

Нормализация на реляционните бази

Богдан Шишеджиев -
Нормализация

1

Основни понятия

- Декомпозиция

- Дефиниция

Една декомпозиция на реляцията $R(A_1, \dots, A_n)$ е заместването ѝ с множеството реляции R_1, \dots, R_n , получени чрез проекции така, че R и $R_1 * R_2 * \dots * R_n$ имат една и съща схема

- без загуба на информация

$$R = R_1 * R_2 * \dots * R_n$$

- Със запазване на функционалните зависимости.

Богдан Шишеджиев -
Нормализация

3

Основни понятия

- Създаване на добра схема

- Представяне.
- Без излишък
- Разделяне

- Универсална реляция

FIRM (N_employee, N_department, chef, contract_type)
HOSPITAL(N_patient, Name, Ward, Doctor)

- аномалия при обновяване
- аномалия при изтриване
- аномалия при вмъкване
- излишък

Богдан Шишеджиев -
Нормализация

2

Декомпозиция

R	N°	name	age	address	speciality
	1	A	30	x1	S1
	2	B	30	x2	S2
	3	C	30	x3	S2
	4	D	40	x4	S1

R1	age	speciality	R2	N	name	addr	R3	N	age
	30	S1		1	A	x1		1	30
	30	S2		2	B	x2		2	30
	40	S1		3	C	x3		3	30
				4	D	x4		4	40

R=R1*R2*R3	N°	name	address	age	speciality
	1	A	x1	30	S1
	1	A	x1	30	S2
	2	B	x2	30	S1
	2	B	x2	30	S2
	3	C	x3	30	S1
	3	C	x3	30	S2
	4	D	x4	40	S1

Декомпозиция

R	N°	name	age	address	speciality
	1	A	30	x1	S1
	2	B	30	x2	S2
	3	C	30	x3	S2
	4	D	40	x4	S1

R1	N°	speciality	R2	N	name	addr	R3	N	age
	1	S1		1	A	x1		1	30
	2	S2		2	B	x2		2	30
	3	S2		3	C	x3		3	30
	4	S1		4	D	x4		4	40

R=R1*R2*R3	N°	name	address	age	speciality
	1	A	x1	30	S1
	2	B	x2	30	S2
	3	C	x3	30	S2
	4	D	x4	40	S1

Богдан Шишеджиев -
Нормализация

5

Първа нормална форма

- Една релация е в първа нормална форма, ако всичките атрибути съдържат атомарна стойност (базирана върху прост домен) и няма повтарящи се атрибути

EMPLOYEE (No, name, Child (firstname, age))

EMPLOYEE (No, name); CHILDREN (No, firstname, age)

No	NAME	CHILD	
		FIRSTNAME	AGE
500	DUPONT	ANDRE	10
501	DURAND	JEAN	11
501	DURAND	PIERRE	12
510	LEFEBVRE	PAUL	13
510	LEFEBVRE	JACQUES	14

EMPLOYEE	No	NAME	CHILDREN	No	FirstName	age
	500	DUPONT		500	André	10
	501	DURAND		501	Jean	11
	510	LEFEBVRE		501	Pierre	12
				510	Paul	13
				510	Jacques	14

Богдан Шишеджиев -
Нормализация

7

Нормални форми

- Правила за декомпозиция на релациите, така че схемата да се подобри, т.е. да се отстранят дефектите и да се намалят рисковете от аномалии
- Недостатъци
 - Усложняване
 - Загуба на яснота и отдалечаване от реалния обект
 - намаляване на ефективността (трябва да се използват много съединения)

Богдан Шишеджиев -
Нормализация

6

Първа нормална форма

MOVIE (No, Name, Director, Actor1, Actor2, Actor3)

MOVIE (No, Name, Director)

CAST (NoMovie, Actor)

MOVIE	No	Name	Director	Actor1	Actor2	Actor3
	1	The Silence of the Lambs	Jonathan Demme	Jodie Foster	Anthony Hopkins	
	2	The Sixth Sense	M. Night Shyamalan	Bruce Willis		

MOVIE	No	Name	Director	CAST	NoMovie	Actor
	1	The Silence of the Lambs	Jonathan Demme		1	Jodie Foster
					1	Anthony Hopkins
	2	The Sixth Sense	M. Night Shyamalan		2	Bruce Willis

Богдан Шишеджиев -
Нормализация

8

Втора нормална форма

- Определение:**

Ще казваме, че Y напълно зависи от X , ако $X \rightarrow Y$ и няма подмножество $Z \subset X$, за което $Z \rightarrow Y$, т.е. Y зависи от целия X , но не зависи от никоя негова част

- Една релация R е във 2NF ако и само ако:

- Тя е в 1NF
- Всички неключови атрибути са напълно зависими от ключа на R .

