

## УДК 378.147



**М.С. Сафонов,**  
к.т.н., доцент Одеського  
національного  
політехнічного  
університету  
e-mail:  
nemko85@gmail.com



**Т.Є. Багмет**  
викладач, Херсонський  
політехнічний коледж  
Одеського національного  
політехнічного  
університету  
e-mail:  
tbagmet@gmail.com

## МЕТОДИ ДЕНОРМАЛІЗАЦІЇ БАЗ ДАНИХ ДЛЯ ЗМЕНШЕННЯ ЧАСУ ВИКОНАННЯ ЗАПИТІВ

*М.С. Сафонов, Т.Є. Багмет.*  
*Методы денормализации баз данных для уменьшения времени выполнения запросов.* Проведен експеримент с пониженіем уровня нормализации баз данных разных архитектур. Представлены методы денормализации для уменьшения времени выполнения запросов.

*M.S. Safonov, T.E. Bagmet. Methods of a denormalization of databases for reduction of time of performance of inquiries.* Experiment with lowering of the level of normalization of databases of different architecture is made. Denormalization methods for reduction of time of performance of inquiries are presented.

**Вступ.** Нормальна форма зберігання інформації в базах даних (БД) припускає уникання дублювання даних. Ключових правил два:

- Атомарність означає, що всі сутності зберігаються в неподільному виді. Наприклад, якщо ми зберігаємо адресу, то він швидше за все буде поділений на назву міста, країни й вулицю. Всі вони повинні бути представлені окремими таблицями.

- Унікальність вимагає, щоб кожна сутність була визначена тільки один раз.

Існують загальноприйняті правила проектування БД. Трактуються вони чотирма нормальними формами.

**Визначення проблеми та постановка задачі дослідження.** Аналіз літератури в області проектування БД із 2003 по 2015 показав, що нічого практично не змінилося. Якщо БД не відповідає хоча б трьом нормальним формам, це приводить до:

- неощадливому використанню дискового простору через наявність даних, що дублюються;

- проблеми, пов'язані з оновленням даних через порушення фундаментального принципу «Кожний факт - в одному місці»;

- пов'язані з ними проблеми посиленого контролю цілісності даних, що піддається серйозним випробуванням при наявності даних, що дублюються.

Отже з літератури можна зробити висновок, що для БД, що динамічно оновлюються, нормалізація є необхідністю, що дозволяє уникнути багатьох проблем у процесі функціонування інформаційної системи (ІС).

**Аналіз виконання денормалізації.** У запитах до повністю нормалізованої бази часто доводиться з'єднувати до десятка, а те й більше, таблиць. А кожне з'єднання — операція досить ресурсномістка. Як наслідок, такі запити потребують більше ресурсів сервера й виконуються повільно[1].

У такій ситуації може допомогти:

- Денормалізація шляхом скорочення кількості таблиць. Краще поєднати в одну кілька таблиць, що мають невеликий розмір, що містять рідко змінювану (умовно-постійну або нормативно-довідкову) інформацію, причому інформацію, за змістом тісно зв'язану між собою. У загальному випадку, якщо у великій кількості запитів потрібно об'єднувати більше п'яти або шести таблиць, варто розглянути варіант денормалізації бази даних.

- Денормалізація шляхом введення додаткового поля в одну з таблиць. При цьому з'являється надлишковість даних, потрібні додаткові дії для збереження цілісності БД.

Був проведений експеримент. Створена БД із структурою, представленою на рис. 1.

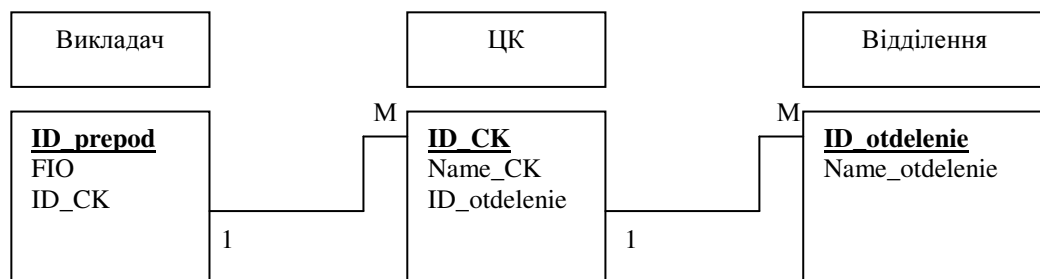


Рис. 1. Структура нормалізованої БД

Далі було згенеровано за допомогою спеціальної програми 1 млн записів у дану БД в таблицю «Викладач». І виконано запит на вибірку всіх викладачів конкретного відділення:

```

SELECT
  prepod.FIO, ck.name_ck, otdelenie.name_otd
FROM (otdelenie INNER JOIN ck ON otdelenie.id_otd = ck.id_otd)
INNER JOIN prepod ON ck.id_ck = prepod.id_ck
WHERE otdelenie.name_otd='Програмна інженерія';

```

Експеримент проводився для трьох СКБД: BDE Paradox, Interbase 7, MySQL 5.

