

Кафедра управління інформаційною безпекою
Спеціальність «Кібербезпека», 5 курс навчання

ЛЕКЦІЯ. НОРМАЛІЗАЦІЯ

План:

1. Постановка задачі.
2. Нормальні форми.
3. Денормалізація.

Література

1. Гайна Г.А. Основи проектування баз даних: Навчальний посібник. К.: КНУБА, 2005. – 204 с.
2. *Гайна Г.А.* Організація баз даних і знань. Мови баз даних: Конспект лекцій.–К.:КНУБА, 2002. – 64 с.
3. *Гайна Г.А., Попович Н.Л.* Організація баз даних і знань. Організація реляційних баз даних: Конспект лекцій.–К.:КНУБА, 2000. – 76 с.

1. Постановка задачі.

Нормалізація – це процедура визначення того, які атрибути зв'язані у відношенні. Одна з головних задач при розробці реляційної БД – об'єднання в одному відношенні тих атрибутів, які зв'язані між собою (між якими є функціональні залежності). Нормалізація являє собою поетапний процес заміни сукупності відношень іншою сукупністю (схемою), в якій відношення мають просту і регулярну структуру. Результатом нормалізації є логічна модель БД.

Надлишковість даних в БД є небажаним явищем, оскільки призводить до збільшення об'єму пам'яті, уповільнює роботу БД. Надлишковість даних є результатом в першу чергу дублювання даних. Розрізняють *незбиткове* та *збиткове* дублювання даних. Повністю усунути надлишковість не потрібно, оскільки при цьому неможливо буде підтримувати БД як єдине ціле. Слід тільки мінімізувати надлишковість, залишивши необхідне дублювання даних.

Дублювання даних створює проблеми при виконанні операцій з БД. Ці проблеми виникають при спробі зробити операції: редагування, додавання або вилучення даних.

Аномаліями називається така ситуація в БД, яка призводить до протиріч у БД, або суттєво ускладнює обробку даних. Розрізняють аномалії модифікації, додавання і вилучення.

Приклад. Розглянемо відношення *Студент* (табл. 1).

Таблиця 1

Студент

Номер залікової книжки	Прізвище	Група	Факультет	Декан
1010	Бойко	ІТП-31	АІТ	Барков
2020	Лемешко	ІУСТ-22	АІТ	Барков
1030	Шевченко	ІТП-31	АІТ	Барков
1121	Петренко	БМО-32	АІТ	Барков
2231	Іванченко	ТБ-21	БТ	Тимчук

Аномалія модифікації виникає при спробі змінити прізвище декана. В цій ситуації необхідно переглянути всі кортежі. При великих розмірах БД це

потребує значного часу, при цьому можливі помилки (у разі невірною введення прізвища), які порушують цілісність БД.

Аномалія додавання виникає при додаванні інформації про нового студента, при цьому необхідно вводити інформацію, яка вже є в БД: назва факультету, прізвище декана. Крім того неможливо створити нову групу поки не існує студентів, які в ній займаються.

Аномалія вилучення виникає при спробі вилучити дані про студента, який в групі поки ще один, наприклад Лемешко. В цьому випадку зникне інформація про групу ІУСТ-22. Виконання декомпозиції наведеного відношення дозволяє позбутися вищезначених аномалій (табл. 2...4).

Таблиця 2

Студент

Номер залікової книжки	Прізвище	Група
1010	Бойко	ІТП-31
2020	Лемешко	ІУСТ-22
1030	Шевченко	ІТП-31
1121	Петренко	БМО-32
2231	Іванченко	ТБ-21

Таблиця 3

Група

Група	Факультет
ІТП-31	АІТ
ІУСТ-22	АІТ
БМО-32	АІТ
ТБ-21	БТ

Таблиця 4

Факультет

Факультет	Декан
АІТ	Барков
БТ	Тимчук

Процес проектування БД з використанням декомпозиції являє собою процес послідовної нормалізації схем відношень, при цьому кожна наступна ітерація відповідає нормальній формі більш високого рівня і має кращі властивості у порівнянні з попередньою. Кожній нормальній формі (НФ) відповідає деякий набір обмежень. Визначають такі нормальні форми: 1НФ, 2НФ, 3НФ, НФБК (нормальна форма Бойса-Кодда), 4НФ, 5НФ.

При виконанні декомпозиції зберігається множина вихідних функціональних залежностей між атрибутами і виконується зворотність. *Зворотність* означає можливість відновлення вихідної схеми. *Функціональні залежності* відображають зв'язки між атрибутами, які властиві реальному об'єкту.

