

ПОДАННЯ ФОРМУВАННЯ МНОЖИНИ ПОЛІВ-КАНДИДАТІВ НА СТВОРЕННЯ КЛАСТЕРНОГО ІНДЕКСУ

Манчук А.О.

**Науковий керівник – ст. викл. каф. «Системне програмне забезпечення»,
канд. техн. наук Зіноватна С.Л.**

В основі більшості інформаційних систем (ІС) лежать бази або сховища даних. В активно використовуваній ІС із зростанням обсягу даних в базі даних (БД) збільшується час виконання запитів. У результаті може виникнути ситуація, коли показники продуктивності системи перестануть задовольняти користувачів. Одним з ефективних способів підвищення продуктивності системи є використання індексів.

Індекс - це структура даних, яка допомагає СУБД швидше виявляти окремі записи у файлі, а тому дозволяє скоротити час виконання запитів користувачів [1]. Індеси бувають кластерні і некластерні. При наявності кластерного індексу рядки таблиці фізично зберігаються в заданому порядку і безпосередньо пов'язані з елементами індексу, тому значно прискорюється доступ до даних при використанні запитів, що використовують даний індекс. У кожній таблиці може бути лише один кластерний індекс. Некластерний індекс містить тільки покажчики на записи таблиці, у зв'язку з чим при вибірці необхідно принаймні ще одне звернення до диску для отримання власне запису.

У літературі наводяться неформальні критерії, що дозволяють визначити стовпці, які необхідно індексувати: стовпці, що використовуються для об'єднання таблиць; стовпці, що використовуються для обмеження діапазону даних, які вибираються при виконанні запитів; стовпці, що використовуються в директивах ORDER BY і GROUP BY запитів; стовпці, що використовуються у функціях підсумовування і підбиття підсумків.

Також описані випадки, у яких доцільно застосовувати кластерні індекси: стовпці використовуються в широкому діапазоні запитів; стовпці використовуються в директивах ORDER BY і GROUP BY запитів; стовпці використовуються для об'єднання таблиць; використовуються запити, які повертають великий результуючий набір даних.

На даний момент існують аналоги, які аналізують запити, дозволяють користувачеві переглядати, редагувати, створювати індекси або самостійно програма на основі аналізу запитів та структури БД створює індекси, наприклад, багатофункціональна компонента Microsoft SQL Query Analyzer за допомогою майстра Index Tuning Wizard проводить настройку індексів по запитам на базі зібраної статистики; програма Erwin

створює та модифікує індекси за «вказівкою» користувача. Аналізуючи аналоги можна зробити висновок: відсутні формальні методи відбору індексів, так як ці дані являються комерційною таємницею або програмний аналіз не використовується – адміністратор створює індекси самостійно, основуючись на свій досвід.

Таким чином на даний момент є актуальною розробка моделей та алгоритмів, які дозволять проаналізувати поведінку ІС та отримати множину полів кандидатів для створення кластерного індексу.

Уявімо запит у вигляді множини $Q = \langle T, \langle F, \langle R \rangle \rangle \rangle$, де T – множина таблиць БД, що беруть участь у запиті; F – безліч полів в таблиці T ; R – роль поля у запиті, $R \in \{ \dot{I}, Ord, Gr, Condit \}$, де \dot{I} – проекція в запиті, Ord – сортування, Gr – групування, $Condit$ – умова пошуку.

Статистику за даними таблиці наведемо у вигляді множини $\langle T, Q^T \rangle$, де Q^T – кількість рядків у таблиці. Характеристики поля F таблиці T :

$\langle F \langle f, q^f \rangle \rangle$, де f – значення поля F , q^f – кількість таких значень.

Статистику за запитами представимо у вигляді множини $\langle T, v_{ins}, v_{del}, \langle F, v_{gr}, v_{ord}, v_{sel}, t_{gr}, t_{ord}, t_{sel} \rangle \rangle$, де v_{ins} – відносна частота вставки в таблицю T , v_{del} – відносна частота видалення в таблиці T , v_{gr} – відносна частота групувань по полю F , v_{ord} – відносна частота сортувань по полю F , v_{sel} – відносна частота вибірок по полю F , t_{gr} – відносний час угруповань по полю F , t_{ord} – відносний час сортувань по полю F , t_{sel} – відносний час вибірок по полю F .

На основі отриманої інформації формується множина кандидатів на кластерний індекс $\langle T, F^{cl} \rangle$, де F^{cl} – поле-кандидат на створення кластерного індексу таблиці T .

Таким чином, загальна схема оптимізації структури кластерних індексів включає аналіз запитів, що надійшли в БД протягом періоду P , визначення статистики про використання таблиць і полів у запитах і формування множини полів кандидатів на

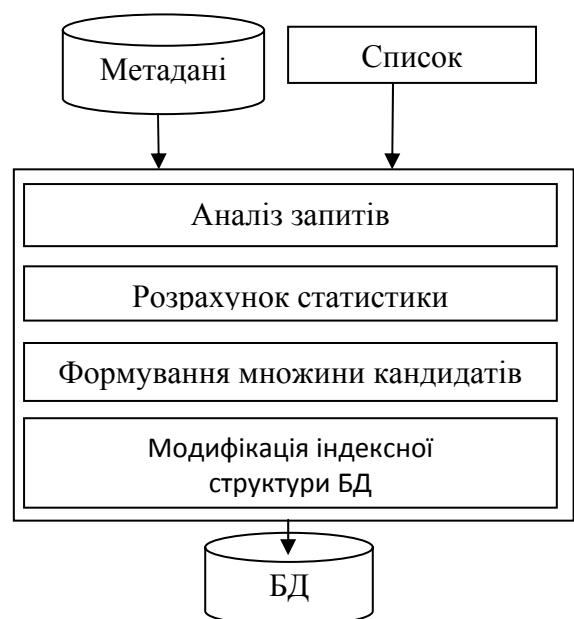


Рис.1. Загальна схема пошуку полів-кандидатів на створення кластерного індексу

кластерний індекс для кожної таблиці (див.рис. 1).

Запропоновані подання запиту і поведінки ІС дозволяють формалізувати пошук оптимальної структури індексного, при якій можливе максимальне зменшення часу виконання множини запитів. Опис правил переходу від запропонованої статистичної інформації до множини полів кандидатів дозволить завершити розробку формального методу пошуку оптимального кластерного індексу.

СПИСОК ЛІТЕРАТУРИ

Т. Конноллі, К. Бегг, А. Страчан Бази даних: проектування, реалізація й супровід. Теорія і практика /. Конноллі Т., Бегг К., Страчан А. - М.: Видавн. дім "Вільямс", 2001. - 1090 с.