

ПРОБЛЕМИ ОПЕРАТИВНОГО АНАЛІЗУ В ГІПЕРВЕЛИКИХ БАНКІВСЬКИХ СУБД

Сергій Піппа, Яків Смолій

Сергій Петрович Піппа – д.е.н., завідувач відділу інформаційних технологій Академії Державної податкової служби України, науково-дослідний центр, 08201, Україна, м. Ірпінь, вул. К. Маркса, 31, e-mail: pippa_serg@yahoo.com

Яків Васильович Смолій – Заступник голови правління Акціонерного Поштово-пенсійного банку “АВАЛЬ”, Україна, 01011, Київ, вул. Лескова, 9, e-mail: smolij@head.aval.kiev.ua

Анотація: дана праця представляє підходи до конструювання систем оперативного аналізу в банківській галузі на основі надвеликих систем управління базами даних (НВСУБД). Розглядається теоретичний і практичний аспект моделювання сховищ даних у фінансовому обліку зі сторони специфіки. Пропонуються методи комбінування концептуальних та релятивних моделей баз даних (БД) в технологіях оперативного аналізу створення сховищ даних (СД) та баз знань.

Ключові слова: оперативний аналіз, сховища даних, концептуальні та реляційні моделі БД.

1. ВСТУП

Інтенсивне впровадження OLAP-технологій крім переваг при складному фінансовому менеджменті (інвестиційний аналіз, стратегічне планування, оцінка ризиків у фінансових розрахунках, прийняття рішень, експертний аналіз) продемонструвало і деякі, дуже істотні проблеми. Найчастіше це – нестача ресурсів пам'яті і складності концептуального моделювання БД навіть для дуже потужних обчислювальних систем і надпотужних СУБД. Спроби систематизувати технології оперативного аналізу в обчислювальних системах (створення так званих FASMi-тестів і т.п.) дозволили не тільки класифікувати вимоги до аналітичних інструментів, але й сформулювати ключові напрямки технологій аналітичної обробки даних із застосуванням надвеликих СУБД для багатомірного аналізу. Найбільш проблемними, як показують дослідження, виявилися два напрямки:

– теоретичні та практичні аспекти моделювання розподіленого доступу в надвеликих БД, де режими багатокористувацького використання даних поєднуються з підвищеними вимогами контролю доступу і захисту інформації;

– підтримка багатоаспектності та багатомірності як основної динамічної характеристики оперативного аналізу в надвеликих розподілених БД.

2. МЕТОДОЛОГІЯ РОЗВ'ЯЗКУ

В архітектурі систем і процесів аналізу з однієї сторони забезпечується ефект надшвидких розрахунків на основі заздалегідь визначених

підсумкових і агрегатних показників багатоаспектного і багатомірного представлення БД. З іншої сторони виникає небезпека перевитрати пам'яті для таких надлишково згенерованих систем даних у виді багатомірних кубів. У БД оперативного аналізу координатами багатомірних даних виступають звичайні ознаки банківських бізнес-процесів – види платіжних операцій, валют, типи клієнтів, різні дати оцінки платежів, аналітичні показники курсів валют і т.п. Проблема полягає в тому, що генерація великої кількості можливих підсумкових і агрегатних показників крім потенційної економії обчислювальних ресурсів значно погіршує не тільки наочність аналітичних даних, але й можливість ефективної орієнтації та навігації в сховищах даних.

Багатомірність на рівні нижніх вузлів розподіленої інформаційної банківської системи (відділень, філій) ще не виявляється в критичному вигляді, однак при підвищенні рівня інтеграції – районні, обласні та центральні офіси банків, відбувається непропорційне збільшення вимог СУБД до пам'яті та обчислювальних ресурсів. Синтез агрегатів приводить до значного збільшення обсягів БД, відбувається також погіршення характеристик цілісності БД, оскільки в алгоритмах багатомірного і багатоаспектного оперативного аналізу розробники програмного забезпечення досить часто використовують алгоритми денормалізації гіпервеликих БД.

