

services help IT deliver business value [White Paper] / IBM Global Technology Services. IBM Corporation, 2017. URL: <https://www.ibm.com/downloads/cas/5GKMXAYZ> (дата звернення 12.02.2021 р.). 3. *Nokes C.* 9 Steps for Selecting the Right IT Services Provider // Сайт "ITSM tools". URL: <https://itsm.tools/9-steps-for-selecting-the-right-it-service-provider/> (дата звернення 12.02.2021 р.). 4. *Viriyasitavat W.* Multi-criteria selection for services selection in service workflow // *Journal of Industrial Information Integration*. 2016. Vol. 1. P. 20-25. DOI: 10.1016/j.jii.2016.03.003 5. *Wang C., Ma H., Chen G.* "Using EDA-Based Local Search to Improve the Performance of NSGA-II for Multiobjective Semantic Web Service Composition", In: Hartmann S., K?ng J., Chakravarthy S., Anderst-Kotsis G., Tjoa A., Khalil I. (eds) *Database and Expert Systems Applications. DEXA 2019. Lecture Notes in Computer Science*, vol 11707. Springer, Cham., 2019. P. 434-451. DOI: https://doi.org/10.1007/978-3-030-27618-8_32. 6. *Zhang N.* Service Discovery and Selection Based on Dynamic QoS in the Internet of Things // *Complexity*. 2021. Vol. 2021. Article ID 6642514. DOI: <https://doi.org/10.1155/2021/6642514>. 7. *Bekkouche A., Benslimane S.M., Huchard M., Tibermacine C., Hadjila F., Merzoug M.* QoS-Aware Optimal and Automated Semantic WebService Composition With User's Constraints // *Service Oriented Computing and Application*. 2017. Vol. 11. P. 183-201. DOI: 10.1007/s11761-017-0205-1. 8. *Petrychenko A., Levykin I., Iuriev I.* Improving a method for selecting information technology services // *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 2. No. 2 (110). P. 32-43. DOI: 10.15587/1729-4061.2021.2299. 9. *Левыкин В.М., Юрьев И.А.* Модель выбора набора ИТ-сервисов для конечных пользователей // *Вісник НТУ "ХПІ"*. Серія: Системний аналіз, управління та інформаційні технології. Х.: НТУ "ХПІ", 2016. №45 (1217). С. 78-84.

Надійшла до редколегії 17.06.2021

Юр'єв Іван Олексійович, старший викладач кафедри інформаційних управляючих систем ХНУРЕ. Наукові інтереси: методи, моделі, інформаційні технології управління інформаційними системами. Адреса: Україна, 61166, м. Харків, пр. Леніна, 14, тел. (057) 702 14 51.

УДК 004.652.4

DOI: 10.30837/0135-1710.2021.177.063

Т. Г. БІЛОВА, В. М. ДЬОМІНА, І. О. ПОБІЖЕНКО

КОМПОЗИЦІЙНЕ ПРОЕКТУВАННЯ ТА РЕСТРУКТУРИЗАЦІЯ ГЛОБАЛЬНОЇ СХЕМИ МУЛЬТИБАЗИ ДАНИХ

Розглянуто особливості створення глобального шару над даними в розподілених базах даних. На основі аналізу варіантів представлення глобальних схем запропоновано поєднання централізованого та децентралізованого підходів. Визначено основні задачі композиційного проектування та реструктуризації глобальної схеми мультибази даних. Формалізовано процеси додавання та видалення локальних вузлів, різних варіантів оновлення з підтримкою обмежень цілісності.

1. Вступ

Проблема інтегрування великих обсягів даних визначає вимоги до представлення їх у вигляді, зручному для використання та перетворення. Користувачу розподіленої бази даних (РБД) надається інформація про дані у вигляді глобальної схеми. Вибір моделі подання такої схеми базується на аналізі вимог користувачів кінцевої системи та не має залежати від типу БД, набору прикладних програм, які створені або будуть створені, мов програмування, що використовуються, самої обчислювальної платформи тощо. Модель представлення даних, отримана в результаті концептуального проектування, є джерелом інформації для фази логічного проектування БД.

Побудова глобальної схеми на основі традиційної реляційної моделі дає можливість використовувати для побудови запитів апарат реляційної алгебри. Недоліком такого підходу є складність приведення моделей нереляційних баз даних до реляційної схеми.

