

РОЗРАХУНОК КОЕФІЦІЕНТА МАТЕРІАЛІЗАЦІЇ ПРИ ОЦІНЦІ ЗАПИТІВ ДЛЯ ОБСЛУГОВУВАННЯ МАТЕРІАЛІЗОВАНИХ ПРЕДСТАВЛЕНЬ

Новохатська К.А.

Одеський національний політехнічний університет, аспірант

I. Постановка проблеми

Із зростанням обсягів даних, що містяться у сучасних СКБД, знадобилися нові підходи до надання ефективного доступу до інформації. Одним із способів підвищення продуктивності роботи СКБД є використання матеріалізованих представлень (МП). Ідея матеріалізації полягає в оптимізації найбільш ресурсоемних і часто виконуваних запитів шляхом збереження попередньо обчислених результатів на диск.

Вибір запитів-кандидатів на матеріалізацію є однією з важливих задач в області автоматизації МП. В першу чергу її нетривіальність обумовлена обчислювальною складністю. Для пошуку запитів-кандидатів на матеріалізацію необхідно проаналізувати журнал транзакцій за тривалий період часу і вибрати запити, згідно кількома критеріям.

Вперше критерії вибору запитів-кандидатів на матеріалізацію були запропоновані в роботах [1-2]. Для оцінки запиту було введено поняття його вартості, яке визначалося як сумарний час виконання всіх однакових запитів. Даний показник не враховував такі важливі характеристики, як кількість витрачених ресурсів при виконанні запиту, частоту його появи і частоту оновлення базових таблиць (БТ).

У роботах [3-4] було продовжено дослідження даного питання. Була запропонована технологія оцінки ефективності МП, що порівнює продуктивність виконання запитів при використанні МП і без нього в різні періоди часу роботи системи. Дана технологія повністю покриває питання обслуговування МП після його створення, однак вона не акцентує увагу на стадії пошуку запитів-кандидатів на матеріалізацію, коли продуктивність виконання запиту з використанням МП ще не відома.

Нами були сформульовані наступні критерії вибору запиту-кандидата на матеріалізацію:

1. Час виконання запиту прагне до максимуму;
2. Частота виконання запиту прагне до максимуму;
3. Число ресурсів, витрачених при виконанні запиту, прагне до максимуму;
4. Частота оновлення БТ прагне до мінімуму.

Останній критерій обумовлений тим, що кожна модифікація даних в БТ вимагає перерахунку МП. Тобто створення МП для часто оновлюваних даних неефективно.

II. Мета роботи

Метою даної роботи є формалізація критеріїв вибору запитів-кандидатів на матеріалізацію та приведення їх до виду чисельного коефіцієнта, який би дозволив порівнювати запити між собою на етапі аналізу журналу транзакцій СКБД.

III. Розрахунок коефіцієнта матеріалізації

В якості вхідних даних пропонується використовувати множину неунікальних запитів різних типів (*SELECT, INSERT, UPDATE, DELETE*) з журналу транзакцій СКБД із зазначенням часу їх виконання та витрачених ресурсів:

$$Q = \bigcup q \langle \text{text}, \tau, b \rangle,$$

де *text* – текст запиту; τ – час виконання запиту; *b* – сумарне число блоків даних, прочитаних з диска або буфера.

Введемо коефіцієнт матеріалізації *K*, що враховує час виконання запиту, число спожитих ресурсів і частоту оновлення БТ. Опишемо його наступним виразом:

$$K = \frac{\tau c n}{F_{\Delta}}, \quad (1)$$

де *n* – число появ запиту; *c* – вартість виконання запиту. Розраховується як сумарна кількість блоків даних *b*, витрачених за *n*-виконань запиту; F_{Δ} – частота оновлення БТ, що беруть участь в даному запиті.

