УДК 004.6

ОБРОБКА ЗАПИТІВ У ХМАРНИХ БАЗАХ ДАНИХ

Ткаченко Є.В. – гр. МгКІ-24, магістрант, eyva.tkachenko@gmail.com
Стаценко Д.В. – к.т.н., доцент., statsenko.dv@knutd.edu.ua
Київський національний університет технологій та дизайну

Мета роботи. Підвищення ефективності обробки даних у розподілених хмарних базах даних шляхом удосконалення методів оптимізації запитів, зменшення часу виконання операцій і підвищення продуктивності системи зберігання даних у хмарному середовищі.

У сучасних інформаційних системах зростання обсягів даних і вимоги до швидкодії призвели до активного переходу від централізованих баз даних до розподілених хмарних систем. Такі системи дозволяють зберігати та обробляти дані на кількох вузлах мережі, забезпечуючи масштабованість, відмовостійкість і паралельність обчислень [1].

Основою функціонування таких систем є механізми реплікації та фрагментації, які дозволяють розподіляти дані між вузлами. Реплікація підвищує доступність і надійність, тоді як фрагментація зменшує обсяг переданих даних і підвищує швидкість виконання запитів [2].

Під час виконання SQL-запитів у розподіленому середовищі головною проблемою є мінімізація міжвузлового обміну. Найбільші витрати часу пов’язані саме з передаванням даних між серверами, тому оптимізація плану виконання запитів має вирішальне значення [3].

Для цього застосовуються:

Евристичні методи оптимізації, що наближено визначають найефективніший план виконання без повного перебору варіантів;

Логічна оптимізація – спрощення структури запиту без зміни результату;

Семантична оптимізація – виключення надлишкових операцій і умов;

Використання статистичних даних з системного каталогу для оцінки вартості операцій та вибору ефективного плану.

Архітектура системи управління розподіленою базою даних передбачає роботу оптимізатора запитів, який формує, оцінює та вибирає план виконання запиту з урахуванням особливостей хмарного середовища. У роботі розглянуто типову архітектуру системи з контролером оптимізації, що координує виконання запитів між вузлами та використовує алгоритми розподілу навантаження для балансування обчислень [4].

Результати дослідження показують, що застосування комбінованого підходу до оптимізації – логічного, семантичного та евристичного — дозволяє скоротити середній час виконання запитів до 30 %, а також підвищити ефективність використання ресурсів у хмарних базах даних.

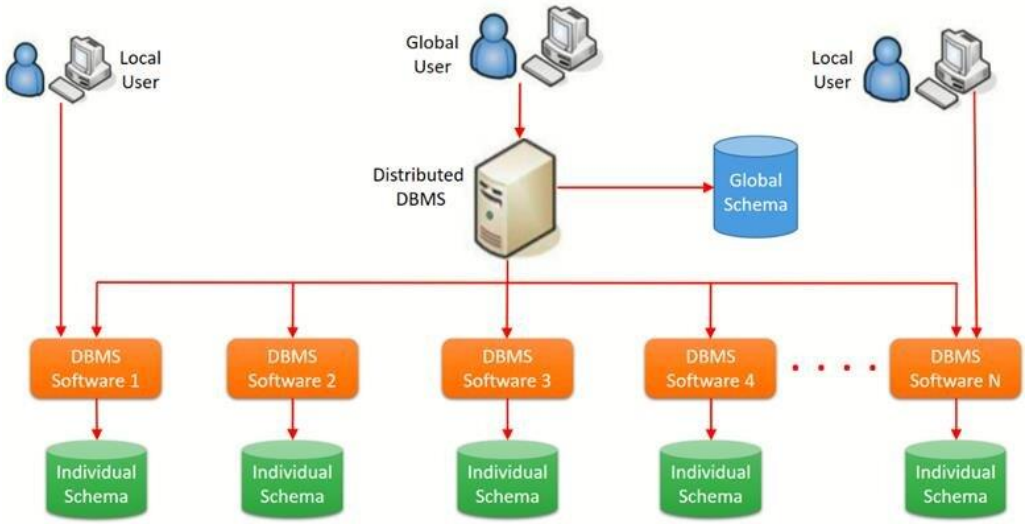

Рисунок 1 – Типова архітектура гетерогенної розподіленої бази даних (Farley J. Distributed Computing in Java 9. O'Reilly Media, 2018)

Запропоновані рішення сприяють зменшенню енергоспоживання дата-центрів за рахунок збалансованого розподілу навантаження і скорочення обсягів переданих даних, що відповідає концепції енергоефективних IT-систем.

Висновки

Оптимізація обробки запитів у хмарних розподілених базах даних є ключовим чинником підвищення їх енергоефективності. Запропоновані алгоритми дозволяють зменшити кількість операцій обміну, покращити швидкодію та забезпечити раціональне використання обчислювальних ресурсів. Це робить можливим створення більш стійких і продуктивних інформаційних систем для енергетичного сектору та управління інфраструктурою.

Список використаних джерел:

1. Özsu M. T., Valduriez P. Principles of Distributed Database Systems. 4th ed. – Cham: Springer, 2020. – 674 p.

2. Garcia-Molina H., Ullman J. D., Widom J. Database Systems: The Complete Book. 2nd ed. – Upper Saddle River, NJ: Prentice Hall, 2008. – 1080 p.

3. Kossmann J., Färber F., Lehner W. Data Dependencies for Query Optimization: A Survey. // The VLDB Journal. – 2022. – Vol. 31. – P. 1471–1494.

4. Pavlo A., Arulraj J., Lin H., Menon P., Mowry T., Pavlo R., Quah J., Stonebraker M. Self-Driving Database Management Systems. // Proc. CIDR Conference. – 2017.

5. Farley J. Distributed Computing in Java 9. O'Reilly Media, 2018.[Електронний ресурс]. – Режим доступу: https://www.oreilly.com/library/view/distributed-computing-in/9781787126992/b9234144-2b0c-4ebe-9638-b84ce094794f.xhtml