

# Бази от данни

## Нормализация (втора част)

доц. д-р Димитър Димитров

# Въведение (1)

- Всеки служител може да има един или повече телефонни номера
- Как да представим това в релация?

# Въведение (2)

name	phone
Елена	123
Елена	4567
Иван	23456
Георги	7890
Георги	567

# Въведение (3)

- Всеки служител освен телефони може да има и едно или повече умения

# Въведение (4)

name	phone	skill
Елена	123	Java
Елена	123	SQL
Елена	4567	Java
Елена	4567	SQL
Иван	23456	Python
Иван	23456	R
Георги	7890	Node.js
Георги	7890	React
Георги	7890	Python
Георги	567	Node.js
Георги	567	React
Георги	567	Python

# Въведение (5)

- Очевидни проблеми, например
  - Как ще добавим нов телефон или умение?
  - Как ще изтрием телефон или умение?
- Но: няма основание да свържем телефон с едно умение и да не го свържем с друго
  - Как бихме намерили по даден списък с телефони кой от служителите знае SQL?

# Въведение (6)

- В коя нормална форма е тази релация?
- В BCNF!
  - Следователно и в 3NF, 2NF и 1NF
  - Защо е в BCNF?
    - От кои атрибути се състои ключът?
      - От всички
    - Имаме ли нетривиални функционални зависимости?

# Въведение (7)

- Решение: четвърта нормална форма (4NF)
- Как предполагахме, че трябва да се представи примерната информация?



# Многозначни зависимости

# Многозначни зависимости

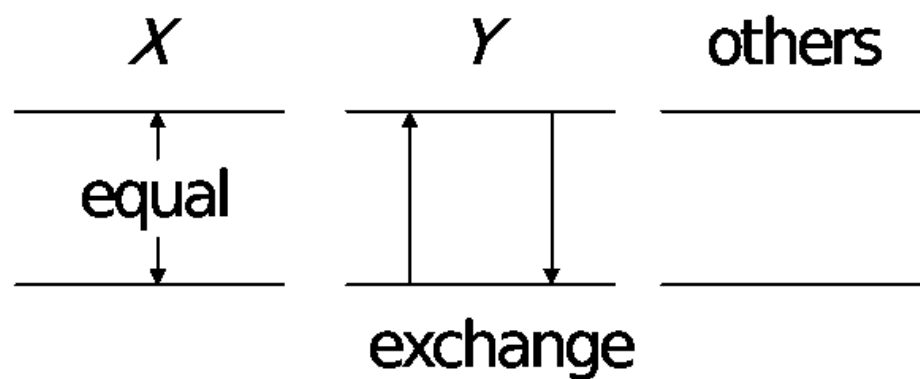
- Два атрибута (или множества от атрибути) са независими помежду си
- Обобщение на функционалните зависимости
- Наличие на ситуации, при които независимостта на атрибутите не се изразява чрез ФЗ

# Дефиниция

- Multivalued dependency (MVD)
- $X \twoheadrightarrow Y$  утвърждава, че ако два кортежа в една релация съвпадат по всички атрибути на  $X$ , то техните компоненти от множеството атрибути  $Y$  могат да бъдат разменени и резултатът ще даде два нови кортежа, които също принадлежат на релацията

# Наглядно представляне

- $X \Rightarrow Y$



# Формална дефиниция

- $A_1A_2...A_n \twoheadrightarrow B_1B_2...B_m$  е многозначна зависимост в  $R$ , ако:

за всяка двойка кортежи  $t, u$  от  $R$ , за които

$$t[A_1A_2...A_n] = u[A_1A_2...A_n],$$

съществува кортеж  $v$  от  $R$ , за който:

$$(1) v[A_1A_2...A_n] = t[A_1A_2...A_n] = u[A_1A_2...A_n]$$

$$(2) v[B_1B_2...B_m] = t[B_1B_2...B_m]$$

$$(3) v[C_1C_2...C_k] = u[C_1C_2...C_k],$$

където  $C_1C_2...C_k$  са всички атрибути от  $R$  с изключение на

$$A_1A_2...A_nB_1B_2...B_m$$

# Пример

- name → street, city
- Ако съществуват следните кортежи, съвпадащи по name:

name	street	city	title	year
C. Fisher	5 Locust Ln.	Hollywood	Star Wars	1977
C. Fisher	123 Maple Str.	Malibu	Empire Strikes Back	1980

