

О.Б. Кунгурцев, к.т.н., проф.
С.Л. Зіноватна, інж.

Одеський національний політехнічний університет

ПОРІВНЯННЯ ВАРТОСТІ ВИКОНАННЯ ЗАПИТІВ ДО Й ПІСЛЯ ДЕНОРМАЛІЗАЦІЇ РЕЛЯЦІЙНОЇ БАЗИ ДАНИХ

Розроблено метод аналізу інформаційної системи з метою виявлення можливостей виконання денормалізації відношень бази даних за рахунок дослідження журналу запитів до системи. Розроблено алгоритм скорочення множини запитів, яка аналізується. Наведено результати експериментальних досліджень роботи алгоритму скорочення множини запитів, яка аналізується.

Постановка задачі в загальному вигляді та її актуальність. Одним зі способів підвищення продуктивності інформаційної системи (ІС) є зниження рівня нормалізованості відношень у базі даних (БД) [1], [2]. У випадку ухвалення рішення про денормалізацію проводиться реструктуризація схеми відношень, і, отже, відбувається заміна текстів певного набору запитів до системи. Дослідження всієї множини запитів до системи протягом певного часового періоду дозволяє виявити прийнятні для даної системи варіанти денормалізації з множини можливих. Кожен варіант необхідно оцінити, зрівнявши витрати на виконання запиту до й після реструктуризації. Витрати можуть вимірюватись з погляду використання системних ресурсів, і, зокрема, вартість виконання запиту може оцінюватись за кількістю операцій доступу до диска.

Аналіз існуючих рішень в області дослідження запитів до БД. При обробці запиту система управління базами даних (СУБД) виконує його оптимізацію з метою вибору найбільш ефективного плану виконання. При цьому виконується нормалізація предикатів запиту, перевірка семантичної коректності запиту і його спрощення. Потім СУБД, використовуючи правила перетворення операцій реляційної алгебри (РА), будує множину планів виконання запитів, з яких вибирає найбільш ефективний [2], [3]. Алгоритми роботи оптимізатора конкретної СУБД невідомі її користувачеві, також невідомо, який із планів виконання вибере система. Однак відомі загальні правила оптимізації запитів.

Метою проведених досліджень є одержання формального способу виявлення можливостей денормалізації схеми відношень БД із метою підвищення продуктивності інформаційної системи.

Викладення основного матеріалу досліджень. Для порівняння витрат на виконання запитів у БД до й після денормалізації пропонується використовувати методи оцінки ефективності планів виконання, застосовувані при оптимізації запитів. На відміну від роботи оптимізатора в цьому випадку порівнюються не різні плани виконання одного запиту, а плани виконання запитів, які надають однакову інформацію, але при цьому звертаються до різних схем відношень.

Як вихідні дані для аналізу роботи ІС використовується журнал системи, що розглядається як множина кортежів $L = \{ \langle d_{st}, t_{st}, q, t_q, wp \rangle \}$, де d_{st} – дата виконання запиту, t_{st} – час запуску, q – текст запиту, t_q – тривалість його виконання й wp – ідентифікатор робочого місця, з якого виконується запуск запиту.

Для ІС необхідно визначити період спостереження P . У переважній кількості випадків спостереження за системою протягом року дасть достовірні результати, оскільки це максимальний період, що визначає господарську діяльність організації, погодні умови, періоди відпусток і т. ін. Однак у великій організації за цей період будуть зроблені сотні тисяч і мільйони запитів до системи, обробка яких буде являти певну проблему, тобто потужність множини L може виявитися дуже великою, і, отже, час аналізу також виявиться значним. Як рішення завдання пропонується обробляти не всю множину запитів, а тільки певним чином відібрані окремі запити до системи.

Для реалізації цього пропонується наступний алгоритм.

1. Визначається номер ітерації $h = 1$.

2. Установлюється мінімальна кількість вибірок $N_h = k_{wp} N_a^d$, де k_{wp} – кількість робочих місць, N_a^d – середня кількість запитів до системи за робочий день від одного робочого місця.

