

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОЕКТУВАННЯ ЛОГІЧНОЇ МОДЕЛІ ГЕТЕРОГЕННОЇ БАЗИ ДАНИХ

Перепичай О. І.

Науковий керівник – к.т.н., доц. Мазурова О. О.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: oksana.mazurova@nure.ua

One of the important stages in the development of information systems with a high load is the design of distributed databases or the reengineering of already existing databases. Today, the use of distributed heterogeneous databases, which use the advantages of various logical database models for data storage and processing, has become widespread. The proposed information technology for designing a logical model of a distributed heterogeneous database, which takes into account the variety of NoSQL and other data models that exist today; combines the design of business logic with the development of traditional elements of logical DB models.

В наш час активного використання різних інтернет-застосунків, в тому числі, систем електронної комерції, важливим є питання масштабування таких систем. В напрямку створення високо масштабованих систем все більш поширеними стають інформаційні системи, що побудовано на основі мікросервісної архітектури.

Одним із важливих етапів розробки таких систем є проектування розподілених баз даних (БД), що покладено в їх основу, або реінженіринг вже існуючих БД [1]. На сьогодні широкого розповсюдження набуло використання розподілених гетерогенних БД, які для зберігання та обробки даних використовують переваги різних логічних моделей БД.

Існуючі підходи до моделювання розподілених гетерогенних БД в основному представлені у вигляді загальних рекомендацій щодо проектування мікросервісних архітектур та «кращими практиками» використання NoSQL [2] та інших систем керування базами даних (СКБД) для вирішення окремих практичних задач. Компанії-розробники нових NoSQL та інших моделей, як правило, не пропонують методологій проектування логіки своїх моделей, а надають інструментар для їх фізичного створення. Далі розробникам БД пропонується оцінити якість отриманого рішення за тими чи іншими моделями [3].

Єдина методологія логічного моделювання гетерогенних баз даних повинна дозволити врахувати різноманіття з'явившихся на сьогодні моделей даних; позбавитися існуючого відокремлення проектування бізнес-логіки (а саме транзакцій БД) від розробки логічної моделі БД. Основні результати досліджень в цій області мають неформальний

характер, і вимагають розробки відповідного інформаційного забезпечення.

Метою даної роботи є розробка інформаційної технології, яка дозволять підвищити ефективність систем автоматизації проектування розподілених гетерогенних баз даних, що будуть підтримувати проектування або реінженіринг БД, яке відповідає вище переліченим вимогам на основі багатокритеріальної оцінки якості такої моделі.

Запропонована інформаційна технологія проектування логічної моделі розподіленої гетерогенної бази даних основана на використанні розширеної інфологічної моделі БД, методів проектування, оптимізації та оцінки якості розподіленої гетерогенної БД.

Запропонована інформаційна технологія містить такі етапи.

- відбір та попередня підготовка вхідних даних; на даному етапі для обраної предметної галузі проводяться аналіз, а також етапи концептуального та інфологічного моделювання БД, що відповідають традиційній методології проектування БД, а саме формується ER-діаграма та відповідні UML-діаграми, що описують бізнес-процеси в предметній галузі;

- побудова розширеної інфологічної моделі; на даному етапі до промодельованих сутностей та зв'язків ER-діаграми додаються елементи моделювання розподілених транзакцій, що притаманні визначеному бізнес-процесу; в тому числі, моделюється інтенсивність, обсяг та характер потоків даних, що складають основу транзакцій, тощо;

- декомпозиція отриманого моноліту розширеної інфологічної моделі; на даному етапі проводиться декомпозиція (відсікання від моноліту) розширеної інфологічної моделі в систему взаємозв'язаних структурних одиниць на основі отриманих раніше характеристик інтенсивності, обсягу та характеру потоків даних між цими структурними одиницями;

- визначення типів логічних моделей; на даному етапі визначається множина ефективних типів логічних моделей БД для відокремлених зв'язок структурних одиниць з урахування особливостей логічних моделей, а також CAP та PACELC теорем щодо створення розподілених систем;

- оцінка якості розподіленої гетерогенної БД та її оптимізація; на даному етапі проводиться багатокритеріальна оцінка якості спроектованої розподіленої гетерогенної БД; в разі необхідності на основі отриманих функціонально-вартісних характеристик проводиться оптимізації логічних структур, що входять до складу моделі.

Для того, щоб довести ефективність та практичну цінність розробленої інформаційної технології, планується провести експериментальне дослідження в наступних умовах:.

- в якості предметної області для проектування БД буде використана область електронної комерції;

- будуть промодельовані типові для електронної комерції транзакції з підтримки торгових операцій, запити на отримання типових статистик;
 - буде програмно реалізовано множину розподілених транзакцій та розподілених запитів для гетерогенної розподіленої БД в області електронної комерції;
 - всі заміри щодо продуктивності БД планується виконувати на кластерах серверів баз даних, які будуть розташовано на віртуальних машинах різного розміру (в якості технічних умов проведення експериментів будуть розглянуті комбінації різних показників vCPU, RAM (ГБ), кількості вузлів та кількості підключень);
 - під час експериментів буде знято заміри за наступними метриками продуктивності БД: місце, яке займає БД (МБ); час виконання розподіленого запиту (мс); час виконання розподіленої транзакції (мс); споживання оперативної пам'яті (МБ); споживання процесорного часу (%), тощо.
- Використання інформаційної технології передбачає отримання наступних результатів:
- скорочення часу виконання розподілених транзакцій та розподілених запитів в спроектованій гетерогенній системі за рахунок усунення витрат, які викликані необхідністю перевірки розподілених обмежень цілісності;
 - скорочення часу виконання CRUD-операцій на окремих логічних структурах БД за рахунок використання ефективно підібраних логічних моделей БД (і відповідних СКБД);
 - скорочення часу виконання розподілених запитів до гетерогенної БД за рахунок використання моделі, що враховує глибинні знання про потоки даних в предметній галузі;
 - відповідно забезпечення високої масштабованості інформаційної системи, що розроблена на основі запропонованої розподіленої гетерогенної БД.

Список використаних джерел:

1. Filatov V., Semenets V. (2018). Methods for Synthesis of Relational Data Model in Information Systems Reengineering Problems. International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). IEEE.
2. Gomes C., Borba E., Tavares E., Junior M. N. de O. (2019). Performability Model for Assessing NoSQL DBMS Consistency. IEEE International Systems Conference (SysCon). <https://doi.org/10.1109/syscon.2019.8836757>.
3. Mazurova O., Syvolovskyi I., Syvolovska O. (2022). NOSQL database logic design methods for MONGODB and NEO4J. Innovative technologies and scientific solutions for industries, 2(20), 52–63. <http://journals.uran.ua/itssi/article/view/262054>.