Таким чином підвищується ймовірність неадекватної реалізації процесу денормалізації в зв'язку з тим, що розміщувана в сховищі

інформація не може містити в собі всі можливі представлення майбутніх аналітичних запитів навіть для однорівневої агрегації.

Процедура денормалізації істотно знижує ефективність технологій оперативного аналізу. Фактично відступ від нормальних форм для реляційної DB охоплює випадки багатомірних і гібридних моделей даних. Режими підтримки таких структур у DW для технологій оперативного аналізу здатні змінити результати тонкого фінансового аналізу, інвестиційних розрахунків з точністю до зворотного результату. При обчисленні показників фінансового ринку, оцінці фінансових активів фондового ринку, дисконтних перетвореннях проявляються проблеми нееквівалентності математичних результатів і їхньої економічної інтерпретації для різних моделей аналітичних даних. Тобто вибір різних сполучень реляційних, багатомірних і гібридних структур у OLAP-технологіях може привести, наприклад, до діаметрально протилежних висновків для однакових інвестиційних проектів. Ще одна проблема виникає при необхідності обліку факторів інфляції, що особливо актуально для умов перехідної економіки. Експонентні методи, що лежать в основі актуарних розрахунків, орієнтовані у своїй більшості на нормативні економіки (низькі відсотки, помірні інфляція, висока кон'юнктура). Їхнє застосування в ненормативних умовах фінансових ринків Східно-європейських країн також приводить до істотних помилок у технологіях традиційного оперативного аналізу для DW. Пошук і добір методів для ліквідації небажаних ефектів DW-додатків, для підвищення точності та економії комп'ютерних ресурсів відіграють важливу роль у розвитку технологій VL DBMS.

Ефективне сховище даних – це завжди компроміс збереження в мінімальному наборі нормалізованих таблиць з характеристиками економічних об'єктів і процесів (у нашому випадку – банківських) максимальної кількості заздалегідь підготовлених можливих вимірів для оперативного аналізу. У реляційній базі даних, де фактично втрачається можливість наочного узагальнення складних моделей даних, навіть на теоретичному рівні досить важко представити деякі обмеження чи залежності для додаткових вимог цілісності. Перехід до концептуальної моделі в таких випадках теж утруднений, оскільки повертаючись до розподілених систем, часто буває важко визначити, які групи даних (аспекти, виміри) потраплять у загальне представлення аналітичної БД.

Конструктивним виходом у такій ситуації може бути спеціальна техніка представлення додаткових моделей сполучення концептуальних і реляційних схем БД у OLAP-процесах.

Ефект застосування додаткових моделей ґрунтується на розширенні стандартної концепції моделювання і адміністрування розподілених баз даних у теорії реляційних моделей. Необхідність розширення поняття розподіленої бази даних і процедур обчислення запитів у таких структурах зв'язана зі специфікою представлення реляційних моделей, що орієнтовані на OLAP-технології і сховища даних. Формулювання запиту в розподіленій базі даних $K \leftrightarrow W_{(e=f)}(L_1, L_2, \dots, L_n)$ де цільове відношення K визначається через операцію об'єднання (\leftrightarrow) безлічі відносин (L_1, L_2, \dots, L_n) розподіленої структури даних по деякій умові ($e=f$). Порушення стандартної ситуації обчислення запиту в розподіленій системі відбувається, як у формулюванні умови, так і у встановленні відносин розподіленої бази даних. Характерно, що проблеми застосування традиційного апарата реалізації запитів у розподілених системах можуть виникнути ще на попередній стадії – при плануванні його виконання.

3. ОСНОВНІ РЕЗУЛЬТАТИ

Перший випадок – якщо в умові e елементи з розподіленої БД. Тоді необхідне виконання стандартної процедури обчислення запиту в розподіленій структурі даних і для самої умови. Складність характеризується тим, що обчислення в умові розподіленого випадку може вплинути на наступний процес виконання основного запиту. Теоретична можливість реалізації таких обчислень для умов стикається зі значними труднощами в контексті застосування оперативно-аналітичних даних.