3. Виконується вибірка кожного K -го кортежу з L , де $K = \frac{N_P}{N_h}$, а N_P – загальна кількість запитів, зафіксованих за період спостереження P . Виконується заповнення структур даних M_1^h і M_2^h , наведених на рис. 1. Опис вимірюваних величин наведено в таблиці 1.

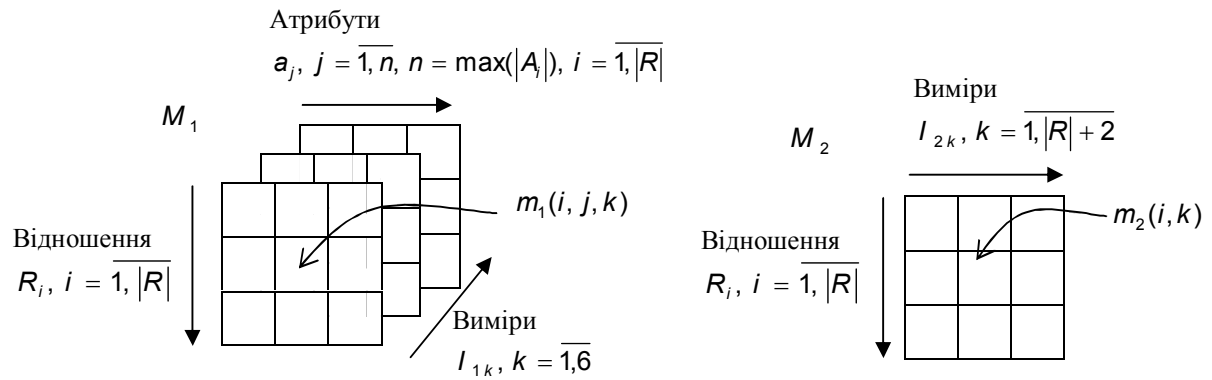


Рис. 1. Структура даних для алгоритму скорочення множини запитів, яка аналізується

Таблиця 1

Опис характеристик, вимірюваних під час роботи алгоритму скорочення множини запитів, яка аналізується

Вимір	Характеристика	Опис
l_{11}	N_{R,a_j}	відносна кількість запитів на вибірку з відношення R_i з операціями проєкції або вибірки, що включають як операнд атрибут a_j
l_{12}	N_{U,R,a_j}	відносна кількість запитів на відновлення атрибута a_j відношення R_i
l_{13}	N_{Gr,R,a_j}	відносна кількість запитів на вибірку з використанням атрибута a_j відношення R_i як атрибута угруповання
l_{14}	t_{R,a_j}	відносний час виконання запитів на вибірку з відношення R_i з операціями проєкції або вибірки, що включають як операнда атрибут a_j
l_{15}	t_{U,R,a_k}	відносний час виконання запитів на відновлення атрибута a_j відношення R_i
l_{16}	t_{Gr,R,a_j}	відносний час виконання запитів на вибірку з використанням атрибута a_j відношення R_i як атрибута угруповання
l_{21}	N_{I,R_i}	відносна кількість додавань кортежів у відношення R_i
l_{22}	N_{D,R_i}	відносна кількість видалень кортежів з відношення R_i
l_{2k}	N_{R_i,R_k}	відносна кількість запитів на вибірку з операцією з'єднання, у якій відношення R_i виступає як батьківське до R_k

4. Виконується аналіз для ітерації $h + 1$, $N_{h+1} = 2N_h$; заповнюються M_1^{h+1} і M_2^{h+1} .

5. Якщо розбіжність у всіх відповідних значеннях пар $(m_1^h(i, j, k), m_1^{h+1}(i, j, k))$ і пар $(m_2^h(i, k), m_2^{h+1}(i, k))$ не більше припустимого значення $|m^h - m^{h+1}| < \varepsilon$, то процес мінімізації L_σ завершується. У протилежному випадку $N_h = 2N_{h+1}$ ($N_h < N_p$), $h = h + 1$, і повторюються кроки 3–4.