- То със сигурност съществуват и:

name	street	city	title	year
C. Fisher	123 Maple Str.	Malibu	Star Wars	1977
C. Fisher	5 Locust Ln.	Hollywood	Empire Strikes Back	1980

# Пояснение към примера

- Всяка филмова звезда може да има няколко адреса и няколко филма
- Множеството от адресите на дадена звезда е напълно независимо от множеството от филмите, в които е играла тя

# Свойства

- Тривиални зависимости
- Правило за транзитивност
- Правило за попълнение
- Правило за обединение



# Тривиални многозначни зависимости

- Нека  $A$  и  $B$  са множества от атрибути
- Дадена е МЗ  $A \twoheadrightarrow B$  в релация  $R$
- Тривиална МЗ, ако:
  - $B$  е подмножество на  $A$  или
  - $A \cup B$  съдържа всички атрибути на  $R$
- Нетривиална МЗ, ако:
  - Нито един от атрибутите  $B$  не принадлежи на  $A$
  - Не всички атрибути на  $R$  принадлежат на  $A \cup B$

# Правило за транзитивност

- Ако  $A \Rightarrow B$  и  $B \Rightarrow C$ , то  $A \Rightarrow C$

# Правило за попълнение

- Ако  $A \Rightarrow B$ , то  $A \Rightarrow C$ , където  $C$  е множеството от всички атрибути на  $R$  с изключение на  $A \cup B$
- В сила ли е аналогично правило за  $\Phi_3$ ?

# Правило на попълнението – пример

- Многозначна зависимост в Stars:
  - name → street city
- Според правилото на попълнението имаме и:
  - name → title year

# Правило за обединение

- Ако  $A \Rightarrow B$  и  $A \Rightarrow C$ , то  $A \Rightarrow B \cup C$

# Още свойства

- Както и при ФЗ, не можем да разделяме лявата част на МЗ
- За разлика от ФЗ, не можем да разделяме и дясната част – понякога се налага да оставяме няколко атрибута в дясната част

# Пример

- name ➔ street, city
- name ➔ street ?

name	street	city	title	year
C. Fisher	5 Locust Ln.	Hollywood	Star Wars	1977
C. Fisher	123 Maple Str.	Malibu	Star Wars	1977

# Връзка с ФЗ

- Всяка ФЗ е и многозначна зависимост
  - Ако  $A \rightarrow B$ , то  $A \twoheadrightarrow B$
- Доказателство – чрез (1), (2) и (3) от дефиницията на МЗ



## Четвърта нормална форма (4NF)

- Излишеството на данни, което произтича от многозначните зависимости, не може да се отстрани чрез привеждане в BCNF
- Необходима е по-строга нормална форма, наречена 4NF, която третира МЗ като ФЗ по отношение на декомпозицията, но не и по отношение на ключовете

# Преговор

- Дефиниция на BCNF
- Релацията  $R$  удовлетворява BCNF, ако за всяка нетривиална фз  $A \rightarrow B$  е изпълнено, че  $A$  е суперключ

# 4NF – дефиниция

- Релацията  $R$  удовлетворява 4NF, ако за всяка нетривиална МЗ  $A \twoheadrightarrow B$  е изпълнено, че  $A$  е суперключ
  - Понятието ключ се основава на дефиницията на ФЗ

# Пример

- Релация, нарушаваща 4NF
  - Employees (name, phone, skill)

# 4NF – свойства

- След като всяка ФЗ е и МЗ, то привеждането в 4NF осигурява и BCNF
- Всяко нарушение на BCNF е и нарушение на 4NF
- Разбира се, не всяка релация в BCNF е и в 4NF – МЗ са неуловими от BCNF
- Всяка релация с два атрибута е в 4NF

# Декомпозиция в 4NF

- Аналогична на декомпозицията в BCNF
- Ако  $A \twoheadrightarrow B$  нарушава 4NF за релацията  $R$ , то създаваме следните две релации:
  - Първата съдържа атрибутите  $A \cup B$
  - Втората съдържа всички атрибути на  $R$  без тези от  $B$
- Повтаряме процеса, докато осигурим, че всички нови релации са в 4NF

# Пример

- Stars не е в 4NF
- name  $\twoheadrightarrow$  street city
  - Нетривиална MVD
  - name не е суперключ
- Прилагаме декомпозиция:
  - StarAddresses (name, street, city)
  - StarsIn (name, title, year)
- name  $\twoheadrightarrow$  street city в StarAddresses?
- name  $\twoheadrightarrow$  title year в StarsIn?



