

ВИКОРИСТАННЯ ДЕНОРМАЛІЗАЦІЇ РЕЛЯЦІЙНОЇ БАЗИ ДАНИХ З МЕТОЮ ПІДВИЩЕННЯ ШВИДКОДІЇ ЗАПИТІВ У СИСТЕМАХ З ВЕЛИКИМ ОБСЯГОМ ДАНИХ

Кравченко Є. О.

Науковий керівник – к.т.н., доц. Мазурова О. О.

Харківський національний університет радіоелектроніки, каф. ПІ
м. Харків, Україна

e-mail: yevhen.kravchenko@nure.ua

Modern database management systems (DBMS) incorporate advanced caching, indexing and data storage optimisation technologies. However, as data volumes increase, the efficiency of query processing in relational databases decreases. While vertical scaling may be limited due to high costs, and horizontal scaling carries risks for operational projects, properly implemented denormalisation offers an effective solution for resource optimisation and performance improvement. This paper presents the results of a planned experimental study of denormalisation methods such as vertical table partitioning and summary tables.

Останнім часом NoSQL БД здобувають все більшу популярність [1], але у багатьох галузях необхідні реляційні бази з повною підтримкою вимог ACID. Сучасні системи керування базами даних (СКБД) інтегрують розширені технології кешування, індексації та оптимізації збереження даних. Проте, зі збільшенням обсягу даних, ефективність обробки запитів у реляційних базах даних знижується. У той час, як вертикальне масштабування може бути обмеженим через високі витрати, а горизонтальне розширення несе ризики для діючих проєктів, адекватно реалізована денормалізація пропонує ефективне рішення для оптимізації ресурсів і підвищення продуктивності.

Денормалізація давно використовується на практиці як засіб оптимізації структури БД задля покращення швидкодії запитів [2] чи вирішення специфічних задач [3], але недостатньо висвітлена на академічному рівні. Зазвичай роботи щодо архітектури БД не рекомендують використовувати навмисну денормалізацію доки не вичерпано всі можливості оптимізації в рамках класичної нормалізованої парадигми та, відповідно, не дають оцінку потенціалу ефективності.

Використання методів денормалізації у дизайні БД можливе з умов розуміння логіки функціонування ПЗ та зазвичай вимагає створення додаткових програмних рішень на стороні серверу БД, що окрім покращення швидкості обробки запитів може привести до погіршення виконання транзакцій в БД.

Отже, для отримання реальних рекомендацій щодо вибору ефективних підходів до денормалізації була поставлена задача дослідити методи денормалізації реляційних баз даних, а саме:

- проаналізувати існуючі методи денормалізації реляційних баз даних та обрати найбільш перспективні для подальшого дослідження;
- провести планування експерименту, а саме обрати прикладну предметну область для дослідження, спроектувати базу даних, обрати метрики для оцінки результатів;
- обрати СКДБ у якості середовища для проведення дослідження;
- розробити нормалізовану до 3НФ та денормалізовану БД на основі обраних методів денормалізації;
- спроектувати та реалізувати програмні рішення для підтримки ненормалізованої БД;
- провести експерименти, проаналізувати отримані результати та сформулювати рекомендації щодо використання методів денормалізації.

У якості СКДБ для проведення дослідження було обрано MS SQL Server 2016. За результатами проведеного аналізу було обрано методи фізичної та логічної денормалізації для подальшого дослідження, а саме метод вертикального розбиття та метод зведених таблиць.

Фізичний метод вертикального розбиття підходить для оптимізації таблиць, що добре сегментуються, або значна частина даних у яких може бути визначена, як архівна. Такі дані не підлягають регулярним змінам та можуть бути виключені з операцій модифікації. Велика перевага метода полягає у збереженні сумісності з похідною таблицею, бо значення ключових полів не змінюються.

Метод зведених таблиць головним чином використовується у побудові аналітичних звітів, коли є можливість заздалегідь підготувати значення обчислюваних метрик, виконати операції поєднання та зберегти результати до таблиці фактів, що оптимізована для операцій читання та є основою для побудови звітів. Таким чином найбільш повільні операції виконуються на попередньому етапі, де вимоги до швидкодії більш м'які.

Було проведено планування експериментальних досліджень, а саме:

- в якості предметної області, на якій буде проводитися дослідження, обрано область діяльності великої мережі лікарень щодо операційної діяльності, в якій ведеться облік та аналітична обробка карток пацієнтів, проведених обстежень, встановлених діагнозів, виконаних процедур, стану фінансових розрахунків зі страховими компаніями, тощо;
- розроблена БД, яка включає в себе такі сутності, як фінансовий баланс за пацієнтами у розрізі часу, реєстр виконаних послуг, що чекають на оплату, реєстр платіжних вимог, реєстр документів щодо підтвердження чи відмови оплати згідно відповідності процедур діагнозам, повноті заповнення супровідних документів, відповідності виставлених рахунків ринковим показникам, тощо;

– в якості метрик для проведення замірів обрано час виконання запиту, завантаження процесору, обсяг використаної оперативної пам'яті, обсяг використаного дискового простору, кількість операцій фізичного читання, ефективність використання кешу.

– заплановано умови проведення експериментів, а саме змодельовано типовий сценарій роботи системи, що включає імпорт та обробку нових даних, збереження оброблених даних у таблиці фактів, побудова плоских звітів з клієнтського веб-інтерфейсу, побудова кубів для зведених звітів з боку OLAP модулю тощо.

На основі найбільш перспективних у контексті предметної області методів денормалізації розроблено структуру БД, яка дозволить дослідити теоретичні переваги денормалізованої структури під час операцій читання та вивчити практичний вплив денормалізації на операції запису та модифікації.

Розроблено програмні рішення для проведення експериментів, а саме створено однаковий за функціями ETL-процес для завантаження даних в фінальні таблиці обох архітектур, зовнішні інтерфейси, що дозволяють робити уніфіковані запити, та середа для генерування одночасних запитів.

Під час проведення експериментів очікується отримати результати, які дозволять оцінити практичний вплив методів денормалізації на швидкодію обраних запитів та навантаження на сервер БД.

Отримані в результаті рекомендації будуть корисні для інженерів, що розробляють нові програмні продукти на базі реляційних БД, або займаються підтримкою систем, що мають проблеми зі швидкодією. Рекомендації щодо ефективності використання досліджених методів дозволять розробникам БД спростити та прискорити такі процеси, як оптимізація реляційних БД задля підвищення їх продуктивності в цілому чи досягнення потрібних значень за певними метриками використання.

Список використаних джерел:

1. Kuzochkina, A., Shirokopetleva, M., Dudar, Z. (2019). Analyzing and Comparison of NoSQL DBMS. 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 Proceedings (pp. 560–564). DOI 10.1109/INFOCOMMST.2018.8632133.

2. Bock, D. B., Schrage, J.F. (2002). Denormalization guidelines for base and transaction tables. ACM SIGCSE Bulletin, Volume 34, Issue 4 (pp.129–133). <https://dl.acm.org/doi/10.1145/820127.820184>.

3. Rathika, V. (2019). Graph-Based Denormalization for Migrating Big Data from SQL Database to NoSQL Database. ICICV 2019: Intelligent Communication Technologies and Virtual Mobile Networks (pp. 546–556). https://doi.org/10.1007/978-3-030-28364-3_56.