БД відповідає нормальній формі. Структура має потрібну вкладеність.

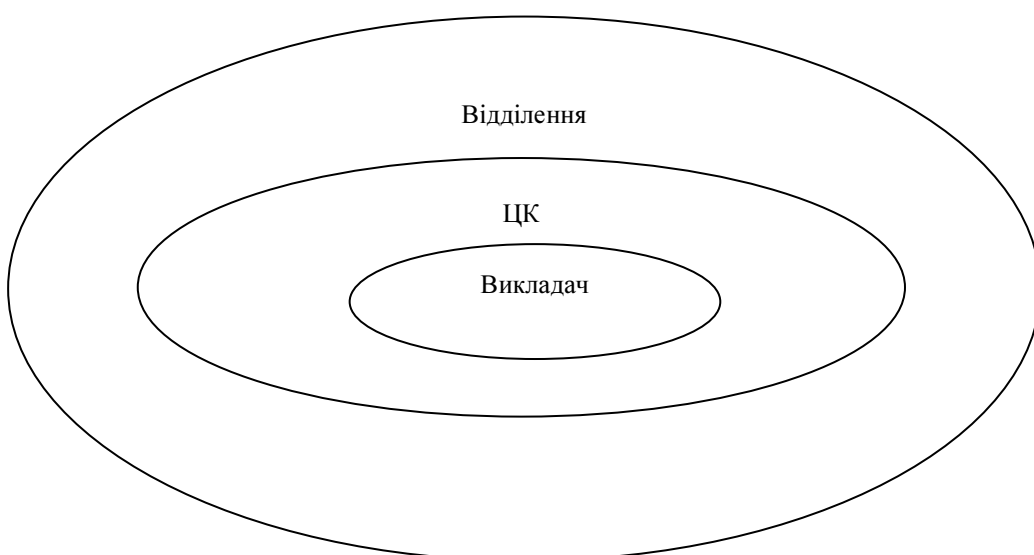


Рис. 2 – Потрійна вкладеність нормалізованої БД

Таблиця 1

**Результати виконання запитів в нормалізованій БД із  
потрійною вкладеністю**

1. БД	1. Розмір (Мбайт)	1. Час виконання запити (сек)
1. BDE Paradox	2. 581	2. 423
1. Interbase 7	3. 768	3. 319
1. MySQL 5	4. 813	4. 198

У рамках експерименту було запропоновано денормалізувати БД із четвертої до другої нормальної форми (рис. 3).

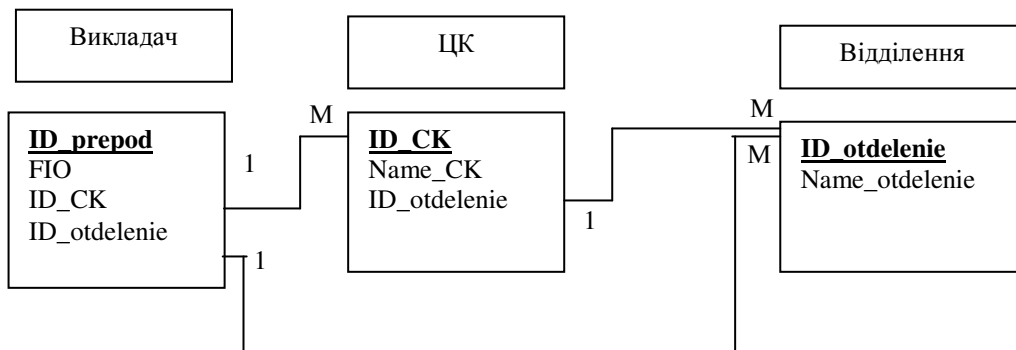


Рис. 3. – Структура денормалізованої БД до другої нормальної форми

Дана БД не відповідає навіть третій нормальній формі. Структура має подвійну вкладеність.

