

ЕФЕКТИВНІСТЬ ВИКОРИСТАННЯ ДЕНОРМАЛІЗАЦІЇ БАЗ ДАНИХ

В роботі розглядається питання швидкодії використання баз даних. На швидкість в певному степені впливає нормалізація. З метою прискорення виконання запитів здійснений порівняльний аналіз швидкості виконання запитів в нормалізованій і денормалізованій базі даних. Визначені фактори, які впливають на тривалість виконання запитів і ступінь їх впливу. В дослідженнях розглянута база даних, в якій в нормалізованому вигляді присутні дві таблиці, що дає можливість виокремити вплив різних факторів. Такими факторами є наявність зв'язків, що передбачають здійснювати звертання з однієї таблиці до іншої та здійснення пошуку в двох таблицях. Отримано, що нормалізація досить суттєво впливає на тривалість виконання запитів у відносному порівнянні. Поставлені задачі подальших досліджень.

Ключові слова: база даних, система керування базами даних, нормалізація, денормалізація, запити на вибірку, групування, швидкість виконання.

V. L. MARCHENKO, D. I. STOPCHAK,

Khmelnytsky National University

EFFICIENCY OF DATABASE NORMALIZATION USAGE

This paper examines the performance of using databases. Normalization in some amount affect the speed. The rate in some amount affect normalization. In order to accelerate query we have done comparative analysis of speed querying in denormalized and normalized databases. The factors affecting the duration of the queries and the amount of their influence. In studies we've examined a database in which there are two tables in normalized form which enables identification of various factors of influence. These factors are evidence of relations that providing appeals from one table to another and searching two tables. We've examined that normalization gives fairly significant impact on the duration of the queries in relative comparison. The task of further research given.

Keywords: database, database management system, normalization, denormalization, sample requests, grouping, speed of execution.

Вступ

Зазвичай вважається загальноприйнятим виконання нормалізації баз даних, що надає значні переваги. Насамперед це полегшує адміністрування та зменшує необхідну пам'ять для зберігання даних. Однак, останнім часом з'являється все більше досліджень, в яких вказується на недоцільність, в певних випадках, детальної нормалізації [1–5]. Обмеження кількості таблиць в схемі бази даних всупереч прийнятим нормам отримала назву денормалізація. Це стосується, головним чином, тих випадків, коли суттєво значення набуває швидкість роботи системи управління базами даних.

Попередні дослідження

Відомі дослідження несуть більш якісні рекомендації і не дають кількісної оцінки використання денормалізації. В даній роботі вказані випадки її доцільності. Зазначається, що звернення між таблицями є операціями, що потребують додаткового часу. Наприклад, в роботах [1, 3–5] вказується в більшості на технологію і варіанти використання денормалізації і зазначається прискорення використання запитів в випадку її впровадження. В роботі [2] зазначається, що наявність зв'язків між таблицями суттєво уповільнюють роботу з базою даних, але кількісної оцінки впливу наявності зв'язків не наведено. Теж саме можна сказати про рекомендацію використання денормалізації в базах даних з графічними полями. В роботі [2] зазначається про суттєвий вплив на тривалість виконання запитів з використанням групування і агрегатних функцій, але немає оцінки ступеня їх впливу. Питання доцільності її використання вимагають більш уважного вивчення і розглянути в даній роботі. Звісно, автори не ставлять під сумнів доцільність нормалізації, але намагаються визначити критерії встановлення ступеня нормалізації.

Постановка задачі

Метою даної роботи є визначення впливу нормалізації на тривалість роботи бази даних і, як результат цієї оцінки, ефективності роботи бази даних. Ефективність визначається витратами часу на виконання запитів різного типу. Витрати часу необхідно представити у вигляді залежностей від об'єму вибірки і кількості записів в таблицях.

Задача поставлена наступним чином – дослідити, які параметри нормалізації і яким чином впливають на швидкість роботи та порівняти отримані результати з денормалізованою базою даних.

Основна частина

Дослідження виконувалися в реляційній системі керування базами даних MySQL. Об'єктом досліджень прийнята створена база даних, яка використовувалась в двох варіантах – у вигляді двох та однієї таблиці (нормалізованої і ненормалізованої бази даних відповідно). Використання більш складної схеми даних вважалось недоцільним на даному етапі досліджень. Це завадить виокремити вплив певних параметрів на швидкість роботи з базою. Такими параметрами є виконання звертань між таблицями та необхідність здійснення пошуку в двох таблицях нормалізованої бази даних. Основна увага приділялась створенню запитів на вибірку в різних варіантах реалізації.