- Пример**

STOCK (part, warehouse, quant, address) не е във 2NF

ФЗ : part, warehouse \rightarrow quant ; warehouse \rightarrow address

Декомпозиция

STOCK (part, warehouse, quant)

LOCAL(warehouse, address)

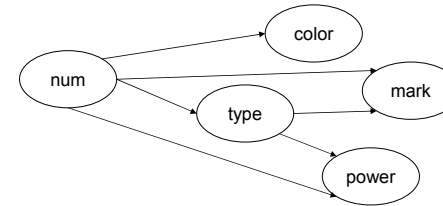
Богдан Шишеджиев -
Нормализация

9

Трета нормална форма

- Пример**

CAR (num, mark, type, power, color)



Решение 1

R1(num, type, color)

R2 (type, mark, power)

Решение 2

R'1(num, type)

R'2(num, power, color)

R'3 (type, mark)

Богдан Шишеджиев -
Нормализация

11

Трета нормална форма

- Една релация е в 3NF, ако и само ако:

- Тя е във 2NF
- Всички атрибути не принадлежащи на един ключ не зависят от неключов атрибут (няма транзитивна ФЗ)

- Пример**

PERSONAL(employee, name, FirstName, Service, Address)

ФЗ : employee \rightarrow name, FirstName, Service, Address;

Service \rightarrow Address

Декомпозиция

PERSONAL(employee, name, FirstName, Service)

LOCAL(Service, Address)

Богдан Шишеджиев -
Нормализация

10

Нормална форма на Бойс-Код (Boyce-Codd)

- Една релация е в BCNF, ако и само ако:

- Тя е във 3NF
- Елементарните ФЗ са само такива, при които ключ определя атрибут (премахват се частичните и транзитивните ФЗ).

- Пример**

POSTCODE	code	town	street
	59650	LA	Sunset
	59650	LA	Hollywood

Декомпозиция

CODE_TOWN(code, town)

CODE_STREET(code, street)



Богдан Шишеджиев -
Нормализация

12

Нормална форма на Бойс-Код

- Пример

Информация за записване:

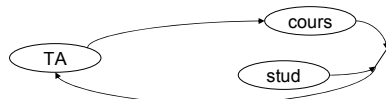
- Във всеки курс могат да се запишат много студенти
- Всеки студент може да се запише в повече курсове
- Един тютор е утвърден само за един курс
- Всеки студент за всеки курс има по един тютор

Релационна схема

Inscription(Cours, Stud, TA)

Кандидат ключове

(Stud, Cours) (Stud, TA)



Богдан Шишеджиев -
Нормализация

13

Четвърта нормална форма

- Многозначни зависимости

- Дефиниция

Нека имаме схема R и $X \subseteq R$ и $Y \subseteq R$, $X \cap Y = \emptyset$, $Z = R - (X \cup Y)$. Казваме, че $X \twoheadrightarrow Y$, ако за всеки 2 кортежа t_1 и t_2 от R , за които $t_1(X) = t_2(X)$, в R съществува кортеж t_3 , за който $t_3(X) = t_1(X)$, $t_3(Y) = t_1(Y)$, $t_3(Z) = t_2(Z)$. От симетрията се вижда, че също има t_4 , $t_4(X) = t_1(X)$, $t_4(Y) = t_2(Y)$, $t_4(Z) = t_1(Z)$.

- Следствие

Ако $X \twoheadrightarrow Y$, то и $X \twoheadrightarrow R - (X \cup Y)$

Богдан Шишеджиев -
Нормализация

15

Нормална форма на Бойс-Код

INSCRIPTION	Course	Stud	TA	TUTEURS	Course	TA
	ENG101	Jones	Clark		ENG101	Clark
	ENG101	Grayson	Chen		ENG101	Chen
	ENG101	Samara	Chen		ENG101	Chen
	MAT350	Grayson	Powers		MAT350	Powers
	MAT350	Jones	O'Shea		MAT350	O'Shea
	MAT350	Berg	Powers		MAT350	O'Shea

INSCRIPTION	Course	Stud
	ENG101	Jones
	ENG101	Grayson
	ENG101	Samara
	MAT350	Grayson
	MAT350	Jones
	MAT350	Berg

