</i>	<i>year</i>	<i>starName</i>
Star Wars	1977	C. Fisher
Star Wars	1977	M. Hamil
Star Wars	1977	H. Ford
Mighty Ducks	1991	E. Estevez
Wayne's World	1992	M. Meyers

Пример – особености (1)

- Множествата от атрибутите на новите релации може да се пресичат
 - напр. title, year участват и в двете


Пример – особености (2)

- Има ли аномалии в Movies?
- Има ли аномалии в Movies1 и Movies2?

Movies	<i>title</i>	<i>year</i>	<i>length</i>	<i>filmType</i>	<i>studioName</i>	<i>starName</i>
	Star Wars	1977	124	color	Fox	C. Fisher
	Star Wars	1977	124	color	Fox	M. Hamil
	Star Wars	1977	124	color	Fox	H. Ford
	Mighty Ducks	1991	104	color	Disney	E. Estevez
	Wayne's World	1992	95	color	color	M. Meyers

Movies1				
<i>title</i>	<i>year</i>	<i>length</i>	<i>filmType</i>	<i>studioName</i>
Star Wars	1977	124	color	Fox
Mighty Ducks	1991	104	color	Disney
Wayne's World	1992	95	color	color

Movies2		
<i>title</i>	<i>year</i>	<i>starName</i>
Star Wars	1977	C. Fisher
Star Wars	1977	M. Hamil
Star Wars	1977	H. Ford
Mighty Ducks	1991	E. Estevez
Wayne's World	1992	M. Meyers



Пример – особености (3)

- При конкретната декомпозиция в новите релации:
 - Няма излишества
 - Няма аномалия при обновяване
 - Няма аномалия при изтриване

Пример – особености (4)

- Какво ще се получи, ако приложим естествено съединение върху двете нови релации?
- Или какво ще върне следната заявка:

```
SELECT Movies1.title, Movies1.year,  
length,  
filmType, studioName, starName  
FROM Movies1  
JOIN Movies2  
ON Movies1.title = Movies2.title  
AND Movies1.year =  
Movies2.year;
```

Movies	title	year	length	filmType	studioName	starName
	Star Wars	1977	124	color	Fox	C. Fisher
	Star Wars	1977	124	color	Fox	M. Hamil
	Star Wars	1977	124	color	Fox	H. Ford
	Mighty Ducks	1991	104	color	Disney	E. Estevez
	Wayne's World	1992	95	color	color	M. Meyers

Movies1	title	year	length	filmType	studioName
	Star Wars	1977	124	color	Fox
	Mighty Ducks	1991	104	color	Disney
	Wayne's World	1992	95	color	color

Movies2	title	year	starName
	Star Wars	1977	C. Fisher
	Star Wars	1977	M. Hamil
	Star Wars	1977	H. Ford
	Mighty Ducks	1991	E. Estevez
	Wayne's World	1992	M. Meyers

- Всичко дотук е много хубаво, но как да разберем кои са **правилните** декомпозиции?

Първа нормална форма (1NF)

Първа нормална форма (1NF)

- Една релационна схема R се намира в първа нормална форма (1NF), ако всичките ѝ домейни се състоят от атомарни (неделими) стойности
 - А не множества, списъци, други релации/таблицы и т.н.
- Без нея заявките на SQL биха били по-трудни

1NF – пример (1)

- Employees (emp_id, name, phone, skills)
- Не е в 1NF заради атрибута skills

emp_id	name	phone	skills
1	Елена	+359812345678	Java SQL
2	Иван	+359887654321	Python
3	Георги	+359811223344	Node.js React Python

1NF – пример (2)

- Вариант 1: две отдельные реляции

emp_id	name	phone
1	Елена	+359812345678
2	Иван	+359887654321
3	Георги	+359811223344

emp_id	skill
1	Java
1	SQL
2	Python
3	Node.js
3	React
3	Python

1NF – пример (3)

