



Харківський національний економічний університет імені Семена Кузнеця
(ХНЕУ ім. С. Кузнеця)

Організація і збереження баз даних

Лекція №3

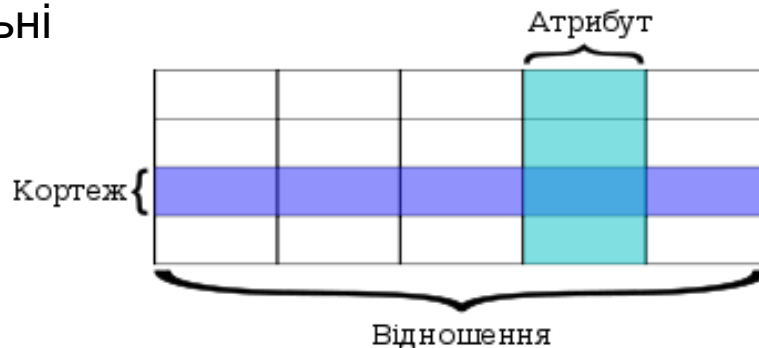
Нормалізація баз даних

Кафедра кібербезпеки та інформаційних
технологій

к.т.н., доц. каф. КІТ
Венгріна Олена Сергіївна

olena.venhrina@hneu.net

- Терміни та визначення.
- Загальні відомості про функціональні залежності.
- Нормальні форми.



Нормалізація являє собою процес розділення даних по окремим пов'язаним таблицям. Нормалізація усуває надлишковість даних (data redundancy) і завдяки цьому дозволяє уникнути порушення цілісності даних під час їх змінювання, тобто уникнути аномалій зміни (update anomaly).

У **ненормалізованій формі** таблиця може зберігати інформацію про дві або більше сутностей. Також вона може вміщувати стовпці що повторюються, або в стовпцях можуть зберігатись повторюванні значення. В **нормалізованій формі** кожна таблиця зберігає інформацію тільки про одну сутність.



Під час нормалізації моделі бази даних слід мати на увазі наступне:

- Найвищій рівень нормалізації не завжди є доцільним.
- Чим вище нормальна форма, тим більше ресурсів потрібно для виконання запиту.
- Не у кожного замовника є можливість фінансувати створення «ідеальної» бази даних.
- Іноді бувають випадки коли модель бази даних потрібно навпаки денормалізувати.



Ключове поняття нормалізації - функціональна залежність.

Функціональна залежність – концепція, що лежить в основі багатьох питань, пов'язаних з реляційними базами даних, включаючи, зокрема, їхнє проектування. Математично являє собою бінарне відношення між множинами атрибутів даного відношення і є, по суті, зв'язком «один-до-багатьох». Функціональна залежність забезпечує основу для наукового підходу до розв'язання деяких проблем, оскільки володіє багатим набором цікавих формальних властивостей.



6

Функціональна залежність

Course	Teacher	Position
Математика	Шевченко Тарас	доцент
Теорія алгоритмів	Шевченко Тарас	доцент
Програмування	Франко Іван	професор



Функціональна залежність

$$r(R), A \subseteq R, B \subseteq R \\ (A \rightarrow B) \Leftrightarrow ((\forall t_1, t_2 \in r : t_1(A) = t_2(A)) \Rightarrow (t_1(B) = t_2(B)))$$

Функціональна залежність вважається тривіальною за умови

$$(B \subseteq A) \Rightarrow (A \rightarrow B)$$



Замикання множин залежностей та атрибутів

8

Замикання множини залежностей

$$(A \rightarrow B) \wedge (B \rightarrow C) \Rightarrow (A \rightarrow C)$$

Замикання множини атрибутів

$$r(R), S, Z \subseteq R, A_i \subseteq R, i = \overline{1, n}$$
$$Z^+ = \{A_i : (Z \rightarrow A_i) \in S^+\}$$

Функціональна залежність



Незвідні множини залежностей

9

Якщо будь-яка функціональна залежність з S_1 входить і в S_2 , тоді S_2 називають покриттям функціональних залежностей S_1

Якщо S_2 – покриття для S_1 , а S_1 – для S_2 (тобто $S_1^+ = S_2^+$), тоді такі множини називаються еквівалентними.

