

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ**

**Харківський національний  
університет внутрішніх справ**

**Кафедра інформаційних технологій факультету № 4**

## **ТЕКСТ ЛЕКЦІЇ**

**з навчальної дисципліни**

**«Теорія розподілених інформаційних ресурсів, захист баз даних»  
вибіркових компонент освітньої програми  
другого (магістерського) рівня вищої освіти**

**125 «Кібербезпека» (Безпека інформаційних та комунікаційних систем)**

**за темою – «Процеси нормалізації та основні її положення»**

**Харків 2018**

**ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від \_\_\_\_\_ № \_\_\_\_

**СХВАЛЕНО**

Вченою радою  
факультету № 4  
Протокол від \_\_\_\_\_ № \_\_\_\_

**ПОГОДЖЕНО**

Секцією Науково-методичної ради  
ХНУВС з технічних дисциплін  
Протокол від \_\_\_\_\_ № \_\_\_\_

Розглянуто на засіданні кафедри інформаційних технологій факультету № 4 (*протокол від \_\_\_\_\_ № \_\_\_\_*).

**Розробники:**

1. Доцент кафедри інформаційних технологій факультету № 4, кандидат технічних наук, доцент Тулупов В.В.

**Рецензенти:**

1. Професор кафедри кібербезпеки факультету № 4, кандидат технічних наук, доцент Носов В.В.
2. Завідувач кафедри систем інформації Національного технічного університету «Харківський політехнічний інститут», д.т.н., професор Серков О.А.

## План лекції

1. Мета нормалізації
2. Процес нормалізації

## Рекомендована література:

### Основна:

1. Дейт К.Дж. Введение в системы баз данных. 6-е издание. - Киев – Москва: Діалектика, 1998. - 784 с.
2. Хансен Г., Хансен Дж. Базы данных: разработка и управление. – Москва: Бином, 1999. - 700 с. Пер. с англ.
3. Коннолли Т., Бегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика. 2-е издание. Москва – Санкт – Петербург - Киев: Вильямс, 2000. - 1111 с.

### Додаткова:

4. Д. Кренке. Теория и практика построения баз данных. 8-е изд. – СПб.: Питер, 2003. – 800 с.: ил. – (Серия «Классика Computer Science»).
5. Гарсиа-Молина, Гектор, Ульман, Джефри, Д., Уидом, Дженифер. Системы баз данных. Полный курс.: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 1088 с.: ил.
6. Кириллов В.В. Основы проектирования реляционных баз данных. Учебное пособие. Санкт-Петербургский Государственный институт точной механики и оптики (технический университет). Кафедра вычислительной техники. - <http://www.cs.ifmo.ru>
7. Кузнецов С.Д. Основы современных баз данных. Информационно-аналитические материалы. - <http://www.citmg.ru/>

## Текст лекції

При проектуванні бази даних у реляційній СУБД основною метою розробки логічної моделі даних є створення точного представлення даних, зв'язків між ними і необхідними обмеженнями. Для досягнення цієї мети необхідно, насамперед, визначити придатний набір відношень. Метод, що використовується для рішення останньої задачі, зветься **нормалізацією** (normalization). Дуже важливо правильно розуміти призначення методів нормалізації і способи їхнього ефективного застосування на практиці.

### 1.1. Мета нормалізації

Нормалізація - **метод створення набору відношень із заданими властивостями на основі вимог до даних, встановлених у деякій організації.**

Процес нормалізації був уперше запропонований **Э. Ф. Коддом (Codd, 1972, б)**. Нормалізація часто виконується у виді послідовності тестів для деякого відношення з метою перевірки його відповідності (чи невідповідності) вимогам заданої нормальної форми. Спочатку були запропоновані тільки три види нормальних форм: перша (1НФ), друга (2НФ) і третя (3НФ). Потім Р. Бойсом і Э. Ф. Коддом (Codd, 1974) було сформульовано більш строге визначення третьої нормальної форми, що одержало назву нормальної форми Бойса-Кодда (НФБК). Усі ці нормальні форми засновані на функціональних залежностях, що існують між атрибутами відношень.

Слідом за НФБК з'явилися визначення четвертої (4НФ) і п'ятої (5НФ) нормальних форм. Однак на практиці ці нормальні форми більш високих порядків використовуються вкрай рідко.