- Вариант 2: повторение на кортежи
- Въвеждат се аномалии
 - Но могат да бъдат отстранени със следващите нормални форми

emp_id	name	phone	skill
1	Елена	+359812345678	Java
1	Елена	+359812345678	SQL
2	Иван	+359887654321	Python
3	Георги	+359811223344	Node.js
3	Георги	+359811223344	React
3	Георги	+359811223344	Python

1NF – размишления

- Звучи елементарно, но...
- Релация с атрибут от някой от следните домейни удовлетворява ли 1NF?
 - Текст, съдържащ JSON или XML? - И не, и да
 - Пример (JSON): конфигурация на визуален елемент -
“{ "fontSize": 14, "colors": ["blue", "green", "red"] }”
 - Не, ако го използваме вместо по-сложни схеми и правим заявки върху свойствата/таговете (SQL има вградени функции за XML)
 - Да, ако напр. само се праща цялата стойност на клиентското приложение
 - Geography (полигон)? - Да
 - Текст, ако върху него се прилага LIKE (но не се използва като CSV) или Full-text search? - Да
 - BLOB (Binary large object)? - Да
- За изпита по БД не е необходимо да се изучават термините JSON, XML, CSV, BLOB и т.н.

1NF – пример №2

- Movies (title, year, length, rating)
- rating – списък от числови оценки за филма

1NF – пример №3

- Йерархия от служители, всеки служител може да има 0 или повече пряко подчинени други служители

Втора нормална форма (2NF)

Преговор

- Функционална зависимост (ФЗ)
- Намиране на ФЗ в релация
- Тривиална ФЗ
- Нетривиална ФЗ
- Напълно нетривиална ФЗ
- Суперключ
- Намиране на всички ключове по дадени ФЗ

Втора нормална форма (2NF)

- Релация R е във 2NF, ако са изпълнени:
 - R е в 1NF
 - Всеки неключов атрибут е в пълна функционална зависимост от (някой кандидат) ключ на R
 - т.е зависи от целия ключ, а не от някакво (собствено) подмножество на ключа
- Друга дефиниция:
 - R е във 2NF, ако за всяка нетривиална функционална зависимост $A_1A_2...A_n \rightarrow V$, която е изпълнена за R, имаме, че V е елемент на ключ или $\{A_1, A_2, ..., A_n\}$ не е собствено подмножество на някой ключ на R

2NF – пример

- Ключ: {customer_id, store_id}
- Релацията е в 1NF, но нарушава 2NF
- Ф3: store_id → store_location

customer_id	store_id	store_location
1	5	Sofia
2	5	Sofia
1	6	Plovdiv

Алгоритъм за привеждане във 2NF

- Дадена е релация R в 1NF
- Намираме всички (кандидат) ключове на R
- За всяка ФЗ $A \rightarrow B$, където A е собствено подмножество на ключ AC:
 - По правилото за комбиниране добавяме в дясната част B всички атрибути, които са функционално определени от атрибутите в A
 - Премахваме атрибутите B от R
 - Създаваме нова релация с атрибути A U B и с ключ A
- Повтаряме процеса за новополучените релации

2NF – пример

customer_id	store_id	store_location
1	5	Sofia
2	5	Sofia
1	6	Plovdiv

- След прилагане на алгоритъма:

customer_id	store_id
1	5
2	5
1	6

store_id	store_location
5	Sofia
6	Plovdiv

2NF – размишления

- Ако приложим естествено съединение върху новите релации, ще получим ли оригиналната?
 - Съединение без загуба – ще го разгледаме и по-късно
- Ако дадена релация е в 1NF и има само един ключ, който се състои от точно един атрибут, релацията ще бъде ли във 2NF?