Для аналізу витрат часу на здійснення звертань між двома таблицями виконано розрахунки

швидкості виконання запиту на вибірку в нормалізованій базі даних, в якій одна з таблиць (головна) має один запис. Кількість записів в іншій таблиці (підпорядкованій) приймала значення 50000. В запит включено поля з двох таблиць, а критерій вибірки заданий для поля підпорядкованої таблиці. Об'єм вибірки змінювався в межах 0 – 50000 записів. Відповідні обчислення виконані для денормалізованої бази даних.

Результати в графічному вигляді наведено на рис. 1. Зважаючи на те, що час виконання запиту є незначним і на отримані результати може впливати виконання інших процесів операційної системи кожна точка на графіках є середнім з тисячі проведених аналогічних обчислень. На рисунку експериментальні значення апроксимовані логарифмічною залежністю. Суцільною лінією показані витрати для нормалізованої бази даних, а пунктиром – денормалізованої.

Аналізуючи наведені результати можна зазначити, що різниця в отриманих результатах для нормалізованої і денормалізованої баз даних збільшується із збільшенням об'єму вибірки і досягає максимуму при кількості записів, що дорівнює 10000. Подальше зростання об'єму вибірки практично не впливає на час виконання запиту. Це може бути пояснено алгоритмами пошуку, які входять в програмне забезпечення системи керування базами даних. З порівняння наведених результатів видно, що в нормалізованій базі швидкість виконання запиту на 30% більший, ніж у відповідній ненормалізованій. Очевидно, що збільшення тривалості виконання запиту обумовлено звертанням до головної таблиці при виконанні запиту. Необхідно звернути увагу, що на виконання пошуку в головній таблиці витрат

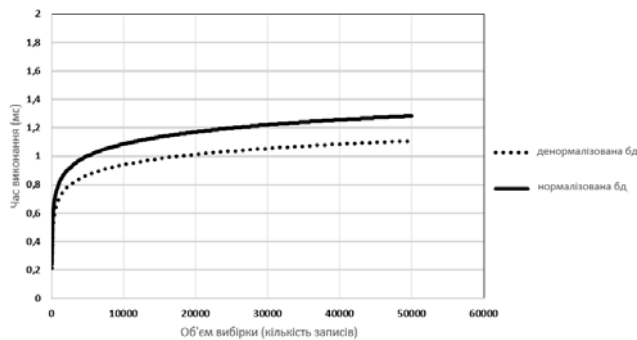


Рис. 1. Залежність тривалості використання запиту від об'єму вибірки (1 таблиця – 50 000 записів, 2 таблиця 1 запис)

збільшенням об'єму вибірки і при виборці 10000 записів досягає трьох мілісекунд, тобто втричі менша. Результати досліджень наведено на рис.2. Де, як і в попередньому випадку, апроксимація експериментальних даних є логарифмічною. Суцільною лінією показані витрати для нормалізованої бази даних, а пунктиром – денормалізованої бази даних.

Подальше зростання об'єму вибірки, як і в попередньому випадку, практично не впливає на результати.

В наведених на рис. 2 графічних залежностях кількість записів головної таблиці становила 5000. При меншій кількості записів головної таблиці різниця тривалості виконання запитів зменшується. Але навіть при незначній кількості записів головної таблиці (в декілька десятків) залишається суттєвою. Інтенсивність зростання різниці швидкостей виконання запиту в денормалізованій і нормалізованій базі даних показана на рис. 3. Наведено залежність різниці термінів виконання запиту з об'ємом вибірки 10000 записів в денормалізованій базі даних з 50000 записами і в нормалізованій, де кількість записів головної таблиці змінюється в межах від 0 до 50000.

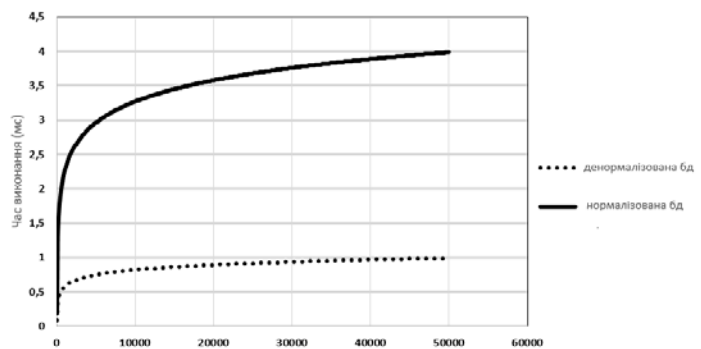


Рис. 2. Залежність тривалості використання запиту від об'єму вибірки (1 таблиця – 50 000 записів, 2 таблиця 5000 записів)