Процес нормалізації є формальним методом, що дозволяє ідентифікувати відношення на основі їх первинних ключів (чи потенційних ключів, як у випадку НФБК) і функціональних залежностей, що існують між їхніми атрибутами.

### 1.2. Процес нормалізації

Нормалізація - **це формальний метод аналізу відношень на основі їх первинного ключа і існуючих функціональних залежностей.** Він включає ряд правил, що можуть використовуватися для перевірки окремих відношень таким чином, щоб уся база даних могла бути нормалізована до

бажаного ступеня нормалізації. Якщо деяка вимога не задовольняється, то відношення, що порушує дану вимогу, повинне бути **декомпозоване** на відношення, кожне з яких (окремо) задовольняє усім вимогам нормалізації.

Найчастіше нормалізація здійснюється декілька етапів, кожний з яких відповідає деякій нормальній формі, що володіє відомими властивостями. У ході нормалізації формат відношень стає усе більш строгим і менш уразливим стосовно аномалій відновлення. При роботі з реляційною моделлю даних важливо розуміти, що тільки задоволення вимог першої нормальної форми (1НФ) обов'язково для створення відношень прийнятної якості. Всі інші форми можуть використовуватися за бажанням проектувальників. Однак, для того щоб уникнути аномалій відновлення, описаних раніше, нормалізацію рекомендується виконувати як **мінімум** до 3НФ.

### **1.3. Перша нормальна форма (1 НФ)**

Перед обговоренням першої нормальної форми доцільно попередньо, дати визначення того стану, що передуює їй.

Ненормалізована форма (ННФ) - таблиця, що містить одну чи кілька повторюваних груп даних.

Перша нормальна форма (1НФ) - відношення, у якому на перетинанні кожного рядка і кожного стовпця міститься тільки одне значення.

У цій главі процес нормалізації починається з перетворення даних з формату джерела (наприклад, з формату стандартної форми введення даних) у формат таблиці з рядками і стовпцями. На вихідному етапі таблиця знаходиться в ненормалізованій формі (ННФ) і часто називається **ненормалізованою таблицею**.

Для перетворення ненормалізованої таблиці в першу нормальну форму (1НФ) у вихідній таблиці варто знайти й усунути всі повторювані групи даних.

**Повторюваною групою називається група, що складається з одного і більш атрибутів таблиці, у якій можливе наявність декількох значень для єдиного значення ключового атрибута таблиці.** Зверніть увагу на те, що в даному контексті термін "ключ" так само відноситься і до одного атрибута, і до групи атрибутів, що єдиним образом ідентифікують кожен рядок ненормалізованої таблиці.

### **1.3. Друга нормальна форма (2НФ)**

Друга нормальна форма (2НФ) заснована на понятті повної функціональної залежності, що описується нижче.

#### **1.3.1. Повна функціональна залежність**

Повна функціональна залежність **У** деяким відношенні атрибут **В** називається **цілком функціонально залежним від атрибута А**, якщо атрибут **В** функціонально залежить від повного значення атрибута **А** и не залежить ні від якої підмножини повного значення атрибута **А**.

Функціональна залежність **А → В** є **повною** функціональною залежністю, якщо видалення якого-небудь атрибута з **А** приводить до втрати цієї залежності.

**Частковою** функціональною залежністю називається така залежність **А → В**, якщо в **А** є деякий атрибут, при видаленні якого ця залежність зберігається.

Наприклад, розглянемо наступну функціональну залежність:

**П\_Номер, Прізвище → Ф\_Номер**

Тут кожна пара значень (**П\_Номер, Прізвище**) зв'язана з єдиним значенням **Ф\_Номер**. Однак ця функціональна залежність не є повною, оскільки **Ф\_Номер** також функціонально залежить від підмножини (**П\_Номер, Прізвище**), тобто від атрибута **П\_Номер**. Інші приклади повної і часткової функціональної залежностей описуються далі.

#### **Визначення другої нормальної форми**

Друга нормальна форма застосовується до **відношень зі складеними ключами**, тобто до таких відношень, первинний ключ яких складається з двох чи більше атрибутів. Справа в тім, що відношення з первинним ключем на основі єдиного атрибута завжди знаходиться, принаймні, у 2НФ.

**Друга нормальна форма (2НФ)** - відношення, що знаходиться в першій нормальній формі і кожен атрибут якого, що не входить до складу первинного ключа, характеризується повною функціональною залежністю від цього первинного ключа.