2NF – задачи

- Movies (movieTitle, movieYear, starName, starBirthdate, compensation)
- $R(\underline{A}, B, C, D, E, F, G, H)$
 $CD \rightarrow A, D \rightarrow E, F \rightarrow G$
- Още задачи

Пояснения към задачите

- Всички ключове: A, CD

Трета нормална форма (3NF)

Преговор

- Транзитивност (трета аксиома на Армстронг)
 - Ако са дадени $X \rightarrow Y$ и $Y \rightarrow Z$, то $X \rightarrow Z$
 - $X \rightarrow Z$ е транзитивна зависимост

3NF – дефиниции (1)

- Релация R е в 3NF, ако за всяка нетривиална ФЗ $A_1A_2\dots A_n \rightarrow B$, която е изпълнена за R:
 - B е елемент на някой ключ
 - или $\{A_1, A_2, \dots, A_n\}$ е суперключ

3NF – дефиниции (2)

- Релация R е в 3NF, ако всички атрибути са функционално зависими единствено от първичния ключ
- или:
- Релация R е в 3NF, ако са изпълнени следните две условия:
 - R е във 2NF
 - Никой неключов атрибут на R не е транзитивно зависим от първичния ключ

3NF – пример (1)

- MoviesFull (title, year, length, studioName, studioAddress, producer)
- studioName → studioAddress
- и понеже title, year → studioName, имаме транзитивна ФЗ:
title, year → studioAddress
- Излишества, аномалии

3NF – пример (2)

- MoviesFull (title, year, length, studioName, studioAddress, producer)

→

Movies (title, year, length, studioName, producer)

Studios (studioName, studioAddress)

- Аналогичен алгоритъм като за 2NF

3NF – особености

- $\text{MoviesFull} = \text{Movies} \bowtie \text{Studios}$
- Релация винаги може да се приведе в 3NF
- $3NF \Rightarrow 2NF \Rightarrow 1NF$
- Декомпозицията в 3NF винаги може да бъде избрана така, че ФЗ да се запазват

3NF – примери

- $R(\underline{A}, B, C, D, E, F, G, H)$
 $CD \rightarrow A, D \rightarrow E, F \rightarrow G$

- В повечето релации в 3NF няма аномалии по обновяване, изтриване и добавяне
- Тези, които имат, се срещат рядко в практиката

Нормална форма на Бойс-Код (BCNF)

BCNF – въведение

- По-късно след създаването на 3NF е осъзнато, че тя не елиминира всички аномалии
- Разработва се по-стриктна версия – нормална форма на Бойс-Код (BCNF)
- Може да се нарече и 3.5NF

BCNF – дефиниция

- Прямо дефиницията на 3NF:
- Релация R е в BCNF, ако за всяка нетривиална ФЗ $A_1A_2\dots A_n \rightarrow B$, която е изпълнена за R:
 - ~~B е елемент на някой ключ~~
 - ~~или~~ $\{A_1, A_2, \dots, A_n\}$ е суперключ

BCNF – аналогичен алгоритъм

- Релация $R (A_1, \dots, A_n, B_1, \dots, B_m, C_1, \dots, C_k)$
- Нека $A_1 \dots A_n \rightarrow B_1 \dots B_m$ е нетривиална ФЗ и $\{A_1, \dots, A_n\}$ не е суперключ (и по правилото за комбиниране в дясната част са добавени всички атрибути, които са функционално определени от $A_1 \dots A_n$)
- Декомпозираме релацията R
 - $S (A_1, \dots, A_n, B_1, \dots, B_m)$
 - $T (A_1, \dots, A_n, C_1, \dots, C_k)$
- Ако някоя от новите релации не е в BCNF, прилагаме същия алгоритъм и за нея
- ФЗ в новите релации се изчисляват чрез проектиране на ФЗ от изходната релация

BCNF – пример (1)

- FullMovies (title, year, length, filmType, studioName, starName)
- ФЗ: title year → length filmType studioName
- FullMovies не е в BCNF (а в коя НФ е?)
- Ключ: {title, year, starname}, няма други
- Лявата част на ФЗ не е суперключ
- BCNF е нарушена в случаите:
 - title year → studioName
 - title year → length
 - title year → filmType

BCNF – пример (2)

- Декомпозираме FullMovies в:
 - Movies (title, year, length, filmType, studioName)
 - StarsIn (title, year, starName)
- Новите релации удовлетворяват ли BCNF?
 - Ф3 title, year \rightarrow length, filmType, studioName – лявата част е ключ
- Имат ли излишества?
- Movies \bowtie StarsIn равно ли е на FullMovies?