Питання розробки загального підходу для представлення глобального рівня даних досліджуються в роботах багатьох науковців. Вибір графової моделі [1-3] дозволяє представити відношення між розподіленими даними у вигляді неорієнтованого графа, що визначає зв'язки між даними, розподіленими по різних вузлах системи. Такий підхід дозволяє створити єдиний інтерфейс доступу до усіх даних. Процес управління графовою моделлю здійснюється за допомогою операцій об'єднання, перетину та алгоритмів обходу у глибину [1]. Іншим підхо-

дом є рекурсивне представлення [4] у вигляді багаторівневої інтеграції гетерогенних джерел даних у єдиний інформаційний простір. Перевагою такого підходу є можливість динамічної зміни схеми даних та перспективне масштабування системи за рахунок приєднання нових моделей.

Недослідженими залишаються питання виникнення конфліктів в процесі проведення реструктуризації глобального рівня у разі зміни інформації на одному з вузлів. Актуальною є задача розробки єдиного підходу до проведення проектування та реструктуризації глобальної схеми в реальному часі з урахуванням обмежень цілісності та несуперечливості даних, що оброблюються, та забезпеченням безперервного доступу користувачів до розподіленої системи.

Метою дослідження є визначення представлення глобального шару даних в розподіленій системі таким чином, щоб забезпечити ефективний процес її подальшої реструктуризації.

В рамках поставленої мети потрібно вирішити наступні задачі: проаналізувати можливі варіанти представлення глобального рівню управління даними в РБД; формалізувати задачу проектування глобальної схеми РБД; визначити підтримку процесів управління глобальною схемою з урахуванням вимог цілісності та доступності інформації для користувачів; розробити діаграму станів процесу реструктуризації глобальної схеми.

2. Аналіз варіантів представлення глобальної схеми та постановка задачі проектування

Розподілена система інтегрує різноманітні ресурси у єдине інформаційне середовище. Проектування такої системи приходиться виконувати методом «знизу до гори», тобто шукати механізми об'єднання вже існуючих баз даних таким чином, щоб представити користувачеві можливість доступу до ресурсів у той же спосіб, як і до централізованої системи. Класичне визначення за Дейтом [6] дванадцяти принципів побудови РБД потребує повної автономності кожного локального вузла, але на практиці досягнути цього майже неможливо.

Для забезпечення сумісної роботи в РБД використовуються наступні засоби:

- фрагментація - розбиття БД або таблиці на кілька частин і зберігання цих частин на різних вузлах РБД;
- реплікація - створення і зберігання копій одних і тих же даних на різних вузлах РБД;
- розподілені обмеження цілісності - обмеження, для перевірки виконання яких потрібно звернення до іншого вузла РБД;
- розподілені запити - запити на читання, які звертаються більш ніж до одного вузла РБД;
- розподілені транзакції - команди на зміну даних, які звертаються більш ніж до одного вузла РБД.

Одним з рішень при побудові розподілених систем є створення мультибази даних. Під мультибазою даних з глобальною схемою розуміється реалізація зовнішнього інтерфейсу доступу до множини локальних СУБД, тобто створення глобального рівню над локальними СУБД.

Мультибаза даних є розподіленою гетерогенною системою, до її характеристик слід віднести неоднорідність у представленні та використанні даних, різні схеми представлення даних та різна структура компонентів, різні інтерфейси та протоколи використання ресурсів та ін. Все це повинно бути приховано від користувача, який працює з розподіленою системою в той самий спосіб, що й з централізованою. Додатково може виникати потреба в обробці інших типів даних, що не є структурованими та надходять з різних джерел.

Проектування мультибазової РБД базується на способах та методах її використання, тобто на стратегії формування запитів, що, в свою чергу, визначає схеми фрагментації та реплікації даних. Вимоги до високої продуктивності розподілених баз даних потребують ефективних процедур оновлення схеми даних. В разі зростання кількості вузлів процес оновлення глобальної схеми може негативно вплинути на середній час виконання розподілених запитів.

Глобальна схема мультибази даних визначається як загальне представлення про дані на локальних вузлах та способи доступу до них. Для її підтримки необхідно реалізувати два етапи: проектування та реструктуризації.

Сформулюємо загальну задачу композиційного проектування глобальної схеми бази даних як процес визначення загальної концепції проектування, аналіз компонентів глобальної схеми та визначення і пошук варіантів оптимального розміщення компонентів.

Порівняльний аналіз різних видів представлення глобальної схеми подано у таблиці 1.