**Нормалізація 1НФ** - відношень з утворенням 2НФ - відношень включає усунення часткових залежностей. Якщо у відношенні між атрибутами існує часткова залежність, то функціонально-залежні атрибути віддаляються з нього і містяться в нове відношення разом з копією їх детермінанта.

### Приклад Друга нормальна форма (2НФ)

О_Номер	Н_Номер	Орендар_Прізвище	Вулиця	Початок_оренди	Кінець_оренди	Макс_Рента	В_Номер	Власник_Прізвище
033333	H22222	Апрогамов	вул.Толстого,23	01.01.2004	01.01.2005	400,00 грн.	B22222	Власенко
044444	H22222	Вагонков	вул.Толстого,23	12.12.1999	31.12.2004	500,00 грн.	B22222	Власенко
011111	H44444	Воскобойник	вул.Перемоги,39	11.11.2000	12.12.2001	350,00 грн.	B88888	Лопаків
077777	H33333	Охрименко	вул.Достоевського,45	12.12.2001	12.12.2002	360,00 грн.	B44444	Даниленко
022222	H77777	Мохровський	вул.Докучаєва,39	12.12.2005	12.12.2006	200,00 грн.	B66666	Уваров
055555	H88888	Дмитренко	вул.Антонова,56	02.02.2001	12.12.2001	650,00 грн.	B55555	Янко
066666	H66666	Рубовський	вул.Празська,21	03.03.2001	12.12.2001	420,00 грн.	B77777	Донченко

На рис.11.4 показане відношення оренди нерухомості. Покажемо функціональні залежності (від **fd1** до **fd6**) для

відношення **Оренда** з парою атрибутів **О\_Номер, Н\_Номер** в якості первинного ключа. Відношення **Оренда** володіє наступними функціональними залежностями:

**fd1: О\_Номер, Н\_Номер →**

Початок\_оренди, Кінець\_оренди(Первинний ключ)

**fd2: О\_Номер →**

Орендар\_Прізвище (Часткова залежність)

**fd3: Н\_Номер →**

Вулиця, Макс\_Рента, В\_Номер, Власник\_Прізвище (Часткова залежність)

**fd4: В\_Номер →** Власник\_Прізвище (Транзитивна залежність)

**fd5: О\_Номер, Початок\_оренди →**

Н\_Номер, Вулиця, Макс\_Рента (Потенційний ключ)

**fd6: Н\_Номер, Початок\_оренди →**

О\_Номер, Орендар\_Прізвище, Кінець\_оренди (Потенційний ключ)

Після виявлення функціональних залежностей процес нормалізації відношення **Оренда** продовжується перевіркою його приналежності до другої нормальної форми. Для цього потрібно знайти хоча б один випадок часткової залежності від первинного ключа. Неважко помітити, що атрибут імені клієнта **Орендар\_Прізвище** частково залежить від первинного ключа, інакше кажучи, він залежить тільки від атрибута **О\_Номер** (ця залежність представлена вище як fd2). Крім того, атрибути об'єкта нерухомості (**Вулиця, Макс\_Рента, В\_Номер, Власник\_Прізвище**) також частково залежать від первинного ключа, але цього разу тільки від атрибута **Н\_Номер** (ця залежність представлена вище як fd3). У свою чергу, атрибути орендованих об'єктів нерухомості (**Початок\_оренди** і **Кінець\_оренди**) цілком функціонально залежать від первинного ключа в цілому, тобто від атрибутів **О\_Номер** і **Н\_Номер** (ця залежність представлена вище як fd1).

Зверніть увагу на те, що на мал. 6.5 показане **наявність транзитивної залежності** (transitive dependence) від первинного ключа (ця залежність представлена вище як fd4). Хоча транзитивна залежність також може послужити причиною аномалій відновлення, проте її присутність у відношенні не порушує обмежень для 2НФ. Такі залежності будуть усунуті при переході до 3НФ.

Виявлення часткових залежностей усередині відношення **Оренда** позначає, що дане відношення не знаходиться в другій нормальній формі. Для перетворення відношення **Оренда** у 2НФ необхідно створити нові відношення, причому так, щоб атрибути, що не входять у первинний ключ, були переміщені в них разом з копією частини первинного ключа, від якої вони функціонально залежать. Застосування цього правила в нашому випадку приведе до створення трьох нових відношень – **Орендар, Власники\_Нерухомості, Облік\_оренди**, що представлені на наступному малюнку відповідно.