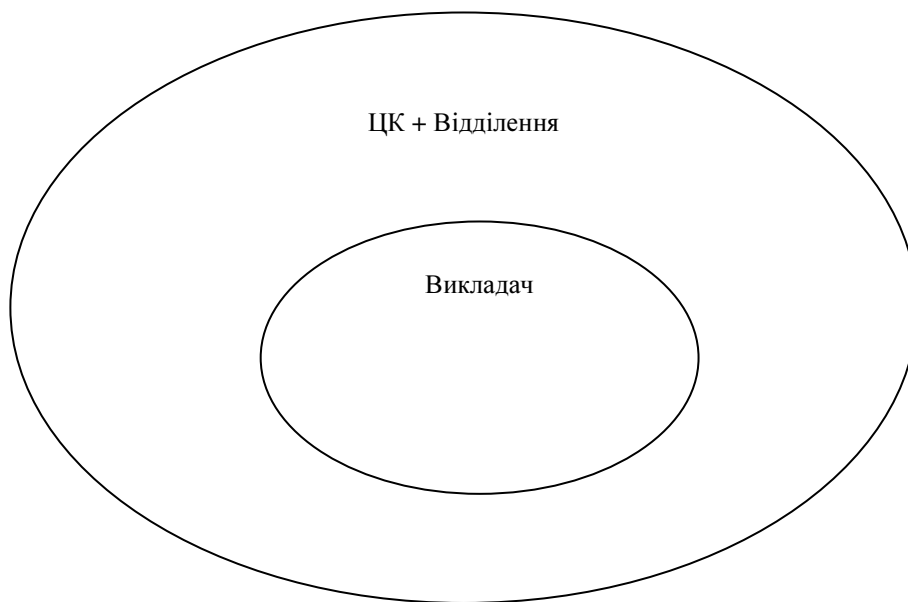


Рис. 4 – Подвійна вкладеність денормалізованої БД

Далі був виконаний запит на вибірку всіх викладачів конкретного відділення:

```

SELECT
FROM otdelenie INNER JOIN prepod ON otdelenie.id_otd =
prepod.kod_otd
WHERE otdelenie.name_otd='Програмна інженерія';

```

Експеримент проводився для трьох СУБД: BDE Paradox, Interbase 7, MySQL 5.

Таблиця 2

### Результати виконання запитів в денормалізованій БД із подвійною вкладеністю

1. БД	1. Розмір (Мбайт)	1. Час виконання запити (сек)
1. BDE Paradox	2. 594	2. 169
1. Interbase 7	3. 779	3. 133
2. MySQL 5	4. 821	4. 87

Найчастіше повільно виконуються й споживають багато ресурсів запити, у яких виконуються складні обчислення, особливо при використанні групувань і агрегатних функцій (Sum, Max, тощо). Іноді має сенс додати в таблицю 1-2 додаткових стовпця, які містять дані, що часто використовуються [2].

Експеримент показав, що час виконання запити не залежить від кількості записів, що повертаються. А залежить від кількості оброблюваних записів.

У зв'язку із цим було запропоноване наступне.

Представлена наступна БД (рис. 5):

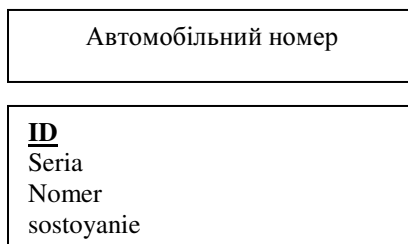


Рис. 5. Незв'язана таблиця в БД

У ній може бути 20 млн записів. Було запропоновано розділити дану таблицю на декілька зв'язаних таблиц, як представлено на рис. 6.

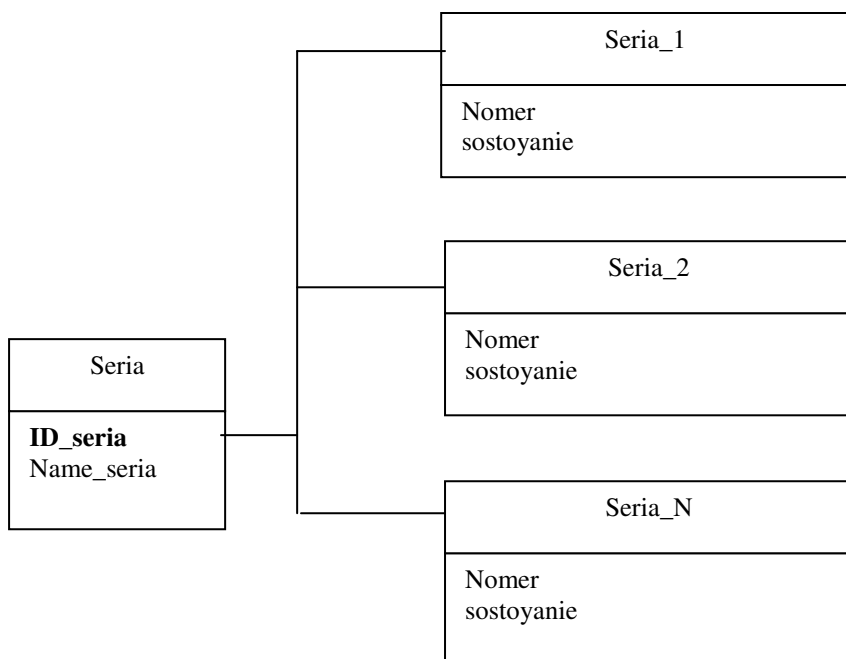


Рис. 6. Групування даних по різним таблицям у відповідності до визначеного признаку

Згідно рис. 6 зв'язок організується з назвою таблиці, а не з її ключем. Порівняльні результати експерименту представлені на рис.7 та рис. 8.