Таблиця 1-

Порівняльний аналіз варіантів представлення глобальної схеми

Глобальна схема	Переваги	Недоліки
Централізована – розміщення на спеціальному виділеному вузлі	– централізоване представлення даних; – швидкість оновлення; – єдині правила оновлення та підтримки цілісності даних	– порушення принципів локальної автономності; – покладання функцій управління на виділений центральний вузол
Реплікована – кожен вузол зберігає власну копію глобальної схеми	– швидкість виконання запитів; – підтримують гомогенні РБД; – єдині правила оновлення та підтримки цілісності даних	– необхідність підтримки усіх реплік в актуальному стані; – надмірні витрати пам'яті та часу на реструктуризацію
Секціонована – на кожному вузлі зберігаються схеми тільки тих вузлів, з яких може бути затребувана інформація	– децентралізація управління даними; – локальна автономія кожного з вузлів; – підтримують гетерогенні РБД	– складність процедур оновлення даних; – відсутність єдиного представлення даних
Секціонована з централізованою реплікацією – на кожному вузлі зберігається його локальна схема, а на центральному вузлі зберігаються репліки локальних схем	– поєднання централізованого та децентралізованого підходів; – підтримують гетерогенні РБД; – є можливість виконання розподілених запитів, які попередньо не були заплановані	– порушення принципів локальної автономності; – відсутність єдиного представлення даних; – покладання функцій управління на виділений центральний вузол

Як видно з табл. 1, кожне представлення має свої переваги та недоліки. Підвищити ефективність представлення глобальної схеми можна за рахунок знаходження компромісу між централізованим та децентралізованим розміщенням глобального шару даних.

Розглянемо постановку задачі управління глобальною схемою у випадку централізованого зберігання на виділеному вузлі з підтримкою для підвищення надійності реплікованих копій на n спеціально виділених «резервних» вузлах, число яких пропорційно кількості вузлів РБД.

В загальному вигляді проектування глобальної схеми S можна представити як ітераційний процес додавання нових вузлів, тобто як реструктуризацію. Тоді задачу додавання локальної схеми з нового вузла в глобальну схему можна представити наступним чином:

а) перевірка схеми на пустоту; якщо схема пуста, то схему локального вузла представляємо як глобальну;

б) якщо схема не пуста, то виконуємо процедуру додавання нового вузла.

Задача реструктуризації глобальної схеми полягає у підтримці основних операцій маніпулювання: оновлення; модифікації, розширення шляхом додавання нових локальних схем.

3. Методи проектування та реструктуризації глобальної схеми

Розглянемо глобальну схему бази даних як структуру, що визначає множину об'єктів - локальних схем БД та операцій над ними. Визначимо необхідні операції для визначеного варіанту представлення глобальної схеми.

Для формалізації процедури додавання нового вузла представимо управління процесом реструктуризації у вигляді трійки параметрів $Z = \langle S, E, \Phi \rangle$, де S - глобальна схема, що є результатом об'єднання множини логічних схем БД локальних вузлів; E - множина схем

сутностей для роботи з даними; Φ - множина відображень схеми представлень локальних вузлів на схеми сутностей

$$\Phi = \{ \phi : S \rightarrow \Xi \mid \Xi \in E \}, \quad (1)$$

де Ξ - елемент множини відображень схеми представлень локальних вузлів на схеми сутностей, $\phi \in \Phi$.

Представлення (1) дозволяє інтегрувати різномірні локальні схеми в єдину, центральну схему для реалізації єдиного механізму управління даними, незалежно від моделей даних та схем локальних даних окремих вузлів.

Даний спосіб представлення дозволяє моделювати схему зв'язку зі схемами даних локальних вузлів, зводячи їх до все простіших елементів. Кожне відношення глобальної схеми даних представляється у вигляді набору трійок: набір атрибутів, клас розбиття домену та ідентифікатор локального вузла, на якому вони розміщені. Крім цього, глобальна схема повинна містити зв'язки між відношеннями. А відношення в свою чергу містять набір атрибутів, домени яких визначають розміщення даних в вузлах мультибази даних.

Дані будь-якого вузла поміщаються в атрибути трійок. Кожному локальному вузлу ставиться у відповідність набір зв'язків з іншими вузлами, набір обмежень цілісності і набір допустимих операцій. Кожен вузол позначимо $S_{k,e}^i$

$$S_{k,e}^i = \left\langle \left\langle \left\langle a_j^i, d_k^i, s_{k,e}^i \right\rangle \mid l \in L^i, L^i < L, k \in K^i \right\rangle, \Lambda_j^i, P_m^i(s_{k,e}^i), O_q^i \right\rangle,$$

де $j, m, q \in [1, \theta]$, $\theta \in \mathbb{R}$, $N : [1, \theta] \rightarrow \mathbb{R} \triangleright \triangleleft \Omega$.

