

2. Primary and Foreign Key Constraints. URL: <https://learn.microsoft.com/en-us/sql/relational-databases/tables/primary-and-foreign-key-constraints?view=sql-server-ver16>

3. Compound key. URL: http://wiki.gis.com/wiki/index.php/Compound_key

4. Candidate key in SQL. URL: <https://datatrained.com/post/candidate-key-in-sql/>

5. Different Keys in SQL (Primary Key, Candidate Key, Foreign Key). URL: <https://www.analyticsvidhya.com/blog/2020/07/difference-between-sql-keys-primary-key-super-key-candidate-key-foreign-key/>

УДК 004.65'416:316.472.4(043.2)

*Назаренко М. С., здобувачка 2 курсу спеціальності 122 Комп'ютерні науки,
Зелінська О. В., канд. техн. наук, доцент, в. о. завідувача кафедри інформаційних технологій*

ОПТИМІЗАЦІЯ БАЗ ДАНИХ ДЛЯ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

Донецький національний університет імені Василя Стуса, м. Вінниця

Вступ. Зі збільшенням кількості соціальних мереж стає все більш важливим всебічний аналіз цих мереж для розуміння різних явищ. Аналіз соціальних мереж (SNA) – це одне з таких досліджень, яке допомагає проаналізувати структуру мереж. SNA може виконувати складні аналізи, як-от прогнозування майбутніх тенденцій та виявлення важливих комунікаційних патернів. Однак обробка даних SNA може бути ресурсомісткою, і для забезпечення ефективності важливо оптимізувати структури баз даних. З огляду на це наше дослідження зосереджується на оптимізації баз даних для забезпечення швидкості та ефективності аналізу соціальних мереж.

Актуальність. Актуальність цієї теми визначається необхідністю вирішення проблем, пов'язаних зі зростанням обсягів інформації у соціальних мережах. Аналіз соціальних мереж має величезне значення для розуміння поведінки користувачів, прогнозування трендів та прийняття управлінських рішень у реальному часі.

Дослідження з питань оптимізації баз даних має таке визначальне значення через кілька аспектів. По-перше, оптимізовані бази даних забезпечують швидкий доступ та оброблення великого обсягу інформації, що є ключовим для вчасного реагування на зміни в соціальних мережах. По-друге, вони дають змогу покращити точність і правдивість аналізу, сприяючи більш точним прогнозам і трендам.

Отже, розглянемо методи оптимізації баз даних для SNA.

Індексація

Створення правильних індексів для таблиць у базі даних є, мабуть, однією з найкращих технік налаштування продуктивності [1]. Без індексів весь доступ до даних у базі повинен здійснюватися шляхом сканування всіх доступних рядків, що дуже нераціонально, особливо для великих таблиць.

Цей процес передбачає створення структур даних, які дають змогу швидко шукати та отримувати певну інформацію. Для аналізу соціальних мереж індексування відіграє важливу роль у ефективному пошуку відносин і зв'язків між окремими особами чи організаціями. Завдяки створенню відповідних індексів СУБД може швидко отримувати доступ і проходити через мережу, забезпечуючи швидший аналіз.

Паралельна обробка

Аналіз великомасштабних соціальних мереж часто вимагає значних обчислювальних ресурсів. Методи паралельної обробки передбачають розподіл робочого навантаження між кількома процесорами або машинами, у такий спосіб скорочуючи час аналізу. Завдяки використанню можливості паралельної обробки, задачі SNA, які раніше вимагали багато часу, тепер можуть виконуватися за менший проміжок часу.

Попередня обробка даних

Попередня обробка даних перед зберіганням і аналізом є ще одним менш важливим аспектом оптимізації. Цей процес передбачає їх очищення, фільтрування нерелевантної інформації та агрегування точок даних для зменшення розміру і складності даних. Виконуючи поставлені завдання, база інформації стає більш оптимізованою, що призводить до швидшого аналізу та підвищення ефективності зберігання. До того ж шляхом видалення нерелевантних даних можна підвищити якість результатів аналізу.

Графові бази даних

Під час роботи з взаємопов'язаними даними, як-от соціальні мережі, традиційні реляційні бази даних можуть не забезпечувати оптимальної продуктивності.

Графові бази даних – це тип баз даних, які зберігають дані як вузли та ребра, що представляють сутність та зв'язки відповідно. Вони дають змогу моделювати й запитувати дані природним та інтуїтивно зрозумілим способом, використовуючи теорію графів і алгоритми [2]. Такий тип БД особливо підходить для аналізу соціальних мереж, оскільки вони може обробляти великий обсяг різноманітних даних соціальних мереж, а також забезпечують швидкий і гнучкий аналіз властивостей мережі, як-от центральність, кластеризація або найкоротші шляхи.

Щоб використовувати бази даних графів для SNA, потрібно визначити схему мережі, тобто типи та атрибути вузлів і ребер мережі (наприклад, можуть бути

вузли для користувачів, публікацій, коментарів і груп, а також межі для вподобайок, підписок та згадок).

Графові БД пропонують власну підтримку для обходу та запиту взаємопов'язаних даних, що робить їх зручними для аналізу соціальних мереж. Їх ефективні механізми обходу графів і вбудовані алгоритми значно підвищують швидкість і продуктивність завдань аналізу соціальних мереж.

Прикладом використання цих БД для SNA є Facebook [3]. Це є один із найбільших сайтів соціальних мереж, і його функція Graph Search дає змогу користувачам шукати вміст на платформі. Щоб створити Graph Search, інженерам Facebook потрібно було оптимізувати свою базу даних для обробки величезної кількості даних, створених їх користувачами.

Graph Search від Facebook використовує Social Graph – структуру бази даних, яка зберігає зв'язки між користувачами. Соціальний графік містить дані про користувачів, публікації, місця і події та ін. Інженери Facebook оптимізували записування Social Graph на диск і читання в пам'яті, щоб мінімізувати перевантаження системи. Щоб зменшити надмірність даних, вони використовували такі методи матеріалізованого представлення та попереднього завантаження даних у розділи. Оптимізація бази даних Social Graph допомогла пошуковій системі Facebook підтримувати десятки мільйонів щоденних запитів.

Висновки. Отже, ефективний аналіз соціальних мереж для отримання важливої інформації вимагає ретельної оптимізації баз даних, які зберігають і обробляють величезну кількість взаємопов'язаних даних. Завдяки використанню вищезазначених методів оптимізація бази даних може значно підвищити швидкість, масштабованість і результативність аналізу соціальних мереж.

Оскільки розмір і складність соціальних мереж продовжують зростати, існує постійна потреба в інноваційних підходах до оптимізації баз даних для SNA. Подальші дослідження та розробки в цій галузі сприятимуть кращій продуктивності та аналізу в реальному часі.

Список використаних джерел

1. Techniques for optimizing databases. URL: <https://www.dbta.com/Columns/DBA-Corner/Techniques-for-Optimizing-Databases-154027.aspx> (дата звернення: 07.07.2022)
2. How Graph databases can boost social network analysis. URL: <http://surl.li/nwidk>
3. Facebook Graph Search. URL: https://en.ryte.com/wiki/Facebook_Graph_Search