Таким чином, для аналізу ІС далі використається множина $L_\sigma \subset L$, отримано в результаті скорочення вихідної множини L описаним вище способом.

Із усієї множини $Q = \{q\}$ різних запитів з L_σ необхідно вибрати такі підмножини Q_i , кожна з яких показує можливість застосування деякого варіанта денормалізації V_i [4], [5]. У результаті буде отримана множина пар $V = \{<Q_i, V_i>\}$.

Наприклад для аналізу можливості застосування висхідної денормалізації (включення в батьківське відношення підсумкових даних, отриманих на основі підлеглих відношень) в Q_i необхідно включити такі запити:

```
SELECT LRgr, f(LR')
FROM JOIN (R')
GROUP BY LRgr,
```

(1)

де R' – підмножина всіх відношень R розглянутої бази даних, які беруть участь у запиті; $f(L_{R'})$ – функція, що обчислює агреговане значення на основі множини атрибутів $L_{R'}$ відношень R' ,

$L_{R'} = L_{R_1} \cup \dots \cup L_{R_i} \cup \dots \cup L_{R_m}$, $|L_{R_i}| \geq 0$, $L_{R_i} \subseteq A_{R_i}$, де A_{R_i} – множина атрибутів відношення R_i ; $L_{R_{gr}}$ – список атрибутів відношення, по якому виконується угруповання $R_{gr} \in R'$; $JOIN(R')$ – результат операцій з'єднання між взаємозалежними відношеннями R' .

Для кожного i -го елемента множини V виконуються наступні дії.

1. Одержати реструктуровану схему відношень R' для V_i , зокрема, для висхідної денормалізації реструктуризація полягає в додаванні атрибута у відношення R_{gr} : $A_{R_{gr}} = A_{R_{gr}} \cup b_{f(L_{R_i})}$.

2. Створити множину Q'_i , кожний j -й елемент якого відповідає j -му запиту з множини Q_i , перетвореному стосовно R' , зокрема запит (1) перетвориться в такий спосіб:

```
SELECT L_{R_{gr}}, b_{f(L_{R_i})}
FROM R_{gr}
```

3. Для кожного $q_{ij} \in Q_i$ й кожного $q'_{ij} \in Q'_i$ побудувати оптимізоване дерево реляційної алгебри (t_{ij} і t'_{ij} відповідно); приклад дерев t_{ij} і t'_{ij} показаний на рис. 2, де σ – операція з'єднання; π – операція вибірки; Π – операція проєкції.



Рис. 2. Приклад БД і дерев реляційної алгебри для запиту до (а) і після (б) денормалізації

При оптимізації дерева реляційної алгебри доцільно використати евристичні правила обробки запитів і правила перетворення операцій реляційної алгебри [2].

Метою даного кроку є одержання можливо мінімальної границі ефективності виконання реструктуризації на наступних кроках.

4. На основі статистичних показників БД оцінити вартість виконання кожної операції реляційної алгебри, представлені в t_{ij} і t'_{ij} ; під час оцінок варто вибирати формули, що відповідають найбільш імовірним стратегіям виконання операції [2], [6]. До статистичних показників належать розміри відношень, розподіл значень в атрибутах, наявність індексних структур для атрибутів і т.ін.

Статистичні показники визначаються для аналізованої БД за станом на сучасний момент або виконується розрахунок цих значень у результаті оцінки імовірнісних характеристик для майбутнього режиму роботи БД.

5. Одержати C_{ij} і C'_{ij} для t_{ij} й t'_{ij} відповідно, де C_{ij} і C'_{ij} – інтегровані оцінки вартості виконання запитів, представлених відповідними деревами.

6. На основі значень C_{ij} і C'_{ij} та витрат на підтримку цілісності даних, яка може бути порушена завдяки введенню надмірності, визначити ефективність виконання реструктуризації для V_i за такими напрямками: для окремого j -го запиту; для окремих робочих місць, з яких надходили запити з множини Q_i ; для системи в цілому.

