



Базы от данни

Релационен модел на данни. Нормализация.

Многозначни зависимости (MVD)

- ▶ 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 името на актьора и филма, в който участва ще се повтарят за всеки различен адрес, на който живее актьора.
- ▶ Поради ограниченията на 1НФ, трябва да имаме по един ред с име на актьор и филм, в който участва за всеки адрес на актьора
- ▶ Това води до повтаряне на информация (излишество) в релацията `stars_in`
- ▶ Ключът за релацията `stars_in` се състои от всичките 5 атрибута на релацията. Така, че релацията се намира в НФБК, защото няма нетриална ФЗ, която да нарушава правилото на Бойс-Код.
- ▶ Как да избегнем тези излишества?

Многозначни зависимости (МЗ)

- ▶ Такива излишества за релации, които са в НФБК и за които има МЗ се избягват с декомпозиция в 4НФ
- ▶ Причината да се появяват МЗ в релациите се дължи на невъзможността атрибутите да съдържат структура (масив, множество и т.н.) като стойност.
- ▶ **МЗ дефиниция:** Нека е дадена релацията $R(A, B, C)$. Казваме, че МЗ $A \twoheadrightarrow B$ (чете се многозначно определя) е в сила за релацията R тогава и само тогава когато, за всеки два кортежа, които съвпадат по атрибута A стойности им в атрибута B и стойностите им в атрибута C се различават (са независими).

МЗ

- ▶ Нека е дадена релацията $R(A, B, C)$. 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 C -тата

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

МЗ - пример

- ▶ Нека разгледаме релацията `stars_in(name, street, city, title, year)`

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

- ▶ МЗ, която е в сила за релацията `stars_in` е:
 - ▶ `name->>street, city`
- ▶ ФЗ която НЕ е в сила:
 - ▶ `name, title, year -> street, city (X)`
- ▶ Ключът за релацията `stars_in` се състои от всичките 5 атрибута, следователно релацията е в НФБК

M3

- ▶ `stars_in(name, street, city, title, year)`
- ▶ Така представената релация не нарушава НФБК.
 - ▶ Определяме ключа на релацията
 - ▶ Търсят нетривиални ФЗ, 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 или при преобразуване на тринарни връзки (n-арни) връзки.

4НФ

- ▶ Такива проблеми се разрешават, като се декомпозира до 4НФ
- ▶ В 4НФ нетривиалните МЗ, които нарушават 4НФ се елиминират, подобно на нетривиалните ФЗ при НФБК. В резултат на това декомпозираната релация няма да има излишества
- ▶ Една МЗ $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ 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 за всяка нетривиална МЗ $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$, която е в сила за релацията R е изпълнено че $\{A_1, A_2, \dots, A_n\}$ е суперключ за релацията R .
- ▶ Пример: Релацията `stars_in(name, street, city, title, year)` не е в 4 НФ.
 - ▶ `name → street, city` е нетривиална МЗ, въпреки това `name` не е суперключ за релацията. Ключът на релацията са всички атрибути взети заедно.
 - ▶ Релацията НЕ е в 4НФ, т.е. Трябва да се декомпозира.

Правила за МЗ

- ▶ **Тривиални МЗ.** Ако $MЗ A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ 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 $MЗ A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ 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 МЗ, те могат да бъдат премахнати от правилото.
- ▶ **Нови правила.** Всяка ФЗ е и МЗ. 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 МЗ

- ▶ **Правилото за разделяне и комбиниране не е в сила за МЗ**
 - ▶ $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ 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 МЗ 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 МЗ, този алгоритъм се казва декомпозиране на релацията в четвърта нормална форма.

Декомпозиция в 4НФ

- ▶ Алгоритъмът за декомпозиране в 4НФ е 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. $\text{name} \rightarrow \text{street, city}$ е нетривиална МЗ, 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 МЗ: $\text{name} \rightarrow \text{street, city}$ е тривиална за първата схема, а МЗ $\text{name} \rightarrow \text{title, year}$ е тривиална за втората схема. Ако една от релациите не беше в 4НФ, трябваше да приложим алгоритъма още веднъж.
 - ▶ `stars(name, street, city)`
 - ▶ `stars_in(name, title, year)`

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 декомпозиция в 4НФ.
- ▶ Когато декомпозираме според МЗ, тази зависимост е достатъчна да ни гарантира, че може да възвърнем оригиналната релация от декомпозираните релации без да получаваме лъжливи кортежи.

МЗ и 4НФ - Пример

branchNo	sName	oName
B003	Ann Beech	Carol Farrel
B003	David Ford	Carol Farrel
B003	Ann Beech	Tina Murphy
B003	David Ford	Tina Murphy

- ▶ Дадена е релацията BranchStaffOwner(branchNo, sName, oName)
 - ▶ sName еднозначно определя всеки Staff
 - ▶ oName еднозначно определя всеки Owner
- ▶ Ключът за релацията е {branchNo, sName, oName}
- ▶ Релацията е в НФБК, но не е в 4НФ
- ▶ В сила са следните МЗ:
 - ▶ branchNo ->> sName
 - ▶ branchNo ->> oName

МЗ и 4НФ - Пример

branchNo	sName	oName
B003	Ann Beech	Carol Farrel
B003	David Ford	Carol Farrel
B003	Ann Beech	Tina Murphy
B003	David Ford	Tina Murphy

- ▶ В сила са следните МЗ:
 - ▶ branchNo ->> sName
 - ▶ branchNo ->> oName
- ▶ Декомпозираме до 4НФ:
 - ▶ BranchStaff(branchNo, sName) – в 4НФ
 - ▶ BranchOwners(branchNo, oName) – в 4НФ

Нормални форми

- ▶ **Първа нормална форма (1НФ)** – изисква всеки компонент в кортежите на релацията да има атомарна стройност.
- ▶ **Втора нормална форма (2НФ)** – изисква релацията да е в 1НФ, да няма атрибути, в които да се пазят изчислими стойности, както и всеки атрибут в релацията да е функционално зависим от атрибутите, съставляващи първичния ключ, но да не е и от негово подмножество.
- ▶ **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 част от ключ.
- ▶ **Нормална форма на Бойс-Код (НФБК)** – изисква релацията да е в 3НФ и за всяка нетривиална ФЗ $A_1, A_2, \dots, A_n \rightarrow V$, която е в сила за R , е изпълнено че $\{A_1, A_2, \dots, A_n\}$ е супер-ключ за R .
- ▶ **Четвърта нормална форма (4НФ)** изисква релацията да е в НФБК и за всяка нетривиална МФЗ $A_1, A_2, \dots, A_n \twoheadrightarrow V$, която е в сила за R , е изпълнено че $\{A_1, A_2, \dots, A_n\}$ е супер-ключ за R .

Връзка между нормалните форми