Орендар : таблиця		Облік_Оренди : таблиця			
О_Номер	Прізвище	Нерухомість	Орендар	Початок_оренди	Кінець_оренди
Q11111	Воскобойник	H22222	O33333	01.01.2004	01.01.2005
Q22222	Мохровський	H22222	O44444	12.12.1999	31.12.2004
O33333	Апрогамов	H44444	O11111	11.11.2000	12.12.2001
O44444	Вагонков	H33333	O77777	12.12.2001	12.12.2002
O55555	Дмитренко	H11111	O88888	12.12.2003	12.12.2004
O66666	Рубовський	H77777	O22222	12.12.2005	12.12.2006
O77777	Охрименко	H88888	O55555	02.02.2001	12.12.2001
*		H66666	O66666	03.03.2001	12.12.2001
Запись: 1		Запись: 1 из 8			

  

Власники_нерухомості : таблиця				
Н_Номер	Нерухомість_адреса	Рента	В_Номер	Власник_Прізвище
H11111	вул.Остроградського,76	230,00 грн.	B33333	Іванов
H22222	вул.Толстого,23	330,00 грн.	B22222	Власенко
H33333	вул.Достоевського,45	250,00 грн.	B44444	Даниленко
H44444	вул.Перемоги,39	400,00 грн.	B88888	Лопаків
H55555	вул.Європейська,56	320,00 грн.	B11111	Петренко
H66666	вул.Празська,21	510,00 грн.	B77777	Донченко
H77777	вул.Докучаєва,39	450,00 грн.	B66666	Уваров
H88888	вул.Антонова,56	360,00 грн.	B55555	Янко
Запись: 1 из 8				

Тепер ці три відношення знаходяться в другій нормальній формі, оскільки кожен атрибут, що не входить в первинний ключ, цілком функціонально залежить від первинного ключа відношення.

### Третя нормальна форма (ЗНФ)

Хоча ЗНФ - відношення в меншому ступені мають надмірність даних, ніж 1НФ-відношення, вони усе ще можуть страждати від аномалій відновлення. Так, при спробі відновлення імені власника нерухомості буде потрібно оновити у декількох рядках відношення

**Власники\_Нерухомості**, представленого на попередньому рисунку. Якщо оновити тільки один з цих двох рядків, база даних потрапить у суперечливий стан. Ця аномалія відновлення викликається транзитивною залежністю, що є присутньою у даному відношенні. Вона може бути усунута шляхом приведення даного відношення до третьої нормальної форми. У цьому розділі транзитивні залежності розглядаються разом із третьою нормальною формою.

#### 2.1. Транзитивна залежність

Транзитивна залежність Якщо для атрибутів А, В і С деякого відношення існують залежності виду  $A \rightarrow B$  і  $B \rightarrow C$ , то говорять, що атрибут С транзитивно залежить від атрибута А через атрибут В (за умови, що атрибут А функціонально не залежить ні від атрибута В, ні від атрибута С).

Транзитивна залежність є описом такого типу функціональної залежності, що виникає при наявності наступних функціональних залежностей між атрибутами А, В і С:

$$A \rightarrow B \text{ и } B \rightarrow C.$$

У даному випадку транзитивна залежність  $A \rightarrow C$  здійснюється через атрибут В. Це твердження справедливе тільки в тому випадку, якщо атрибут А функціонально не залежить від атрибутів В і С. Наприклад, розглянемо наступні функціональні залежності усередині відношення **Персонал\_Філія**, представленого на рис.3.1.

$$П\_Номер \rightarrow Ф\_Номер \text{ і } Ф\_Номер \rightarrow Вулиця$$

У цьому випадку транзитивна залежність  $П\_Номер \rightarrow Вулиця$  здійснюється через атрибут **Ф\_Номер**. Дане твердження справедливе, оскільки атрибут **П\_Номер** не залежить функціонально від атрибутів **Ф\_Номер** і **Вулиця**.

##### 2.1.1. Визначення третьої нормальної форми

Третя нормальна форма (ЗНФ). Відношення, що знаходиться в першій і другий нормальних формах і не має не вхідних у первинний ключ атрибутів, що знаходилися б у транзитивній функціональній залежності від цього первинного ключа.

Нормалізація 2НФ - відношень з утворенням ЗНФ - відношень включає усунення транзитивних залежностей. Якщо у відношенні існує транзитивна залежність між атрибутами, у такому випадку транзитивно залежні атрибути відаляються з нього і містяться в нове відношення разом з копією їхнього детермінанта.