Інший випадок відступу від стандартної ситуації відбувається щодо безлічі відносин в розподіленій БД. Проблема даного випадку зв'язана з необхідністю розгляду множинних переміщень умов і самих відносин між вузлами розподіленої бази даних. Принципово задача раціонального переміщення між структурами даних у розподіленій системі може бути вирішена, однак при допущенні розбіжності розміщень умов і вихідних відношень уже досить складно знайти прийнятний вихід.

Однак, у плані описаних маніпуляцій з розподіленою базою даних, дійсні складності для традиційної техніки застосування напівз'єднань [1, с.346-359] (як основного опера-

ційного апарату в розподілених системах) починаються при додаванні семантичних компонентів для аналітичних операцій зі сховищами даних. Навіть потенційний ввід випадків рекурсивних запитів і предикатних конструкцій (наприклад, у системах підтримки прийняття рішень, експертних системах, застосуваннях баз знань) значно ускладнює формальний апарат модифікації запитів теорії реляційних баз даних. В приведених випадках виникає необхідність пошуку нових інструментальних засобів чи істотного розширення традиційного теоретичного апарату напівз'єднань, їхніх послідовностей і повної редукції.

Для забезпечення можливостей оперативного аналізу накопичених даних в умовах розподілених баз даних при виконанні запитів передбачаються спеціальні процедури, що забезпечують трансформацію концептуальних та реляційних моделей VL DB. Такі трансформаційні підходи в OLAP-аналізі дають ефект при створенні систем підтримки прийняття рішень, експертних систем і баз знань у розподіленому середовищі. Сховища даних таких розподілених систем забезпечують всі основні вимоги сучасних аналітичних інструментів: а) предметна орієнтованість; б) інтегрованість; в) прив'язка в часі і г) незмінюваність.

Інформація в сховищі даних розподіленої DB організована відповідно до загальноприйнятих аспектів банківської діяльності. Однак деякі з них можуть або бути віртуальними, або мати віртуальні значення (для значень, що розміщені в різних пунктах розподіленої DB, відмінних від цільової DB). Предметна організація в сховищі даних розподіленої системи сприяє як значному спрощенню оперативного аналізу, так і підвищенню швидкості виконання запитів при переміщенні даних між сегментами розподіленої DB.

Традиційні для сховищ структури моделей даних в умовах розподіленості (наприклад, моделі “stars” і “snowflakes”) виявляються не настільки ефективними. Проблема полягає в тому, що зазначені структури є незручними при переміщенні між ними даних умов або самих відносин з погляду архітектури і топології мережі. Більш ефективний доступ до розподіленої DB забезпечують структури даних, що враховують фактор розподіленості. Найкращі результати демонструють, за результатами теоретичних досліджень та апробації, трансформаційні структури даних. Вони реалізують можливості попереднього аналізу концептуальної моделі та відкривають шляхи для наступного вибору раціональних сполучень реляційних схем у розподіленій DB. У випадку

використання багатомірних і гібридних моделей механізм концептуально-реляційної трансформації розширюється операціями прискореної навігації в сховищі даних.

Властивість інтегрованості сховища даних засновано на операціях попередньої агрегації. Фактично первинні дані які видобуваються з оперативних DB, перевіряються, очищаються і приводяться до універсального вигляду. Після чого виконуються операції агрегування та обчислення необхідних зведених показників, що, приблизно, на 75% і складають зміст сховища даних для оперативного аналізу. Облік структури розподіленої DB також виконується на цьому етапі, оскільки агрегування може бути зв'язане з виконанням переміщення деяких даних між сегментами DB. У даному випадку застосовується “подвійна” інтеграція: традиційна – для полегшення оперативного аналізу, і “поглиблена” – для пошуку оптимальних схем одержання зведених і агрегатних показників у розподіленій системі. Механізм оптимізації для одержання інтеграції у VL DB базуються на об'єднанні умов і відносин у процесі навігації для виключення зайвих повторень пошуку і переміщень у розподіленій мережі.