Ефективність денормалізації для конкретного запиту q_{ij} з використанням варіанта V_i для конкретного запиту q_{ij} оцінюється за формулою:

$$E^{V_i, q_{ij}} = \frac{T - T^{V_i}}{T^{V_i}} \cdot 100 \% \gg 1, \tag{2}$$

де T – час виконання запиту на вибірку q_{ij} до денормалізації;

T^{V_i} – час виконання запиту на вибірку q'_{ij} після денормалізації.

Ефективність для робочого місця im розраховується за формулою:

$$E^{V_i q_{ij} im} = \frac{N_{q_{ij}} (T - T^{V_i}) - T^{U_v}}{N_{q_{ij}} T^{V_i} + T^{U_v}} \cdot 100 \% \gg 1, \quad (3)$$

де $N_{q_{ij}}$ – кількість запусків запиту q'_{ij} від конкретного робочого місця;

T^{U_v} – час на виконання додаткового коду T^{U_v} , який може знадобитися для підтримки цілісності даних після реструктуризації БД.

Для системи в цілому величини $N_{q_{ij}}$ и T^{U_v} розраховуються для всіх робочих місць.

7. Ухвалити рішення щодо доцільності реструктуризації, представлені варіантом V_i .

Перераховані дії є основою алгоритму для автоматизації денормалізації реляційної бази даних. Схема автоматизованої системи представлена на рис. 3.

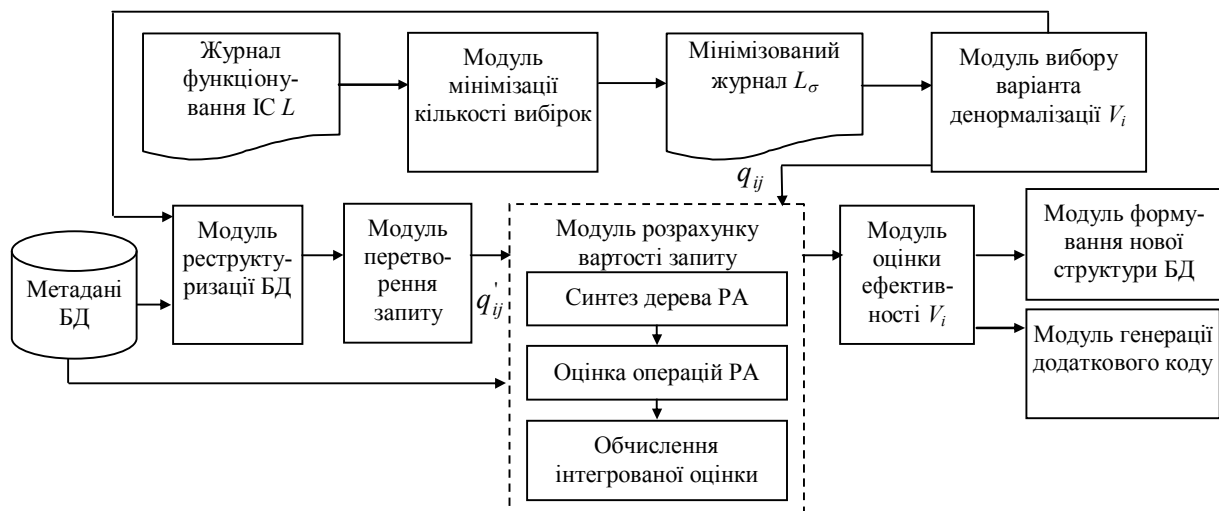


Рис. 3. Схема автоматизованої системи реструктуризації БД

Результати експериментальних досліджень. Робота алгоритму скорочення аналізованої множини запитів була перевірена експериментально в інформаційній системі збору статистичної інформації в лікувально-профілактичній установі. Досліджено журнал з 72322 запитів. Після застосування алгоритму скорочення в журналі залишено 18080 запитів. Час роботи алгоритму скорочення склав 659 мсек. Проведено аналіз на виявлення можливостей застосування різних варіантів денормалізації з використанням вхідного журналу L і скороченого журналу L_{σ} . У результаті множини можливих варіантів реструктуризації БД V_L й $V_{L_{\sigma}}$ збіглися, при цьому час обробки журналу скоротився в 3 рази, з урахуванням часу, витраченого на скорочення L .