TUTORIAL	Stud	TA
	Jones	Clark
	Grayson	Chen
	Samara	Chen
	Grayson	Powers
	Jones	O'Shea
	Berg	Powers

Декомпозиция

Inscription(Cours, Stud)

Tutors(Course, TA)

Tutorial(Stud, TA)

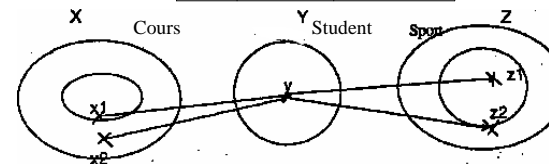
Богдан Шишеджиев -
Нормализация

14

Четвърта нормална форма

- Многозначни зависимости

STUDENTS	N	COURS	SPORT
	100	BD	Tennis
	100	BD	Football
	200	BD	Swimming
	200	AN	Swimming



$Y \twoheadrightarrow Z : x_1yz_1 \text{ et } x_2yz_2 \in R \Rightarrow x_1yz_2 \text{ et } x_2yz_1 \in R$

Богдан Шишеджиев -
Нормализация

16

Четвърта нормална форма

- Аксиоми

- Допълнение : $X \twoheadrightarrow Y \implies X \twoheadrightarrow R \cdot X \cdot Y$
- Мултидобавяне : $(X \twoheadrightarrow Y) \text{ и } W \subseteq R \implies \mathbf{XW} \twoheadrightarrow \mathbf{YW}$
- псевдотранзитивност : $(X \twoheadrightarrow Y) \text{ и } (Y \twoheadrightarrow Z) \implies X \twoheadrightarrow Z \cdot Y$
- репликация : $X \rightarrow Y \implies X \twoheadrightarrow Y$
- сливане : $X \twoheadrightarrow Y \text{ et } Z \subseteq Y \text{ и } \exists W \subseteq R \text{ с } W \cap Y = \emptyset \text{ и } W \rightarrow Z \implies X \twoheadrightarrow Z$
- обединение : $(X \twoheadrightarrow Y) \text{ и } (Y \twoheadrightarrow Z) \implies X \twoheadrightarrow YZ$
- сечение : $X \twoheadrightarrow Y \text{ и } X \twoheadrightarrow Z \implies X \twoheadrightarrow Y \cap Z$
- разлика : $X \twoheadrightarrow Y \text{ и } X \twoheadrightarrow Z \implies X \twoheadrightarrow Y \cdot Z \text{ и } X \twoheadrightarrow Z \cdot Y$

Богдан Шишеджиев -
Нормализация

17

Пета нормална форма

- Зависимост от съединението

R1	STUDENTS	COURS	PROF
	X	CL	Z
	X	CL	T
	X	SIO	T
	Y	SIO	T

- Тази релация не е в 4NF, но няма M3.
- $(X \text{ CL } Z) \text{ и } (X \text{ SIO } T) \neq \implies (X \text{ CL } T) \in R \text{ и } (X \text{ SIO } Z) \in R$
- $\text{STUDENT} \twoheadrightarrow \text{COURS}$ е невярно, защото $(X \text{ SIO } Z)$ не съществува и не може да съществува
- $\text{COURS} \twoheadrightarrow \text{PROF}$ също, защото $(Y \text{ CL } Z)$ не съществува
- $\text{PROF} \twoheadrightarrow \text{STUDENT}$ също, защото $(Y \text{ SIO } T)$ не съществува

Богдан Шишеджиев -
Нормализация

19

Четвърта нормална форма

- Дефиниции

- Една релация е в 4NF ако и само ако е в BCNF единствените елементарни ФЗ са тези, при които един ключ определя един атрибут.
- Една релация е в 4NF ако и само ако е в BCNF и не съдържа повече от една многозначна зависимост

- Пример

STUDENT (num, cours, sport) не е в 4NF

Декомпозиция

num \twoheadrightarrow cours – R1(num, cours)

num \twoheadrightarrow sport – R2(num, sport)

Богдан Шишеджиев -
Нормализация

18

Пета нормална форма

R1	Student	cours	R2	Student	Prof	R3	cours	Prof
	X	CL		X	Z		CL	Z
	X	SIO		X	T		CL	T
	Y	CL		Y	T		SIO	T

В този случай декомпозицията е на 3, а не на 2 релации, защото никоя двойка $R1 \cdot R2$, $R2 \cdot R3$ или $R1 \cdot R3$ не възстановяват R.
 $R = R1 \cdot R2 \cdot R3$ – това се нарича зависимост от съединението (има зависимост между повече от 2 атрибута).