Прив'язка вчасно, як специфічна властивість сховищ даних, також має свої особливості в підтримці VL DB. Добір інформації з оперативних масивів здійснюється у виді “історичних шарів” і розміщується в сховищі стосовно до конкретних моментів часу. Властивість розподіленості найчастіше виявляється як ймовірність одержання тенденцій і тимчасових оцінок з неповним охопленням чи появою можливості різноманітних “історичних шарів”. Забезпечити адекватну прив'язку до часу в розподіленій системі можна з використанням механізмів, аналогічних для розглянутої вище техніки інтегрованості.

Наступне важлива властивість сховищ даних – незмінюваність. На даний момент забезпечення незмінюваності даних розподіленої системи є відкритою проблемою. Складність полягає в необхідності використання процедур агрегації й узагальнення інформації, відмінних від стандартних підходів у сховищах даних. Можливий вибір декількох варіантів – незмінюваність на тривалих історичних етапах за рахунок динаміки середньострокових і короткострокових шарів. Інші варіанти представляються сполученням приведених вище видів тривалості. За результатами останніх досліджень найкращі результати в забезпеченні незмінюваності показали методи диферен-

ційованих підходів до різних інструментів фінансово-кредитних ринків.

4. ВИСНОВКИ

Апробація запропонованого розширення понять розподілених БД для поєднання концептуальних і реляційних моделей була проведена як практична реалізація OLAP-розрахунків у банку “Аваль” на основі систем СУБД Oracle і MS SQL-Server. Її використання показало реальну застосовність описаних вище підходів у розподілених системах навіть в умовах досить нестабільних фінансових ринків і трансформаційних процесів перехідної економіки. Додатковий ефект проявився також завдяки можливостям системної інтеграції різнорідних програмних продуктів СУБД на основі реалізованої моделі концептуальних і реляційних інструментів для підтримки надвеликих інформаційних систем. Досягнута також домовленість між АППБ “Аваль” і науково-дослідним центром Академії ДПС України про розширення

теоретичних і практичних робіт в області поєднання концептуальних і реляційних моделей БД при створенні фінансово-аналітичних баз даних і знань. При цьому найбільш прийнятними областями для спільних робіт в області оперативно-аналітичних технологій і розподілених систем намічені напрямки по створенню інформаційних серверів для експертних і консультаційних послуг в Інтернеті, а також розгортання корпоративних мереж для перепідготовки і підвищення кваліфікації фахівців в області фінансів, банківських послуг і податків.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Мейер Д. *Теория реляционных баз данных: Пер. с англ.* – М.: Мир, 1987. – 608с.
2. A.Sachenko, S.Rippa, J.Smolij. *HPC for Financial Management of the Ukrainian banking System. Advanced Research workshop on High Performance Computing: Technology and Application, Cetraro (Italy) – June 12-15, 2000, p.16.*



Ріппа Сергій Петрович, 1957 року народження, росіянин. Закінчив у 1979 році факультет механізованої обробки економічної інформації Ростовського-на-Дону інституту народного господарства за спеціальністю “Організація механізованої обробки економічної інформації”. Кандидат економічних наук з 1985 року. Доктор економічних наук з 1999 року. Стажувався в університетах ФРН Констанс, Регенсбург в 1989-1990 роках, Німецькій академії менеджменту Нижньої Саксонії, м. Целле; навчальному центрі Торгово-промислової палати, м. Мюнхен 1994-1995 рр. Працює на посаді професора кафедри інформаційних систем і технологій Академії державної податкової служби України. Викладає дисципліни: “Інформаційне забезпечення ДПС України”, “Автоматизація роботи в органах ДПС”, “Адміністрування баз даних Oracle”, “Адміністрування БД Державного реєстру фізичних осіб”.

Яків Васильович Смолій, 1961 року народження, освіта вища математична. Заступник голови правління банку “АВАЛЬ”. Відповідає за роботу операційного управління та розвиток інформаційних технологій.