ID_студента	Прізвище	Ім'я	Група	Спеціальність
1001	Сидоров	Іван	Гр-101	Інформатика
1002	Петров	Олег	Гр-102	Математика
1003	Коваленко	Марія	Гр-103	Фізика
1004	Іванов	Анна	Гр-101	Інформатика
1005	Гриценко	Павло	Гр-104	Хімія

Функціональна залежність



Множина функціональних залежностей S називається **незвідною** тоді і тільки тоді, коли виконуються наступні вимоги:

- В кожній ФЗ залежна частина містить лише один елемент.
- Детермінант кожної ФЗ є незвідним (ні один атрибут не може бути видаленим з детермінанта без зміни замкнення S^+)
- Жодну ФЗ з S не можна виключити без зміни замкнення S^+

Для будь-якої множини ФЗ існує не менше ніж одна еквівалентна множина, яка є незвідною. Така множина називається **незвідним покриттям**.



Правила виводу Армстронга.

11

$$r(R) \quad A, B, C, D \subseteq R.$$

Правила виводу Армстронга:

- Рефлексивність $(B \subseteq A) \Rightarrow (A \rightarrow B)$
- Поповнення. $(A \rightarrow B) \Rightarrow (AC \rightarrow BC)$
- Транзитивність. $(A \rightarrow B) \wedge (B \rightarrow C) \Rightarrow (A \rightarrow C)$



Правила виводу Армстронга.

12

Правила виводу Армстронга можна охарактеризувати як:

- **Повні** – з їхньою допомогою можна вивести решту ФЗ, що припускаються даною множиною).
- **Надійні** – «зайві» ФЗ вивести не можна, виведена ФЗ справедлива всюди, де справедлива та множина ФЗ, з якої вона була виведена.



Крім того з даних правил можна вивести декілька додаткових правил, які спрощують задачу виведення ФЗ:

- Самовизначення $A \rightarrow A$
- Декомпозиція $(A \rightarrow BC) \Rightarrow (A \rightarrow B) \wedge (A \rightarrow C)$
- Об'єднання $(A \rightarrow B) \wedge (A \rightarrow C) \Rightarrow (A \rightarrow BC)$
- Композиція $(A \rightarrow B) \wedge (C \rightarrow D) \Rightarrow (AC \rightarrow BD)$



- Нехай X – множина атрибутів, яка в решті-решт стане замкненням.
- $B_1 B_2 \dots B_m \rightarrow C, \quad B_1 B_2 \dots B_m \subseteq X, \quad C \notin X.$
- Повторюємо пункт 2, доки до множини X буде неможливо додавати атрибути.
- Множина X , до якої неможливо буде додавати атрибути і буде замкненням.



- Велика кількість ФЗ уповільнює роботу бази даних, зменшення кількості ФЗ може призвести до виникнення помилок в даних.
- Рішення – оптимізація кількості ФЗ, перевірка яких буде гарантувати відсутність помилок.
- Оптимальна кількість – це незвідне покриття початкової множини ФЗ.



- Нормалізація передбачає застосування нормальних форм до структури даних.
- Існує 7 нормальних форм.
- Кожна нормальна форма (за виключенням першої) передбачає, що до даних вже було застосовано попередню нормальну форму.



- Перша нормальна форма (1НФ, 1NF) – є основою для структурованої схеми бази даних.
- Друга нормальна форма (2НФ, 2NF) – вимагає аби дані, які зберігаються в таблицях із композитним ключем, не залежали лише від частини ключа.
- Третя нормальна форма (3НФ, 3NF) - вимагає, аби дані в таблиці винятково від основного ключа.
- Нормальна форма Бойса-Кодда (НФБК, BCNF) – більш сувора версія 3НФ.