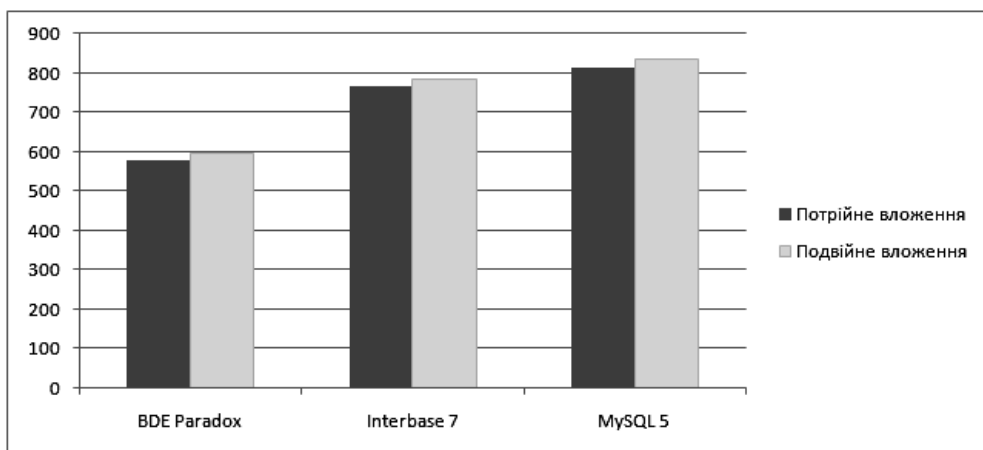


Рис. 7. Розмір БД до та після денормалізації в МБайтах

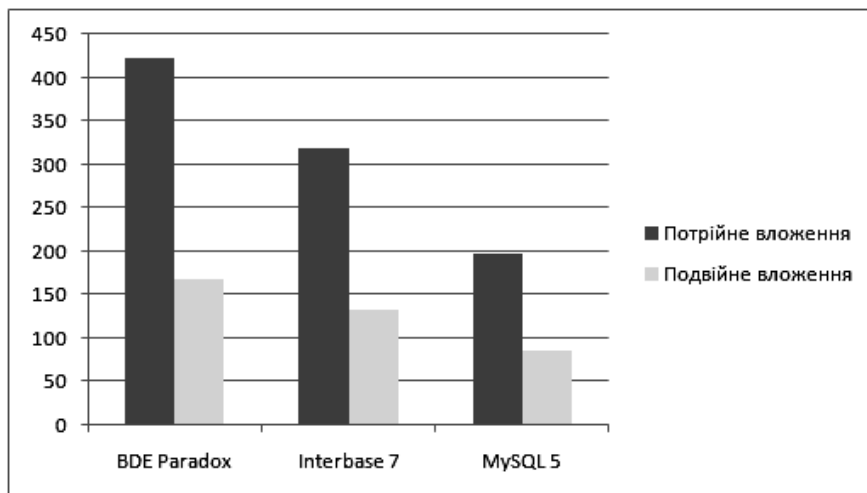


Рис. 8. Час виконання запитів в БД до та після денормалізації в секундах

**Висновки.** З'єднання таблиць - операція, що вимагає певних витрат ресурсів - пам'яті, процесорного часу. Чим вище ступінь нормалізованості БД, тим більше в ній таблиць, отже, тим повільніше виконуються запити на вибірку.

У денормалізованій БД підвищується надмірність даних, що може підвищити продуктивність, але потребує більше зусиль для контролю повнотілісності. Ускладниться процес створення додатків, оскільки дані будуть повторюватися і їх сутужніше буде відслідковувати.

У нормалізованій БД процес контролю цілісності й зв'язків лежить повністю на БД. Денормалізована БД виконує лише функцію зберігання даних. А процес контролю цілісності й зв'язків виконує додаток.

На жаль, не існує ніяких формалізованих методів для досягнення задовільного компромісу між нормалізацією й денормалізацією.

Існує золота середина між нормалізацією й денормалізацією, але щоб знайти її, потрібні чималі зусилля та знання предметної області.

### Література

1. Темушева, Ю., Денормализация БД. Зачем? Когда? Как? [Електронный ресурс] / Режим доступа: <http://habrahabr.ru/post/64524/> (30.11.2015)

2. Кузнецов, С.А. К свободе от проблемы Больших Данных [Текст] / С.А. Кузнецов // Журнал компьютерных технологий, Открытые системы СУБД, - 2012.- № 02 . - С.22 - ISSN 1028-7493.

*Надійшла до редакції 1.12.2015*