Кожен вузол позначається трьома індексами: i - номер батьківського локального джерела даних (вузла); k - індекс відповідного класу розбиття домену на непересічні класи ($K = \{K_1^i, K_2^i, K_3^i, \dots, K_k^i\}$); e - грає роль унікального ключа. Індеси j, e, m, q - номери джерел локальних даних; R - максимальна кількість вузлів; a_j^i - атрибути елементів даних (призначені для зберігання даних); Ω - множина перспективних локальних вузлів, з якими можуть бути встановлені зв'язки, L - максимальна кількість атрибутів.

Множину зв'язків даного вузла позначимо Λ_k^i , $\Lambda_k^i = \left\langle \left\langle \left\langle s_{k,e}^i \right\rangle, \gamma \mid \gamma \in \Gamma, \mu \right\rangle \right\rangle$, де Γ - множина атрибутів зв'язків; μ - назва зв'язку; індекс e грає роль унікального ключа (у всіх вузлів в схемі даний індекс унікальний).

Множину предикатів, які задають обмеження цілісності на множині схем локальних вузлів, позначимо як $P_j^i(S)$. Множина S може містити схеми даних локальних вузлів S_r , тобто її можна представити як вектор $S = \langle S_1, S_2, \dots, S_r \rangle$, $r \in \mathbb{R}$.

Схема локального вузла $s_{k,e}^i$ знаходиться в узгодженому стані, якщо параметри локальної схеми з локального вузла додані до глобальної схеми та задані все предикати з множини предикатів, які задають обмеження цілісності для даного вузла. Таким чином, $s_{k,e}^i$ - в узгодженому стані, якщо $\forall s : s \in S \cup \{s_{k,e}^i\}$ виконується умова $\forall \phi : \phi \in P(s), \phi(s) = \text{true}$, де s - довільна схема локального вузла; $\phi(s)$ - деякий предикат, $P(s)$ - множина предикатів, які задають обмеження цілісності для локального вузла s .

Обмеження на діапазон допустимих значень представляються як $\Xi = a < A < b$, $a, b \in \Xi$, де A - атрибут локального вузла; Ξ - множина, в якій беруть значення атрибути локального вузла. Обмеження посилянь: $\Xi(s^x, s^y) = \forall s^x \exists s^y (s^x \rightarrow s^y)$. Агреговані обмеження: $\Xi = \max(A) > a$, $a \in \Xi$, де A - атрибут локального вузла.

Множину операцій для даного локального вузла позначимо O_k^i :

$$O_k^i = \{ \varphi : S_{k,e}^i \rightarrow S_{k,e}^i \}, k \in K^i .$$

де $s_{k,e}^i$ - множина схем даних для i -го локального вузла, що містить домен K^i . Всі операції з глобальною схемою даних можуть бути розділені на чотири типи: вибірка; оновлення; видалення; додавання.

Нехай s^x - схема локального вузла, що додається. Операція додавання для схеми локального вузла $s_{k,e}^i$ має вигляд

$$\phi : S_k^i \rightarrow S_k^i, k \in K^i, S_k^i = S_k^i \cup \{s^x\} .$$

Нехай s^x - схема з локального вузла, що видаляється. Операція видалення для схеми локального вузла $s_{k,e}^i$ має вигляд

$$\phi : S_k^i \rightarrow S_k^i, k \in K^i, S_k^i = S_k^i \setminus \{s^x\} .$$

Операція поновлення для схеми локального вузла $s_{k,e}^i$ має вигляд

$$\phi : S_k^i \rightarrow S_k^i, k \in K^i, S_k^i = S_k^i .$$

У процесі оновлення у локальній схемі можуть змінюватися атрибути, обмеження цілісності, набір зв'язків, набір операцій. Операції додавання і видалення схем локальних вузлів, що знаходяться в узгодженому стані, зберігають узгодженим стан схеми локального вузла, для якого ці операції виконувалися.

Розглянемо два типи зв'язків між локальними схемами різних вузлів системи: структурні і асоціативні. Структурні зв'язки є неявними і обумовлені структурою схеми даних.

Асоціативні зв'язки дозволяють зв'язати схеми локальних вузлів, які не пов'язані безпосередніми структурними зв'язками. Асоціативні зв'язки з'єднують локальні вузли неявно, через проміжний вузол. Якщо розглядати структурні зв'язки як вертикальні, то асоціативні можна розглядати як горизонтальні.