Висновки. Результати досліджень показують ефективність застосування формалізованого методу оцінки можливості виконання різних варіантів денормалізації в БД. Результати досліджень також свідчать про можливість застосування алгоритму скорочення множини запитів для аналізу інформаційної системи. Застосування даного алгоритму дозволяє зменшити час дослідження системи.

ЛІТЕРАТУРА:

1. Kris Van Thillo. Using Dynamic View Generation to offset Effects of Performance based Denormalisation. – <http://www.fors.com.ru/eoug97/papers/0039.htm>.
2. Томас Коннолли. Базы данных: Проектирование, реализация, сопровождение. Теория и практика. – М.: Издательский дом «Вильямс», 2002. – 1120 с.
3. Выбор и оценка альтернативных планов выполнения запросов. – http://blackhost.narod.ru/sql/info/optim/art_26_5.htm.
4. Зиноватная С.Л. Выбор отношений-кандидатов для выполнения нисходящей денормализации // Тр. Одесск. политехн. ун-та. – Одесса. – 2006. – С. 68–72.
5. Зиноватная С.Л. Анализ информационной системы для принятия решения о выполнении восходящей денормализации // Холодильная техника и технологии: Сборник научных трудов IV семинара «Информационные системы и технологии». – Одесса, 2006. – С. 14–18.

6. Гарсиа-Молина Г. Ульман Дж., Уидом Дж. Системы баз данных: Полн. курс – М.: Издат. дом “Вильямс”, 2003. – 1088 с.

КУНГУРЦЕВ Олексій Борисович – кандидат технічних наук, професор кафедри системного програмного забезпечення Одеського національного політехнічного університету.

Наукові інтереси:

- методи підвищення продуктивності інформаційних систем з реляційними базами даних;
- методи побудови інтерфейсу користувача інформаційної системи.

ЗІНОВАТНА Світлана Леонідівна – інженер, старший викладач кафедри системного програмного забезпечення Одеського національного політехнічного університету.

Наукові інтереси:

- продуктивність інформаційних систем;
- ефективність структур даних.

Подано 19.10.05

Кунгурцев О.Б., Зіноватна С.Л. Порівняння вартості виконання запитів до й після денормалізації реляційної БД

Кунгурцев А.Б., Зиноватная С.Л. Сравнение стоимости выполнения запросов до и после денормализации реляционной БД

Kungurtsev A.B., Zinovatnaya S.L. Comparison of query execution cost before and after denormalization relation data base

УДК 004.658.3

Порівняння вартості виконання запитів до й після денормалізації реляційної БД / О.Б. Кунгурцев, С.Л. Зіноватна

Розроблено метод аналізу інформаційної системи з метою виявлення можливостей виконання денормалізації відношень БД за рахунок дослідження журналу запитів до системи. Розроблено алгоритм скорочення множини запитів, яка аналізується. Наведено результати експериментальних досліджень роботи алгоритму скорочення множини запитів, яка аналізується.

УДК 004.658.3

Сравнение стоимости выполнения запросов до и после денормализации реляционной БД / А.Б. Кунгурцев, С.Л. Зиноватная

Разработан метод анализа информационной системы с целью выявления возможностей выполнения денормализации отношений БД за счет исследования журнала запросов к системе. Разработан алгоритм сокращения анализируемого множества запросов. Приведены результаты экспериментальных исследований работы алгоритма сокращения анализируемого множества запросов.

УДК 004.658.3

Comparison of query execution cost before and after denormalization relation data base / А.В. Kungurtsev, S.L. Zinovatnaya

Method for information system analysis in order to reveal possibilities of execution of data base relations denormalization investigating log of queries to the system was elaborated. Algorithm of reduction of the analyzed query set was elaborated. The outcomes of experimental researches of the operation of the algorithm of the analyzed query set reduction are given.