Атрибут B *функціонально* залежить від A , якщо кожному значенню A відповідає в точності одне значення B .

Математичний запис функціональної залежності (ФЗ): $A \rightarrow B$

Приклад. Функціональні залежності:

Студент \rightarrow *Група*; *Група* \rightarrow *Факультет*;

Викладач, *Студент*, *Дисципліна* \rightarrow *Оцінка*.

Якщо існує ФЗ $A \rightarrow B$, то це означає, що у всіх кортежах з однаковим значенням атрибуту A атрибут B буде мати також одне й те ж значення. A і B можуть складатися з декількох атрибутів.

2. Нормальні форми.

Перша нормальна форма. Відношення знаходиться в 1НФ тоді і тільки тоді, коли всі його атрибути є атомарними.

Значення атрибуту вважається *атомарним*, якщо воно є неподільним у всіх застосуваннях.

Приклад. Представлення даних у таблицях може вважатися як атомарним, так і неатомарним залежно від використання. Засіб представлення визначається необхідним ступенем деталізації і повинен підтримуватися у всіх застосуваннях (табл. 5).

Дата народження

Прізвище	Дата народження
Бойко	15 лютого 1991

Прізвище	Дата і місяць	Рік
Бойко	15 лютого	1991

Прізвище	День	Місяць	Рік
Бойко	15	лютий	1991

Друга нормальна форма. Відношення знаходиться в 2НФ, якщо воно знаходиться в 1НФ і кожен його непервинний атрибут функціонально повно залежить від первинного ключа.

Неповною функціональною залежністю називається залежність неключового атрибуту від частини ключа, що складається з декількох атрибутів. Повна функціональна залежність передбачає залежність неключового атрибуту від всіх атрибутів одночасно, що входять до складу ключа.

Приклад. Розглянемо відношення *Студент* (табл. 6).

Таблиця 6

Студент

Номер залікової книжки	Прізвище	Група	Дисципліна	Оцінка
1010	Бойко	ІТП-31	Бази даних	5

Функціональні залежності:

№ залік. кн., Дисципліна → *Прізвище, Група, Оцінка*

№ залік. кн. → *Прізвище, Група*

Для приведення даного відношення до 2НФ необхідно розбити його на проєкції, при цьому повинна бути виконана умова відновлення вихідного відношення без втрат. Проєкції мають такий вигляд (табл. 7...8).

Таблиця 7

Студент

<u>Номер залікової книжки</u>	<u>Дисципліна</u>	<u>Оцінка</u>
1010	Бази даних	5

Таблиця 8

Група

<u>Номер залікової книжки</u>	<u>Прізвище</u>	<u>Група</u>
1010	Бойко	ІТП-31

Відсутність втрат при декомпозиції відношення $R(x,y,z)$ на відношення $R_1(x,y)$ і $R_2(x,z)$ виконується, якщо від спільного атрибуту двох отриманих відношень (x) залежить хоча б один атрибут з двох, що залишилися (y або z), тобто якщо виконується $(x \rightarrow y)$ або $(x \rightarrow z)$.

Третя нормальна форма. Відношення знаходиться в 3НФ, якщо воно знаходиться в 2НФ і жоден з непервинних атрибутів у відношенні не є транзитивно залежним від первинного ключа.

Атрибут C *транзитивно* залежить від атрибуту A , якщо для атрибутів A, B, C виконуються такі умови $A \rightarrow B$ і $B \rightarrow C$, але зворотня залежність відсутня.

Приклад. Розглянемо відношення *Студент* (табл. 9).

Таблиця 9

Студент

<u>Номер залікової книжки</u>	<u>Прізвище</u>	<u>Група</u>	<u>Факультет</u>
1010	Бойко	ІТП-31	АІТ

Функціональні залежності:

$\text{№ залік. кн.} \rightarrow \text{Прізвище, Група, Факультет}$

$\text{Група} \rightarrow \text{Факультет}$

Між атрибутами існує транзитивна залежність. Для того щоби запобігти цьому необхідно виконати декомпозицію відношення (табл. 10, 11):

Студент

<u>Номер залікової книжки</u>	Прізвище	Група
1010	Бойко	ІТП-31

Таблиця 11

Група

<u>Група</u>	<u>Факультет</u>
ІТП-31	АІТ

Нормальна форма Бойса-Кодда. Відношення знаходиться в НФБК, якщо воно знаходиться в ЗНФ і у ньому відсутні залежності атрибутів первинного ключа від неключових атрибутів.