Кожен зв'язок характеризується атрибутом, який характеризує дії, спрямовані на підтримку посилальної цілісності під час виконання операцій видалення над пов'язаними локальними вузлами. Атрибут може приймати одне з наступних значень:

- каскадне видалення: при видаленні схеми локального вузла s , на який посилається схема локального вузла t , схема локального вузла t також буде видаленою;

- заборона видалення: якщо при видаленні схеми локального вузла s існує схема локального вузла t , який посилається на локальний вузол s , то операція видалення не буде дозволена; щоб її здійснити, необхідно або спочатку видалити схему локального вузлу t , або перенаправити зв'язок від нього на інший вузол;

- видалення зв'язку: при видаленні схеми локального вузла даних s видаляється зв'язок, котрий посилається на нього від деякої схеми локального вузла t ; при цьому схема локального вузла t продовжує існувати незалежно.

Характер зв'язків визначає процедуру оновлення глобальної схеми, яка повинна бути виконана. Структурні зв'язки вимагають оновлення одного вузла, асоціативні зв'язки передбачають більш складне каскадне оновлення, яке передбачає формування глобального плану, що визначає порядок взаємодії вузлів.

В загальному вигляді реструктуризацію глобальної схеми можна визначити як циклічний безперервний процес, що містить наступні етапи:

- стан додавання нового вузла;
- стан моніторингу;
- стан оновлення, що охоплює один локальний вузол;
- стан каскадного оновлення, що охоплює декілька локальних вузлів;

- стан оновлення резервних копій глобальної схеми (ГС).

На рис. 1 представлена діаграма переходів станів процесу реструктуризації глобальної схеми.



Рис. 1. Діаграма переходів станів процесу реструктуризації глобальної схеми

4. Висновки

Розроблений підхід до проектування та реструктуризації глобальної схеми дозволяє виконувати операції над метаданими з підтримкою цілісності та надійності роботи системи. Поєднання централізованого та децентралізованого представлення глобальної схеми надає можливість ефективно проводити процес реструктуризації в режимі обслуговування розподілених запитів. Розроблена діаграма станів реструктуризації відображує безперервний процес функціонування РБД з контролем актуального стану глобального шару опису даних.

Подальшого розвитку потребує вирішення проблеми децентралізації управління глобальною схемою. Питання управління частковою глобальною схемою, коли на кожному вузлі зберігаються схеми тільки тих вузлів, з яких може бути затребувана інформація, будуть розглядатися в наступних дослідженнях.

Список літератури: 1. Афанасьев В.В., Лебедеко Е.В. Графоаналитическая модель процесса формирования глобальной схемы мультибазы данных с учетом этапов ее реструктуризации // Информационные системы и технологии. 2014. № 1 (81). С 12-18. 2. Білова Т.Г. Проектування розподіленої бази даних системи надання електро-них адміністративних послуг // АСУ та прибори автоматики. Харків, 2019. Вип. 176. С. 49-54. № 175. С. 49-53. 3. Лаврентьев К.А., Пономарчук Ю.В., Фалеева Е.В. Применение ориентированных графов для построения схемы гетерогенной распределенной базы данных уровня предприятия // Фундаментальные исследования. 2015. № 6 (ч. 2). С. 257-262. 4. Кашиников А, Лядова Л. Интеграция гетерогенных источников данных на основе рекурсивной декомпозиции // International Journal "Information Technologies & Knowledge". 2011. Number 3, Vol.5. P. 274-284. 5. Дейт К. Дж. Введение в системы баз данных = An Introduction to Database Systems. 7-е изд. "Вильямс", 2001. 1072 с.

Надійшла до редколегії 18.06.2021

Білова Тетяна Георгіївна, кандидат технічних наук, доцент, доцент кафедри системотехніки ХНУРЕ. Наукові інтереси: моделювання бізнес-процесів, хмарні технології, розподілені бази даних. Адреса: Україна, м. Харків, пр. Науки 14, тел. (057) 702 14 51.

Дьоміна Вікторія Михайлівна, кандидат технічних наук, доцент, доцент кафедри інформаційних технологій, консалтинг і туризму Харківського національного аграрного університету ім. В. В. Докучаєва. Наукові інтереси: моделювання процесів стратегічного управління регіоном, математично-статистичні методи і моделі. Адреса: Україна, Харківська обл., селище Докучаєвське, уч. міс. ХНАУ, тел. (057) 709 03 01.

Побіженко Ірина Олександрівна, кандидат технічних наук, доцент, доцент кафедри інформаційних технологій Харківської державної академії культури. Наукові інтереси: моделювання бізнес-процесів, хмарні технології, розподілені бази даних, бізнес аналіз. Адреса: м. Харків, Бурсацький узвіз, 4, тел. (057) 731 32 82.