Виходячи з виразу (1) коефіцієнт матеріалізації K буде прагнути до максимуму для наступних типів запитів:

1. Запити виконуються рідко, але споживають велику кількість ресурсів при обчисленні. До таких запитів відносяться різного виду звіти про роботу підприємства за деякий період (день, тиждень, місяць, квартал).
2. Запити виконуються із середньою частотою і споживають середнє число ресурсів. У цьому випадку оптимізація за допомогою МП найбільш ефективна, оскільки дана група запитів становить основне навантаження СКБД в робочий час підприємства.
3. Запити виконуються дуже часто, але споживають невелику кількість ресурсів. Матеріалізація даної групи запитів менш ефективна, але в цілому дозволить підвищити продуктивність роботи СКБД.

Для наступних типів запитів коефіцієнт матеріалізації K буде прагнути до мінімуму, що дозволить виключити їх з числа кандидатів на матеріалізацію:

1. Запити, розраховані на підставі часто оновлюваних даних. Виняток даної групи дозволить знизити витрати ресурсів на обслуговування МП.
2. Запити, які були виконані один або кілька разів. Оскільки дана група запитів не з'являється в системі циклічно, їх матеріалізація не потрібна.
3. Запити з'являються часто, але виконуються гранично швидко. Їх подальша оптимізація не має сенсу.

Приклад розподілу коефіцієнта матеріалізації K залежно від значень τ , c і $F\Delta$ для журналу транзакцій досліджуваної СКБД представлено на рис. 1.

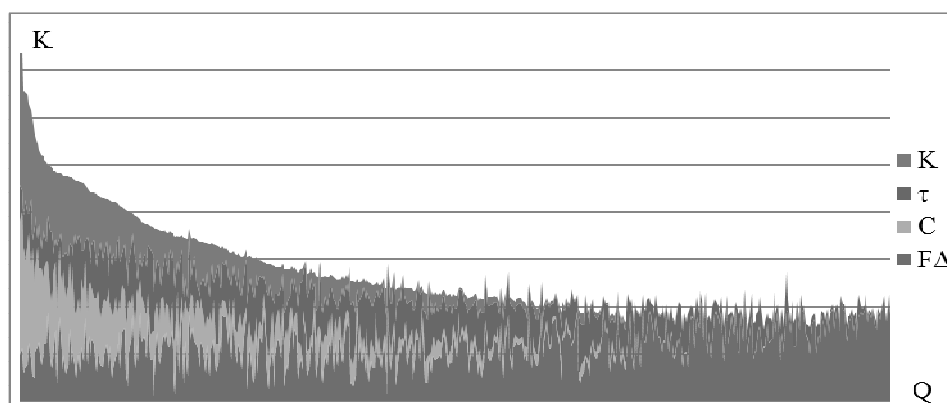


Рисунок 1 - Приклад розподілу коефіцієнта кластеризації

Висновок

У даній роботі був запропонований числовий коефіцієнт оцінки запитів з точки зору можливості створення МП. Отриманий коефіцієнт покрив такі важливі часові та статистичні показники виконання запитів, як обсяг витрачених при формуванні результату ресурсів, частоту поява запитів в ІС, а також частоту оновлення БТ. Останній показник дозволить відсіяти МП, обслуговування яких потребуватиме значного числа обчислювальних ресурсів СКБД.

Список використаних джерел

1. Кунгурцев А.Б. Анализ возможности применения материализованных представлений в информационных системах / А.Б. Кунгурцев, Куок Винь Нгуен Чан // Праці ОПУ. – 2003. – № 2(20). – С. 102-106.
2. Кунгурцев А.Б. Сравнение запросов в реляционных базах данных для построения материализованных представлений / А.Б. Кунгурцев, Куок Винь Нгуен Чан, Блажко А.А. // Праці УНДІРТ. – Одеса. – 2004. – № (39). – С. 35-38.
3. Кунгурцев А.Б. Поддержка эффективности механизма управления материализованными представлениями / А.Б. Кунгурцев, Ю.Н. Возовиков // Електротехнічні та комп'ютерні системи – 2011. – № 4. – С. 136-140.
4. Кунгурцев А.Б. Технология создания материализованных представлений для реляционных баз данных / А.Б. Кунгурцев, Ю.Н. Возовиков // Праці ОПУ – 2012. – № 2. – С. 170-176.