Необхідно звернути увагу на те, що критерій пошуку задавався для поля підпорядкованої таблиці. Тоді при виконанні вибірки пошук здійснювався двічі – в полі підпорядкованої таблиці за заданим критерієм і в головній таблиці в полі ідентифікатора при кожному звертанні. Запит може виконуватися в іншому порядку. Критерій може бути заданим для поля головної таблиці (в нашому випадку ця таблиця менша), а при кожному звертанні до підпорядкованої таблиці в ході виконання запиту пошук в підпорядкованій таблиці буде здійснюватись по полю ідентифікатора. Отже в першому і в другому випадку в нормалізованій базі даних пошук буде здійснюватись двічі. Але в першому випадку (критерій задається в підпорядкованій таблиці) в більшому текстовому полі і меншому полі ідентифікатора. В другому випадку (критерій задається в текстовому полі головної таблиці) в меншому текстовому полі і в більшому полі ідентифікатора. Вказана комбінація полів різних таблиць при виконанні запиту на вибірку може призвести до зниження ефективності денормалізації, якщо типи полів і розмір даних буде значно відрізнятися, але в

проведених в роботі досліджень отримано несуттєвий вплив типу поля на тривалість виконання запиту на вибірку. При здійсненні обчислень були розглянуті поля з цілим типом даних та текстовим, в якому кількість символів було задано 100 і 200. Але при більшій кількості символів ефект може бути протилежним. Тим більше, якщо тип поля буде графічним, а критерій пошуку заданим для цього поля.

Особливістю розглянутих прикладів є те, що результат містить велику кількість даних, пропорційних об'єму вибірки. Доступ до даних є послідовним і не дозволяє безпосередньо звернутися до необхідного елемента. Це призводить до необхідності перебору всіх елементів, які йому передують, що вимагає додаткових витрат часу. В деяких випадках, важливих при

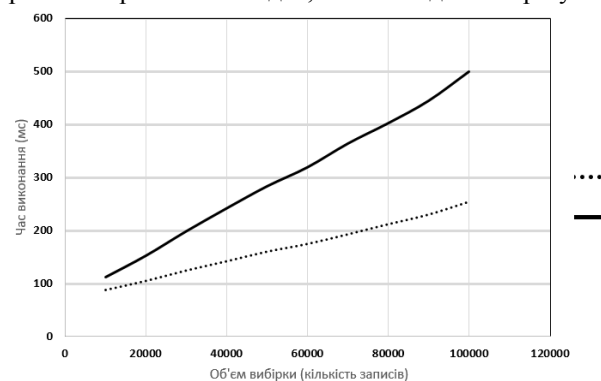


Рис. 4. Залежність використання агрегатної функції (сумування) від кількості записів, що відповідають заданому текстовому критерію.

відповідного йому значення в залежній таблиці). Тобто по осі x представлено кількість записів, які відповідають критерію і були використані для групової функції. Як видно з результатів, наведених на рис. 4 тривалість виконання запитів значно вища, ніж в попередніх випадках. Відмінним є і характер цієї залежності. В нормалізованій і денормалізованій базі даних залежність тривалості виконання запитів від об'єму вибірки є лінійним. Співвідношення між термінами виконання змінюється і становить 26,83% при величині вибірки 10000 записів та 96,79% при виборці 100000 записів.

Висновки

Отримані результати свідчать про те, що нормалізація бази даних вимагає певного зниження швидкодії. При одноразовому виконанні запитів в базі даних з двома таблицями це зниження незначне, але буде збільшуватися із збільшенням кількості таблиць. Зниження швидкості буде зростати при збільшенні кількості звертань до бази даних. На швидкодію впливає тип поля і розмір даних, що якісно змінює ефект нормалізації. При цьому в певних межах, які вказані в роботі і часто використовуються практично, цей вплив незначний. Отже перераховані висновки вказують тільки на те, що питання доцільності застосування надмірної нормалізації неоднозначні і вимагають додаткових досліджень.

Література

1. Серверы корпоративных баз данных [Електронний ресурс]. – Режим доступу :
2. <http://bourabai.kz/dbt/servers/index.htm>
3. Денормализация БД. Зачем? Когда? Как? [Електронний ресурс]. – Режим доступу :
4. <https://habrahabr.ru/post/64524/>
5. The Denormalization Survival Guide – Part I [Електронний ресурс]. – Режим доступу : <http://tdan.com/the-denormalization-survival-guide-part-i/4988>
6. Гандерлой М. Освоение Microsoft SQL Server 2005 / М. Гандерлой, Д. Джорден, Д. Чанц. – М. : Диалектика, 2007. – С. 1103.
7. Databases: Normalization or Denormalization. Which is the better technique? [Електронний ресурс]. – Режим доступу : <http://www.ovaistariq.net/199/databases-normalization-or-denormalization-which-is-the-better-technique/>

Рецензія/Peer review : 3.3.2016 р. Надрукована/Printed : 19.4.2016 р.
Стаття прорецензована редакційною колегією