- Четверта нормальна форма (4НФ, 4NF) - потребує, аби в схемі баз даних не було нетривіальних багатозначних залежностей множин атрибутів від будь чого, окрім надмножини ключа-кандидата.
- П'ята нормальна форма (5НФ, 5NF, PJ/NF) - вимагає, аби не було нетривіальних залежностей об'єднання, котрі б не витікали із обмежень ключів.
- Нормальна форма домен/ключ (DKNF) – ця нормальна форма вимагає, аби в схемі не було інших обмежень окрім ключів та доменів. Зазвичай ця форма ототожнюється з шостою нормальною формою (6НФ, 6NF)



Критерії:

- Кожна таблиця повинна мати основний ключ, мінімальний набір колонок, які ідентифікують запис.
- Уникнення повторень груп (категорії даних, що можуть зустрічатись різну кількість разів в різних записах) правильно визначаючи неключові атрибути.
- Атомарність: кожен атрибут повинен мати лише одне значення, а не множину значень.

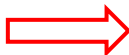


Перша нормальна форма. Приклад.

20

Початкова ненормалізована таблиця

Викладач	Дисципліна
Бондаренко О. О.	«Українська ділова мова»
Коваленко С. М.	«Інформатика» «Системи баз даних»
Ткач Е. М.	«Управління ризиками»
Вакуленко Н. С.	«Вища математика»



Таблиця, приведена до 1НФ

Викладач	Дисципліна
Бондаренко О. О.	«Українська ділова мова»
Коваленко С. М.	«Інформатика»
Коваленко С. М.	«Системи баз даних»
Ткач Е. М.	«Управління ризиками»
Вакуленко Н. С.	«Вища математика»

Нормальні форми



Критерії:

- Схема бази даних повинна відповідати всім вимогам 1НФ.
- Дані, що повторно з'являються в декількох рядках, мають бути винесені в окремі таблиці.



Друга нормальна форма. Приклад 1.

22

Початкова таблиця в 1НФ

Викладач	Дисципліна	Посада
Бондаренко О. О.	«Українська ділова мова»	Доцент
Коваленко С. М.	«Інформатика»	Доцент
Коваленко С. М.	«Системи баз даних»	Доцент
Ткач Е. М.	«Управління ризиками»	Асистент
Вакуленко Н. С.	«Вища математика»	Професор
Вакуленко Н. С.	«Інформатика»	Професор

Таблиці, приведені до 2НФ

Викладач	Посада
Бондаренко О. О.	Доцент
Коваленко С. М.	Доцент
Ткач Е. М.	Асистент
Вакуленко Н. С.	Професор

Викладач	Дисципліна
Бондаренко О. О.	«Українська ділова мова»
Коваленко С. М.	«Інформатика»
Коваленко С. М.	«Системи баз даних»
Ткач Е. М.	«Управління ризиками»
Вакуленко Н. С.	«Вища математика»
Вакуленко Н. С.	«Інформатика»

Нормальні форми



Друга нормальна форма. Приклад 2.

23

Таблиця в 2НФ

Житловий комплекс	Рік введення в експлуатацію	Забудовник	Дата отримання ліцензії
«Казка»	2021	ЖС1	01.06.2020
«Мангеттен».	2023	ЖС2	08.05.2021
«Нодус»	2023	ЖС2	08.05.2021
«Баухаус»	2024	ЖС3	05.12.2021

Нормальні форми



24

Таблиці в 2НФ

Виробник	Модель	Повна назва	Країна виробник
Volkswagen	Polo	Volkswagen Polo	Німеччина
Fiat	Tipo	Fiat Tipo	Італія
Volkswagen	Golf	Volkswagen Golf	Німеччина
Audi	Q5	Audi Q5	Німеччина