# Пример – пояснения

- В новите релации има M3, но те са тривиални
- Какво ще се получи при естественото съединение на двете нови релации?

- Както и при BCNF, декомпозицията в 4NF 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 4NF
- Верността на декомпозицията е в сила и при декомпозиция в 4NF
- 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

# Съединение без загуба

- Декомпозицията на  $R$  до  $S(B_1, \dots)$  и  $T(C_1, \dots)$  е без загуба, ако
$$R = S \bowtie T$$
  - т.с.т.к., когато за  $R$  е изпълнена поне една от следните МЗ:
  - $B \cap C \Rightarrow B-C$  или
  - $B \cap C \Rightarrow C-B$

# Вградени МЗ

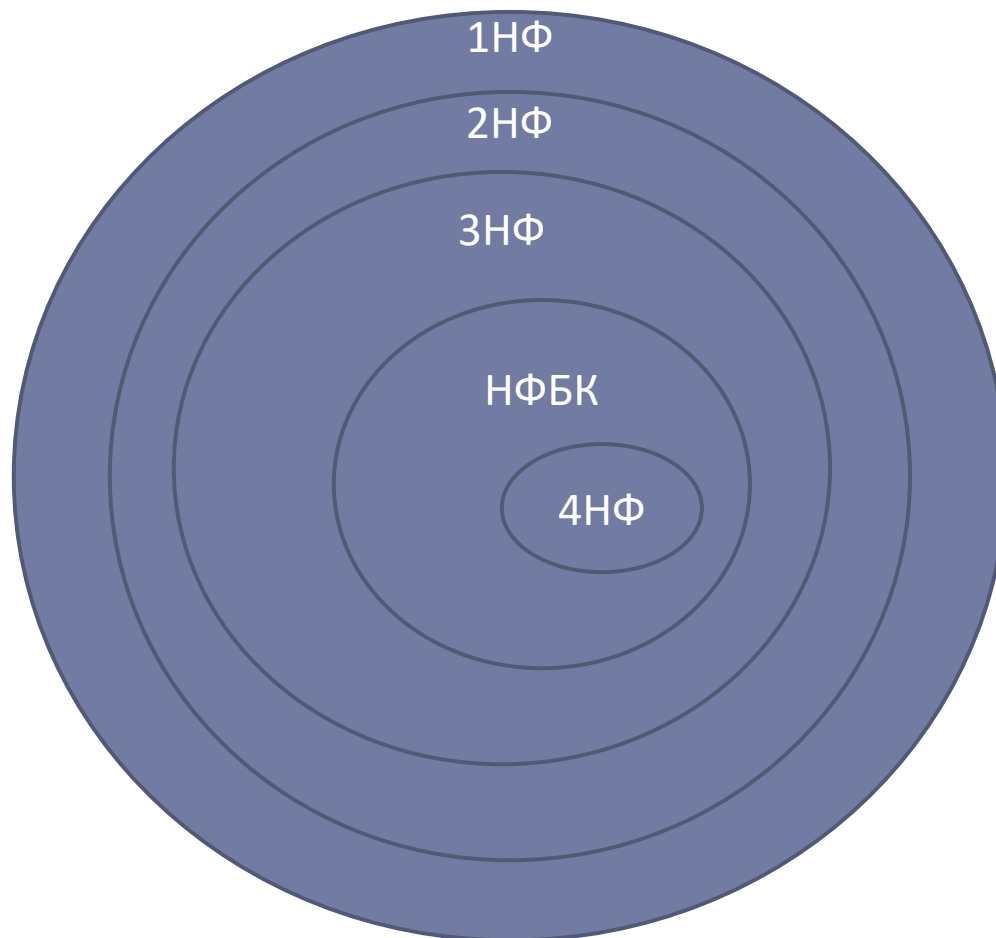
- Нека  $R$  удовлетворява множеството  $D$  от дадени ФЗ и МЗ,  $S$  - проекция на  $R$
- ФЗ, които следват от  $D$  и участват само атрибути на  $S$ , важат и в  $S$
- Аналогично се "проектират" и МЗ, но
- може да има такива МЗ, които са в сила за  $S$ , но не и за  $R$  – вградени