#### Відношення Орендар

$$fd2 \quad O\_Номер \rightarrow Прізвище$$

#### Відношення Облік\_Оренди

$$fd1 \quad O\_Номер, Н\_Номер \rightarrow Початок\_оренди, Кінець\_оренди$$

**fd5** O\_Номер, Початок\_оренди → Н\_Номер, Кінець\_оренди

**fd6** Н\_Номер, Початок\_оренди → O\_Номер, Кінець\_оренди

#### Відношення Власники Нерухомості

**fd3** Н\_Номер → Адреса, Рента, В\_Номер, Власник\_Прізвище

**fd4** В\_Номер → Власник\_Прізвище

#### **Приклад 11.5. Третя нормальна форма (ЗНФ)**

Спочатку розглянемо функціональні залежності, що існують у відношеннях **Облік\_Оренди**, **Орендар** і **Власники\_Нерухомості**.

Усі не входні в первинний ключ атрибути відношень **Облік\_Оренди**, **Орендар** функціонально залежні тільки від **їхніх** первинних ключів. Отже, відношення **Облік\_Оренди**, **Орендар** не мають транзитивних залежностей, а тому вони вже знаходяться в третій нормальній формі (ЗНФ).

Усі не входні в первинний ключ атрибути відношення **Власники\_Нерухомості** функціонально залежать від первинного ключа, за винятком атрибута **Власник\_Прізвище**, що також залежить і від атрибута **В\_Номер** (залежність fd4). Це типовий приклад транзитивної залежності, що має місце при наявності залежності не входного в первинний ключ атрибута (**Власник\_Прізвище**) від одного чи декількох інших атрибутів, що також не входять у первинний ключ (**В\_Номер**).

Для перетворення відношення **Власники\_Нерухомості** у третю нормальну форму необхідно насамперед видалити згадану вище транзитивну залежність шляхом створення двох нових відношень **Власність\_в\_оренду** і **Власник**, що представлені на наступному малюнку. Нові відношення мають вигляд:

Власність_в_оренду : таблиця				Власник : таблиця	
Н_Номер	Вулиця	Рента	Власник	В_Номер	Прізвище
H11111	вул.Остроградського,76	230,00 грн.	B33333	B11111	Петренко
H22222	вул.Толстого,23	330,00 грн.	B22222	B22222	Власенко
H33333	вул.Достоевського,45	250,00 грн.	B44444	B33333	Іванов
H44444	вул.Перемоги,39	400,00 грн.	B88888	B44444	Даниленко
H55555	вул.Європейська,56	320,00 грн.	B11111	B55555	Янко
H66666	вул.Празська,21	510,00 грн.	B77777	B66666	Уваров
H77777	вул.Докучаєва,39	450,00 грн.	B66666	B77777	Донченко
H88888	вул.Антонова,56	360,00 грн.	B55555	B88888	Лопаков
				B99999	Громов

Відношення **Власність\_в\_оренду** і **Власник** знаходяться в третій нормальній формі, оскільки в них немає ніяких транзитивних залежностей від первинного ключа. У результаті виконання нормалізації ми отримали чотири окремих відношення, кожне з яких знаходиться в третій нормальній

формі.

Атрибут **O\_Номер** є первинним ключем відношення **Орендар** і додатково, зовнішнім ключем відношення **Облік\_оренди**. Зверніть увагу на те, що у відношенні **Облік\_оренди** атрибут **Власник** виконує функції як зовнішнього, так і частини первинного ключа. Аналогічним чином, атрибут **Н\_Номер** є первинним ключем відношення **Власність\_в\_оренду** і, додатково, у відношенні **Облік\_оренди** виконує функції як зовнішнього ключа, так і частини первинного ключа.

Інакше кажучи, процес нормалізації полягає в декомпозиції вихідного відношення **Оренда** за допомогою послідовного виконання декількох операцій проекції реляційної алгебри. Отримані в результаті декомпозиції відношення забезпечують виконання їхнього з'єднання без втрат.

#### **Питання для контролю:**

1. Як Ви розумієте аномалію вставки і чому її треба уникати при проектуванні бази даних?
2. На що вказує функціональна залежність у відношенні?
3. Що таке детермінант?
4. З якою метою виконується нормалізація відношень?