Виробник	Країна виробник
Volkswagen	Німеччина
Fiat	Італія
Audi	Німеччина

Виробник	Модель	Повна назва
Volkswagen	Polo	Volkswagen Polo
Fiat	Tipo	Fiat Tipo
Volkswagen	Golf	Volkswagen Golf
Audi	Q5	Audi Q5



Критерії:

- Схема бази даних повинна відповідати всім вимогам другої нормальної форми.
- Будь-яке поле, яке залежить від основного ключа та від будь-якого іншого поля, має вноситись в окрему таблицю.



Третя нормальна форма. Приклад

26

Таблиця в 2НФ

Житловий комплекс	Рік введення в експлуатацію	Забудовник	Дата отримання ліцензії
«Казка»	2021	ЖС1	01.06.2020
«Мангеттен».	2023	ЖС2	08.05.2021
«Нодус»	2023	ЖС2	08.05.2021
«Баухаус»	2024	ЖС3	05.12.2021

Таблиці в 3НФ

Житловий комплекс	Рік введення в експлуатацію	Забудовник
«Казка»	2021	ЖС1
«Мангеттен».	2023	ЖС2
«Нодус»	2023	ЖС2
«Баухаус»	2024	ЖС3

Забудовник	Дата отримання ліцензії
ЖС1	01.06.2020
ЖС2	08.05.2021
ЖС2	08.05.2021
ЖС3	05.12.2021

Нормальні форми



Критерії:

- Детермінант кожної функціональної залежності є потенційним ключем.

Якщо нормалізація виконується послідовно, то в переважній більшості випадків при досяганні 3НФ автоматично будуть задовольнятися умови НФБК.



ЗНФ не збігається з НФБК коли одночасно виконуються наступні три умови:

- Відношення має 2 або більше потенційних ключів.
- Ці потенційні ключі складні (містять більше ніж один атрибут).
- Ці потенційні ключі перекриваються, тобто мають щонайменше один спільний атрибут.

Для приведення до НФБК необхідно таблицю розділити на дві шляхом наступних операцій проєкції на кожну ФЗ, детермінант якої не є потенційним ключем:

- Проєкція без атрибутів залежної частини такої функціональної залежності.
- Проєкція на всі атрибути цієї функціональної залежності.



Таблиця в 3НФ але не в НФБК

Замовлення на день

Сектор	Час початку	Час завершення	Програма лояльності
Зал	9:00	11:00	«-5 %»
Зал	11:30	13:30	«-5 %»
Зал	14:00	16:00	«Стандарт»
Тераса	8:30	10:30	«Стандарт +»
Тераса	11:00	13:00	«Стандарт +»
Тераса	15:00	20:00	«-10%»

Тип оплати

Замовлення на день

Час початку	Час завершення	Програма лояльності
9:00	11:00	«-5 %»
11:30	13:30	«-5 %»
14:00	16:00	«Стандарт»
8:30	10:30	«Стандарт +»
11:00	13:00	«Стандарт +»
15:00	20:00	«-10%»

- 1НФ – всі атрибути відношення атомарні.
- 2НФ – 1НФ + відсутні часткові функціональні залежності.
- 3НФ – 2НФ + відсутні транзитивні залежності неключових атрибутів від ключа.
- НФБК – 3НФ + детермінанти всіх функціональних залежностей є потенційними ключами.
- 4НФ – 3НФ + відсутні нетривіальні багатозначні залежності.
- 5НФ – 4НФ + відновлення з проєкцій відтворює початкове відношення.
- ДКНФ – відношення не має аномалій модифікації.
- 6НФ – 5НФ + проєкції зберігають тимчасові дані.



1. Нормалізація відношень.
2. Нормальні форми.
3. Поняття функціональної залежності.
4. Перша нормальна форма.
5. Друга нормальна форма.
6. Третя нормальна форма.