Нормални форми – обобщение

# Нормални форми – обобщение

- **Първа нормална форма (1NF)** – изисква всеки компонент в кортежите да е атомарна стойност
- **Втора нормална форма (2NF)** – за всяка нетривиална ФЗ  $A_1A_2...A_n \rightarrow B$ , която е изпълнена за  $R$ , имаме, че  $B$  е елемент на ключ или  $\{A_1, A_2, ..., A_n\}$  не е собствено подмножество на някой ключ на  $R$
- **Трета нормална форма (3NF)** – ако  $A_1A_2...A_n \rightarrow B$  е нетривиална ФЗ, която е в сила за  $R$ , то или  $\{A_1A_2...A_n\}$  да е суперключ за  $R$ , или  $B$  да е член на някой ключ. Казано по друг начин, всеки атрибут да не е транзитивно зависим от първичния ключ
- **Нормална форма на Бойс-Код (BCNF)** – изисква ако  $A_1A_2...A_n \rightarrow B$  е нетривиална ФЗ, която е в сила за  $R$ , то  $\{A_1A_2...A_n\}$  да е суперключ за  $R$
- **Четвърта нормална форма (4NF)** – изисква ако  $A_1A_2...A_n \twoheadrightarrow B$  е нетривиална МЗ, която е в сила за  $R$ , то  $\{A_1A_2...A_n\}$  е суперключ за  $R$

# Връзка между нормалните форми



# Сравнение

Свойство	3NF	BCNF	4NF
Отсъствие на излишества, породени от ФЗ	В повечето случай	Да	Да
Отсъствие на излишества, породени от МЗ	Не	Не	Да
Запазване на ФЗ	Да	Невинаги	Невинаги
Запазване на МЗ	Невинаги	Невинаги	Невинаги



# Още едно неформално обобщение

> [hackr.io](https://hackr.io)

## Types of Normal Forms in DBMS

	1NF	2NF	3NF	4NF	5NF
Conditions	<b>R</b>	<b>R<sub>11</sub></b> <b>R<sub>12</sub></b>	<b>R<sub>21</sub></b> <b>R<sub>22</sub></b> <b>R<sub>23</sub></b>	<b>R<sub>31</sub></b> <b>R<sub>32</sub></b> <b>R<sub>33</sub></b> <b>R<sub>34</sub></b>	<b>R<sub>41</sub></b> <b>R<sub>42</sub></b> <b>R<sub>43</sub></b> <b>R<sub>44</sub></b> <b>R<sub>45</sub></b>
	Decomposition of Relation				
	Eliminate Repeating Groups	Eliminate Partial Functional Dependency	Eliminate Transitive Dependency	Eliminate Multi-values Dependency	Eliminate Join Dependency

# Връзка с E/R

- Да анализираме дали при преобразуването на E/R модел до релационен се спазват нормалните форми
  - Връзка M:1
  - Връзка M:N
  - Йерархии
  - и т.н.

# Заклучение

- Препоръчва се релационните схеми да са в (поне) 3NF
- Следването на подхода от E/R модел към релационен ни помага да постигнем нормализирана БД

# Допълнителен материал

- Други нормални форми
- Денормализирани бази
  - Предимства на аномалиите  
(недостатъци на нормализацията)

# Упражнения

- Задачи за МЗ
  - Ако  $AC \twoheadrightarrow B$ , в сила ли е  $AC \twoheadrightarrow BC$ ?
  - Ако  $A \twoheadrightarrow B$  и премахнем тези атрибути от  $B$ , които принадлежат и на  $A$ , ще се получи ли отново МЗ?
- Задачи за 4NF

# Въпроси?