

## **Бази даних та інформаційні системи**

### **ЛАБОРАТОРНА РОБОТА №5**

#### **Нормалізація відношень бази даних**

Виконав(ла):

Ст. \_\_Прізвище Ім'я\_\_\_\_\_

Група \_\_\_\_\_

Оцінка

Прийняв:

доц. Малець Р.Б.

**Тема:** Нормалізація відношень бази даних.

**Мета роботи:** Ознайомлення з поняттям нормалізації відношень бази даних та власне самим процесом нормалізації.

## Теоретичний матеріал

Перелік понять, з якими необхідно самостійно ознайомитись, використовуючи лекційний матеріал та додаткову літературу, для виконання завдання лабораторної роботи: 1. Поняття функціональної залежності

- 1.1. Суперключ
- 1.2. Детермінант і залежна частина
- 1.3. Тривіальна, нетривіальна залежності
- 1.4. Повна функціональна залежність
- 1.5. Транзитивна функціональна залежність
2. Багатозначні залежності
3. Надлишковість даних і аномалії поновлення даних в БД.
4. Нормалізація відношень

Основною проблемою реляційної моделі БД є надлишковість даних, яка у свою чергу призводить до аномалій поновлення даних – фактично зникання постійних (для предметної області) даних під час поновлення або ж поява надлишкових кортежів під час загалом коректних сполучень.

Цілі нормалізації: позбавлення аномалій при модифікації даних (insert, update, delete), мінімізація реструктурування бази даних при додаванні нових типів, забезпечення нейтральності відношень до статистики запитів з бігом часу (запити виконуються однаково незалежно від змін в базі даних)

Кажуть, що відношення знаходиться в певній нормальній формі, якщо задовольняється певний набір умов:

**Перша нормальна форма.** Відношення відповідає  $1NF$  тоді, коли є ключовий атрибут, значення атрибутів тільки елементарні (неподільні, атомарні) і не містяться груп, що повторюються.

**Друга нормальна форма.** Відношення знаходиться в  $2NF$ , якщо виконуються обмеження  $1NF$  і кожен неключовий атрибут функціонально повно залежить від первинного ключа (у тому числі і складеного).

**Третя нормальна форма.** Відношення знаходиться у  $3NF$ , якщо виконуються обмеження  $2NF$  і всі неключові атрибути відношення взаємно незалежні і повністю залежать від первинного ключа, тобто кожний неключовий атрибут не транзитивно залежить від ключа.

**Нормальна форма Бойса — Кодда.** Відношення знаходиться в НФБК, тоді і лише тоді коли детермінант кожної функціональної залежності є потенційним ключем. Якщо це правило не виконується, тоді щоб привести вказане відношення до НФБК його слід розділити на два відношення шляхом двох операцій проекції на кожен функціональну залежність детермінант, якої не є потенційним ключем:

1. Проекція без атрибутів залежної частини такої функціональної залежності;
- 2.

Проекція на всі атрибути цієї функціональної залежності.

Визначення НФБК не потребує жодних умов попередніх нормальних форм. Якщо проводити

нормалізацію послідовно, то в переважній більшості випадків при досягненні 3НФ автоматично будуть задовольнятися вимоги НФБК.

3НФ не збігається з НФБК лише тоді, коли одночасно виконуються такі 3 умови: 1.

Відношення має 2 або більше потенційних ключів.

2. Ці потенційні ключі складені (містять більш ніж один атрибут)

3. Ці потенційні ключі перекриваються, тобто мають щонайменше один спільний атрибут.

**Четверта нормальна форма (4NF).** Відношення знаходиться в 4NF, якщо воно знаходиться в BCNF і в ньому відсутні багатозначні залежності, які не є функціональними залежностями.

*Четверта нормальна форма* стосується відношень, в яких є повторювані набори даних. Декомпозиція, заснована на функціональних залежностях, не призводить до виключення такої надмірності. У цьому випадку використовують декомпозицію, засновану на багатозначних залежностях. Багатозначна залежність є узагальненням функціональної залежності і розглядає відповідності між множинами значень атрибутів. Приведення відношення до 4NF дозволяє виключити тип аномалій оновлення. Для приведення відношення з BCNF до 4NF слід виконати проєкції вихідного відношення на пари атрибутів, що створюють багатозначні залежності.

*Четверта нормальна форма (4NF)* є окремим випадком 5НФ, коли повна декомпозиція повинна бути з'єднанням рівно двох проєкцій. Досить не просто підібрати реальну таблицю, яка перебувала б у 4NF, але не була б у 5НФ. Дана Нормальна форма має більший інтерес для теоретичних досліджень, ніж для практики проектування баз даних.

**П'ята нормальна форма (5НФ, 5NF, PJ/NF)** У всіх розглянутих до цього моменту випадках нормалізація відношення відбувається декомпозицією одного відношення на два. Іноді нормалізувати відношення шляхом декомпозиції на два відношення без втрат не вдається, але є можливість декомпозиції вихідного відношення без втрат на більшу кількість відношень, тобто  $n > 2$ .