BCNF – свойство

- Всяка релация с два атрибута е в BCNF
- Доказателство:
 - Нека $R(A,B)$
 - Ако ключът е $\{A, B\}$, то няма нетривиални зависимости \Rightarrow няма нарушаване на BCNF
 - Ако $A \rightarrow B$, но не и $B \rightarrow A$, то само A е ключ \Rightarrow всички нетривиални ФЗ съдържат A в лявата си част
 - Аналогично за $B \rightarrow A$
 - Ако $A \rightarrow B$ и $B \rightarrow A$, то A и B са ключове и всички нетривиални ФЗ съдържат A или B в лявата си част
- Ще се запазят ли обаче ФЗ на първоначалната релация?
- Процесът по прилагане на алгоритъма за декомпозиция в BCNF е краен, тъй като винаги получаваме релации с по-малко атрибути, а всяка релация с два атрибута е в BCNF

Съединение без загуба на ФЗ

- Релация R е декомпозирана на S и T
- F – ФЗ на R
- F_1 – проекция на F върху S
- F_2 – проекция на F върху T
- Съединението на S и T е без загуба на функционалните зависимости, ако:
- $F_1 \cup F_2 = F$

3NF vs BCNF

- Пример за релация в 3NF, която не удовлетворява BCNF:
 - $R(\underline{A}, \underline{B}, C)$ ($\{A, B\}$ е ключ)
 - фз $C \rightarrow B$
 - C не е суперключ

Загуба на ФЗ (1)

- БД за резервиране на билети за филми, прожектирани в различни киносалони
 - Всеки киносалон се намира в строго определен град
 - Един филм може да се прожектира в няколко киносалона едновременно и в няколко града едновременно, но не може в няколко салона в един и същ град
 - Градовете обикновено имат повече от един киносалон
 - Има киносалони с няколко зали, в които могат да се прожектират няколко филма едновременно

Загуба на ФЗ (2)

- Релация Booking (заявки за билети) със следните атрибути:
 - title – име на филм
 - theater – име на киносалон
 - city – град, в който се намира киносалонът
- Функционални зависимости:
 - theater \rightarrow city
 - title, city \rightarrow theater
- Отговаря ли конкретният пример на абстрактния пример – $R(\underline{A}, \underline{B}, C)$ с ФЗ $C \rightarrow B$?

Загуба на ФЗ (3)

- (Кандидат) ключове в Booking:
- Нито един единичен атрибут не е ключ
- {title, city} е ключ заради дадената ФЗ
- {theater, title} е ключ, защото theater функционално определя city
- {city, theater} функционално не определят title и затова не формират ключ
- Ключове: {title, city} и {theater, title}

Загуба на ФЗ (4)

- Декомпозиция в BCNF
- theater \rightarrow city
- Декомпозираме в следните две релационни схеми:
 - R (theater, city)
 - S (theater, title)

Загуба на ФЗ (5)

- ФЗ title, city → theater не е валидна за нито една от новите релации
- Никой не ни спира да добавим следните кортежи:

theater	city
El Capitan	LA
New Beverly Cinema	LA

theater	title
El Capitan	The Net
New Beverly Cinema	The Net

- Съединение:

title	city	theater
The Net	LA	El Capitan
The Net	LA	New Beverly Cinema

- ФЗ title, city → theater е нарушена

BCNF – заключение

- Може да приемем, че 3NF е смекчена форма на BCNF
- BCNF не гарантира запазването на ф3
- Съществуват релации, които не са в BCNF, за които по-нататъшна декомпозиция е нецелесъобразна или невъзможна

Упражнения (1)

- Дадени са релации, коя НФ удовлетворяват?
 - Да разгледаме и по-голям пример – напр. Northwind

Упражнения (2)

- Привеждане на релации в определена НФ
 - От Wikipedia: Books (Title, Author, Author Nationality, Format, Price, Subject, Pages, Thickness, Publisher, Publisher Country, Genre ID, Genre Name)

Въпроси?