

В. Ярута, М. Панасюк

МОЖЛИВОСТІ СУЧАСНИХ СИСТЕМ УПРАВЛІННЯ БАЗАМИ ДАНИХ З ПІДТРИМКИ РОБОТИ З ДОКУМЕНТАМИ

V. Yaruta, M. Panasiuk

CAPABILITIES OF MODERN DATABASE MANAGEMENT SYSTEMS IN SUPPORTING WORK WITH DOCUMENTS

Проблема опрацювання великих обсягів даних, значна кількість яких представлена в документній формі в різноманітних галузях, таких як: банківська справа та фінанси, телекомунікації, Інтернет та соціальні мережі, логістика і транспорт, наука та дослідження, медицина тощо — з часом лише посилюється. Нині для її вирішення використовуються різноманітні системи управління базами даних (СУБД), що підтримують роботу з моделями баз даних (БД) різних типів.

Класичними є ієрархічна, мережева та реляційна моделі. Ієрархічна ефективно працює з даними, що можуть бути представленими у вигляді дерева, забезпечує їх швидкий аналіз, але має незручний пошук та обмежену гнучкість. Мережева ефективно працює з даними, що можуть бути представленими у вигляді графа, та, на відміну від попередньої, забезпечує інтегрованість та швидкий доступ до даних, але має складну структуру та високі вимоги до апаратного забезпечення. Реляційна, хоча на сьогодні є найпоширенішою та дозволяє ефективно працювати з добре структурованими даними, представленими у виді системи пов'язаних таблиць, але має підвищені вимоги до обсягів зовнішньої пам'яті та не підходить для неструктурованих даних, таких як фото, відео, аудіофайли.

Найбільш ефективними під час роботи з документами різних типів (вебсторінки, текстові файли, зображення, аудіо- і відеофайли, таблиці тощо) є документ-орієнтовані БД, які з'явились у 1980-х рр. та набули найбільшого поширення з появою Інтернету та вебтехнологій. Серед сучасних СУБД такого типу виокремлюють MongoDB, RavenDB та CouchDB.

MongoDB — СУБД з відкритим кодом, написана з використанням мов програмування JavaScript, C++, Python, розповсюджується відповідно до ліцензії Server Side Public License, працює на операційних системах Linux, Windows, Mac OS, Solaris, підтримує роботу у хмарі. Розроблена у 2007 р. компанією MongoDB Inc, стабільна версія MongoDB 4.4 випущена 30 липня 2020 р. Використовує власний формат для зберігання інформації — Binary JavaScript Object Notation (BSON), має високу продуктивність, дозволяє швидко знайти дані завдяки індексованому пошуку. До основних переваг цієї СУБД належить можливість виконувати горизонтальне масштабування через розгортання на множині машин, організувати документо-орієнтоване сховище, виконувати динамічні запити, повну підтримку індексів, швидке оновлення «на місці», ефективне зберігання бінарних даних великих обсягів, журналювання операцій, що модифікують дані в БД, підтримку відмовостійкості. На думку розробників, MongoDB має посісти місце між малими та великими СУБД. Вона використовується, наприклад, банком Nationwide Building Society для швидшого створення додатків та управління ними; компанією медичного страхування Humana для моделювання даних з основних платформ, щоб зберігати їх у більш стандартизованому форматі; компанією Noodoe для організації основної БД при управлінні зарядними станціями електромобілів.

RavenDB — СУБД з відкритим кодом, яка написана на мові C Sharp, розповсюджується відповідно до ліцензії AGPLv3, працює на операційних системах Windows, Linux, Mac OS, Docker та Raspberry Pi, підтримує роботу в хмарі. Розроблена у 2008 р. компанією Hibernating Rhinos, стабільна версія RavenDB 5.3 випущена у листопаді 2021 р. Використовує динамічну таблицю індексів, що дозволяє обробляти дані швидше за інші СУБД, має вбудовану функцію контролю помилок та повторюваних запитів, а також автономну архітектуру, яка не потребує зовнішніх серверів. До основних особливостей цієї СУБД належить те, що запити формуються за допомогою мови LINQ або власної мови Raven Query Language (RQL), використовуються графові запити, механізм проєкції (для перетворення індексованих даних, виконання обчислень, агрегування та виконання коду JavaScript з боку сервера), а також повнотекстовий пошук. СУБД RavenDB використовується, наприклад, компанією TOAD LLC для спрощення вибору матеріалів відповідно до місцевих будівельних норм у галузі меблевого виробництва та компанією FinStat для спрощення організації та агрегування великих обсягів публічних статистичних даних у платформі HitHorizon's.

CouchDB — розподілена документо-орієнтована СУБД з відкритим кодом, що не вимагає опису схеми даних, ядро якої написане на мові програмування Erlang, розповсюджується відповідно до ліцензії Apache License 2.0, працює на операційних системах Linux, Microsoft Windows і Mac OS. Розроблена у 2005 р. фірмою Apache Software Foundation, перша стабільна версія випущена 11 липня 2011 р. У цій СУБД одиницею збереження даних є документ, що має унікальний ідентифікатор та версію і містить довільний набір полів; моделлю даних є дерево; для їх організації застосовується концепція формування видів; цілісність даних забезпечується на рівні записів, при цьому не підтримуються зв'язки між таблицями або записами; проте підтримується паралельна обробка даних та вертикальна масштабованість, а отже, підтримуються портативні пристрої. СУБД CouchDB використовується, наприклад, авіакомпанією United Airlines для організації системи розваг під час польотів; організацією Червоний Хрест для електронного оформлення справ у зонах стихійних лих; Європейською організацією з ядерних досліджень CERN для управління даними на Великому адронному колайдері.

Отже, сучасні документ-орієнтовані системи управління базами даних є крос-платформовим програмним забезпеченням, що дозволяє індексувати документи за різними критеріями (автор, дата, тип документа, його тематика тощо), налаштувати їх пошук за ключовими словами та тегами, крім того організувати повнотекстовий пошук, що, як порівняти з іншими моделями баз даних, значно спрощує та пришвидшує маніпуляції з документами.