Якщо в процесі природнього з'єднання отриманих відношень в порівнянні з вихідним відношенням генеруються зайві кортежі, то така декомпозиція характеризується залежністю з'єднання.

В відношенні  $R(X, Y, \dots, Z)$  відсутня залежність з'єднання  $*(X, Y, \dots, Z)$ , в тому і тільки в тому випадку, коли  $R$  відтворюється без втрат шляхом з'єднання своїх проєкцій на  $X, Y, \dots, Z$

*Відношення знаходиться в п'ятій нормальній формі, якщо воно не містить залежностей з'єднання.*

**Шоста нормальна форма.** Таблиця знаходиться у 6nf, якщо вона знаходиться у 5nf та задовольняє вимозі відсутності нетривіальних залежностей. Зазвичай 6nf ототожнюють з dknf

**Доменно-ключова нормальна форма (dknf)** - це нормальна форма, що використовується в нормалізації баз даних і вимагає щоб база даних не містила жодних інших обмежень крім обмежень доменів і обмежень ключів. Обмеження домену - обмеження, яке наказує використовувати для певного атрибуту значення тільки з деякого заданого домену. Обмеження по своїй суті є заданням переліку (або логічного еквіваленту переліку) допустимих значень типу і оголошенням про те, що вказаний атрибут має даний тип. Обмеження ключа – обмеження, яке стверджує, що деякий атрибут чи комбінація атрибутів є потенційним ключем. Будь-яка змінна відношення, що знаходиться в ДКНФ, обов'язково знаходиться в 5НФ. Але не будь-яку змінну відношення можна привести до ДКНФ.

### *Неформальний процес нормалізації:*

1. Відношення у 1НФ розбивають на такі проекції, які дозволяють ліквідувати нескоротні функціональні залежності. В результаті отримуємо 2НФ.
2. Отримані відношення у 2 НФ розбиваються на проекції, які дозволяють ліквідувати транзитивні функціональні залежності. Отримуємо 3НФ.
3. Отримані відношення у 3НФ розбиваються на проекції, у яких ліквідовуються функціональні залежності, де детермінанти не є потенційними ключами. Отримуємо НФБК.
4. Відношення у НФБК розбиваються на проекції, у яких ліквідовано усі багатозначні залежності, що не є функціональними. Отримуємо 4НФ.
5. Відношення у 4 НФ розбиваємо на проекції у яких ліквідовано залежності сполучення, що не визначаються потенційними ключами. Отримуємо 5НФ.

На кожному етапі декомпозиція має відбуватися без втрат і зі збереженням залежностей.

Нормалізація корисна, але її не можна вважати панацеєю, в чому легко переконатися, розглянувши її цілі і те, наскільки добре вона їм відповідає:

- прийти до структури, яка б «добре» представляла реальний світ, тобто була б інтуїтивно зрозуміла і була хорошою основою для майбутнього розширення
- зменшити надлишковість даних
- тим самим уникнути деяких аномалій при оновленні
- спростити формулювання і перевірку деяких обмежень цілісності [5]

### **Хід роботи**

1. Опрацювати теоретичний матеріал.
2. Проаналізувати створені відношення (таблиці) бази даних на відповідність нормальним формам. За необхідності провести нормалізацію відношень до певної форми за допомогою декомпозиції . Пояснити необхідність проведеної нормалізації або чому її не проводили, а залишили відношення без змін. Проілюструвати скрінами.
3. Оформити звіт про виконання лабораторної роботи, який має містити:
  - титульну сторінку;
  - тему, мету та завдання лабораторної роботи;
  - пояснення виконаних дій та скріни екрану з отриманими результатами.

Завантажити звіт у вигляді файлу з іменем: Прізвище-ЛР5.pdf.

Література для ознайомлення з теоретичним матеріалом теми «Нормалізація відношень бази даних»:

1. H.Garcia-Molina, J.D.Ullman, J.Widom Database Systems (2nd Edition). 2009 by Pearson Education Inc. —1203с.
2. Гарсиа-Молина Г., Ульман Д., Уидом Д. Системы баз данных. Полный курс.: Пер. с англ. —М.: Издательский дом “Вильямс”, 2003. —1088с.
3. Дейт, К. Дж. Введение в системы баз данных, 8-е издание.: Пер. с англ. — М.: Издательский дом "Вильяме", 2005. — 1328 с.

4. C. J. Date SQL and Relational Theory: How to Write Accurate SQL Code (3rd edition) 2015 by O'Reilly Media, Inc. —563с.
5. К. Дж. Дейт SQL и реляционная теория. Как грамотно писать код на SQL. – Пер. с англ. – СПб.: Символ-Плюс, 2010. – 480 с.