Приклад. Розглянемо відношення *Спеціальність* (табл. 12).

Таблиця 12

Спеціальність

<u>Спеціальність</u>	<u>Дисципліна</u>	<u>Викладач</u>
ІТП	Бази даних	Барко
ІУСТ	Бази даних	Шевченко

Припустимо, що на кожній спеціальності певну дисципліну може викладати тільки один викладач і кожен викладач викладає тільки одну дисципліну. У цьому випадку мають місце такі залежності:

Спеціальність, Дисципліна → *Викладач*

Викладач → *Дисципліна*

Відношення знаходиться в ЗНФ, але неключовий атрибут *Викладач* визначає атрибут *Дисципліна*, що входить у ключ.

Для того щоби позбутися аномалій необхідно виконати декомпозицію відношення (табл. 13, 14).

Таблиця 13

Спеціальність

<u>Спеціальність</u>	<u>Дисципліна</u>
ІТП	Бази даних
ІУСТ	Бази даних

Дисципліна

Викладач	Дисципліна
Барко	Бази даних
Шевченко	Бази даних

Четверта нормальна форма. Відношення знаходиться в 4НФ тоді і тільки тоді, коли у випадку існування багатозначної залежності $A \twoheadrightarrow B$ всі інші атрибути відношення функціонально залежать від A .

У відношенні $R(A,B,C)$ існує багатозначна залежність $A \twoheadrightarrow B$ в тому і тільки в тому випадку, коли множина значень B , що відповідає парі значень A і C залежить тільки від A і не залежить від C .

Відношення $R(A,B,C)$ можна розбити без втрат інформації на відношення $R_1(A,B)$ і $R_2(A,C)$ в тому і тільки в тому випадку, якщо існують багатозначні залежності $A \twoheadrightarrow B$ і $A \twoheadrightarrow C$.

Приклад. Розглянемо відношення *Кафедра* (табл.15).

Кафедра

Кафедра	Викладач	Група
ІТ	Барко	ІТП-31
ІТ	Барко	ІТП-32
ІТ	Шевченко	ІТП-31
ІТ	Шевченко	ІТП-32

У даному відношенні існують дві багатозначні залежності:

Кафедра \twoheadrightarrow *Викладач*

Кафедра \twoheadrightarrow *Група*

Це означає, що кожній кафедрі відповідає перелік викладачів, які на ній працюють і кожній кафедрі відповідає перелік груп, яким ця кафедра викладає дисципліни.

Для того, щоби звести відношення до 4НФ, необхідно виконати його декомпозицію (табл. 16, 17).

Кафедра

Кафедра	Викладач
ІТ	Барко
ІТ	Шевченко

Група

Кафедра	Група
ІТ	ІТІІ-31
ІТ	ІТІІ-32

П'ята нормальна форма. Відношення знаходиться в 5НФ тоді і тільки тоді, коли будь-яка залежність з'єднання у відношенні виходить з існування деякого можливого ключа у відношенні.

Відношення $R(X, Y, \dots, Z)$ задовольняє залежності з'єднання (X, Y, \dots, Z) тоді і тільки тоді, коли R відновлюється без втрат інформації шляхом з'єднання своїх проєкцій на X, Y, \dots, Z . Залежність з'єднання є узагальненням функціональної і багатозначної залежностей.

Приклад. Розглянемо відношення *Заняття*:

Заняття (Студент, Викладач, Дисципліна)

Кожен студент слухає лекції багатьох викладачів, кожен викладач викладає для багатьох студентів, кожен студент вивчає багато дисциплін, кожен викладач викладає багато дисциплін. У відношенні відсутні багатозначні і функціональні залежності й воно знаходиться в 4НФ. У відношенні можливі аномалії, які пов'язані з повтором значень атрибутів в декількох кортежах. Наприклад, якщо студент навчається у багатьох викладачів, то при його відрахуванні з університету необхідно знайти і вилучити декілька записів з відношення.

Утворимо такі складені атрибути відношення:

СВ (*Студент, Викладач*)

СД (*Студент, Дисципліна*)

ВД (*Викладач, Дисципліна*).