- Дефиниция**

Релацията R удовлетворява зависимостта от съединение, ако и само ако е съединение на (R_1, R_2, \dots, R_n) , където $R_i \subseteq R$

Богдан Шишеджиев -
Нормализация

20

Пета нормална форма

- Релацията R е в 5NF, ако за всяка 3C едно от следните е в сила:
 - (R_1, R_2, \dots, R_n) е тривиална 3C, т.е. едно R_i е R
 - Всяко R_i е кандидат ключ на R
- Тази форма се нарича още нормална форма на проекция-съединение (*PJNF*). Ако схемата е в тази форма, повече не може да се декомпозира без загуба на информация

Богдан Шишеджиев -
Нормализация

21

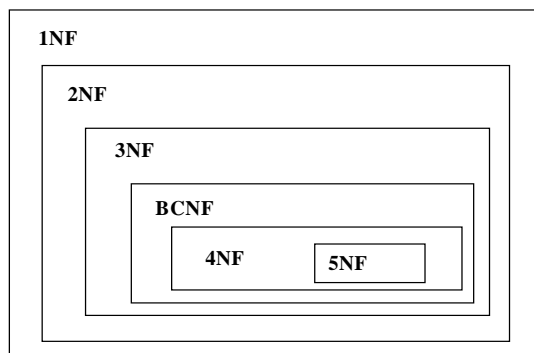
Заклучение

- **Предимства и недостатъци на реляционния модел**
 - **Предимства**
 - Простота за потребителя
 - Независимост на потребителя от логическата и физическата структура, както и от методите за достъп до данните. Понятието файл съществува само за администратора и е грижа на СУБД.
 - Мощност и еднотипност на представянето : математическата теория позволява точно и алгоритмично проектиране на схемата.
 - Мощност на осигуряване на сигурност на данните: контрол според съдържанието, структурата и контекста.
 - Съществуване на непроцедурен интерфейс за неинформатици.
 - Бурно развитие на комерсиални СУБД и създаването на 4-то поколение SQL и QBE.
 - **Недостатъци**
 - Необходимост от мощна СУБД
 - Известна загуба на логическа независимост при нормализацията.

Богдан Шишеджиев -
Нормализация

23

Заклучение



Богдан Шишеджиев -
Нормализация

22

Пример за нормализация

Реалната област е авиокомпания.

Имена на атрибутите:
Имена на компаниите – C
номера на полети – V
Номера на самолети – A
модела на самолети – M
капацитет на самолетите – K
имена на пилотите – P
час на излитане – H
ден на излитане – J
начален град – D
град на пристигане – E

ФЗ :

- (1) $M \rightarrow K$
- (2) $A \rightarrow M, C$
- (3) $V \leftrightarrow H, D, E$
- (4) $J, V \rightarrow P, A$
- (5) $J, H, P \rightarrow D, E, A$
- (6) $J, H, A \rightarrow D, E, P$

R (C,A,V,M,K,P,H,J,D,E) с канд. ключове [VJ], [HDEJ], [PJH],[AJH]

От DF (3) R не е в 2NF:

$R1(V,H,D,E)$ с канд. кл. [V] et [HDE]

$R2(C,A,V,M,K,P,J)$ с канд. кл. [VJ]

От DF (2) R2 не е в 3NF

$R21(V,J,A,P)$ [VJ]

$R22(A,C,M,K)$ [A]

От DF (1) R22 не е в 3NF

$R221(A,C,M)$ [A]

$R222(M,K)$ [M]

Нормализираната схема е :

$R1(\underline{V},H,D,E)$

$R21(\underline{V},J,A,P)$

$R221(\underline{A},C,M)$

$R222(\underline{M},K)$

Богдан Шишеджиев -
Нормализация

24

Пример за нормализация

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
a1	b1	c1	d1	e1
a2	b1	c1	d1	e4
a1	b1	c1	d2	e5
a1	b3	c3	d1	e2
a1	b1	c1	d1	e7
a1	b2	c1	d1	e3
a3	b1	c3	d1	e6
a4	b1	c1	d2	e7

$DE \rightarrow A$; $EA \rightarrow D$; $E \rightarrow B$; $E \rightarrow C$; $AB \rightarrow C$

Отговор : $T1(\underline{A}, \underline{D}, \underline{E})$, $T2(\underline{A}, \underline{B}, \underline{C})$, $T3(\underline{B}, \underline{E})$

или $T1(\underline{D}, \underline{E}, \underline{A})$, $T2(\underline{E}, \underline{B}, \underline{C})$