Якщо відношення R спроектувати на складені атрибути СВ, СД, ВД, то з'єднання цих проєкцій дасть вихідне відношення. Це означає, що у відношенні *Заняття* існувала залежність з'єднання. Результатом декомпозиції відношення *Заняття* буде отримання таких відношень: $R_1(\text{Студент}, \text{Викладач})$, $R_2(\text{Студент}, \text{Дисципліна})$, $R_3(\text{Викладач}, \text{Дисципліна})$.

Для зведення вихідного відношення до 5НФ виконують його декомпозицію на відношення, кількість яких перевищує два.

Результати зведення до нормальних форм наведені в табл. 18.

Таблиця 18

Правила формування нормальних форм

Нормальні форми	Приклад
1НФ	$R(ABCD)$ – відношення A, B, C, D – атомарні атрибути
2НФ	$R(\underline{AB}CD)$ – відношення, \underline{AB} – ключ, $\underline{AB} \rightarrow CD$, неможливі залежності: $\underline{A} \rightarrow CD, \underline{A} \rightarrow C, \underline{A} \rightarrow D,$ $\underline{B} \rightarrow CD, \underline{B} \rightarrow C, \underline{B} \rightarrow D$
3НФ	$R(\underline{AB}CD)$ – відношення, \underline{AB} – ключ, $\underline{AB} \rightarrow CD$, неможливі залежності: $C \rightarrow D, D \rightarrow C$
Нормальні форми	Приклад
НФБК	$R(\underline{AB}CD)$ – відношення, \underline{AB} – ключ, $\underline{AB} \rightarrow CD$, неможливі залежності: $C \rightarrow \underline{A}, C \rightarrow \underline{B}, D \rightarrow \underline{A}, D \rightarrow \underline{B},$ $C \rightarrow \underline{AB}, D \rightarrow \underline{AB}$
4НФ	$R(\underline{ABC})$ – відношення, $A \twoheadrightarrow B, A \rightarrow C,$ неможливі залежності: $A \twoheadrightarrow C$
5НФ	$R(ABC)$ – вихідне відношення; результат декомпозиції: $R_1(\underline{AB}), R_2(\underline{AC}), R_3(\underline{BC})$

3. Денормалізація.

Денормалізація – модифікація реляційної моделі, при якій ступінь нормалізації модифікованого відношення стає нижче, ніж ступінь нормалізації щонайменше одного з вихідних відношень.

Денормалізація застосовується у тих випадках, коли нормалізована БД не задовольняє вимогам, що висуваються до продуктивності системи. Денормалізація може застосовуватися у таких випадках:

- об'єднання таблиць зі зв'язками "один до одного";
- дублювання неключових атрибутів у зв'язках "один до багатьох" для зменшення кількості з'єднань;
- дублювання атрибутів зовнішнього ключа у зв'язках "один до багатьох" для зменшення кількості з'єднань;
- дублювання атрибутів "багато до багатьох" для зменшення кількості з'єднань;
- створення таблиць з даних, що містяться в інших таблицях;
- введення груп полів, що повторюються.

Застосовуючи денормалізацію слід враховувати, що цей процес має такі негативні наслідки:

- призводить до появи аномалій БД;
- знижує гнучкість системи;
- може зменшити час на відповіді до БД, але при цьому уповільнює операції оновлення даних;
- може ускладнити фізичну реалізацію системи.

Контрольні запитання

1. Дати визначення терміну функціональна залежність. Навести приклади функціональних залежностей.
2. У чому полягає збиткове і незбиткове дублювання даних?
3. Що таке аномалії додавання, оновлення, вилучення?
4. Дати визначення 2НФ. Навести приклад відношення, яке знаходиться в 1НФ, але не знаходиться у 2НФ. Звести його до 2НФ.
5. Дати визначення 3НФ. Навести приклад відношення, яке знаходиться в 2НФ, але не знаходиться у 3НФ. Звести його до 3НФ.
6. Дати визначення НФБК. Навести приклад відношення, яке знаходиться в 3НФ, але не знаходиться у НФБК. Звести його до НФБК.

7. Дати визначення терміну багатозначна залежність. Навести приклади багатозначних залежностей.
8. Дати визначення 4НФ. Навести приклад відношення, яке знаходиться в НФБК, але не знаходиться у 4НФ. Звести його до 4НФ.
9. Дати визначення 5НФ. Навести приклад 5НФ.
10. Яке місце займає нормалізація в процесі проектування бази даних?
11. Що таке денормалізація бази даних і які її переваги і недоліки?
12. Навести приклади проекту бази даних, коли доцільно виконати денормалізацію.

Розробив: к.т.н., доц. Полотай О.І.