

УДК 004.891

О.В. Огнєвий,

кандидат технічних наук, доцент

А.М. Огнева,

кандидат технічних наук, доцент

Д.В. Зайцев,

кандидат військових наук

О.В. Банзак,

кандидат технічних наук, доцент

МЕТОДИ ОРГАНІЗАЦІЇ ІНФОРМАЦІЇ В ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМАХ

У статті розглянуті задачі оптимального розподілу інформаційних ресурсів, що виникають під час проектування та адміністрування розподілених баз даних комп'ютерних інформаційно-аналітичних систем. Розроблено методiku розв'язання задачі оптимального розподілу файлів серед вузлів обчислювальної мережі територіально розподіленої системи підприємства.

Ключові слова: база даних, інформаційно-аналітична система, інтегральний показник, інформаційні ресурси, модель, моніторинг.

В статье рассмотрены задачи оптимального распределения информационных ресурсов, возникающие во время проектирования и администрирования распределенных баз данных компьютерных информационно-аналитических систем. Разработана методика решения задачи оптимального распределения файлов среди узлов вычислительной сети территориально распределенной системы предприятия.

Ключевые слова: база данных, информационно-аналитическая система, интегральный показатель, информационные ресурсы, модель, мониторинг.

Paper considers the issue of the optimal allocation of information resources that arise during the design and administration of distributed databases of computer informational analytical systems. The methodology of solving the problem of optimal file allocation amongst computing networks nodes of geographically distributed enterprise system is developed.

Keywords: database, information-analytical system, integral index, information resources, model, monitoring.

Вступ та постановка проблеми. Інформація є важливим ресурсом у господарській діяльності підприємств. Використання інформаційних технологій, персональних комп'ютерів зумовило принципове перетворення відносин і технологічних основ діяльності у сфері економіки. Це відбувається за рахунок поєднання трьох взаємозв'язаних процесів – комп'ютеризації, вдосконалення технологій збору, накопичення і використання інформації та її інтелектуалізації.

Управлінські структури як суб'єкти системи інформаційного забезпечення постійно взаємодіють із інформаційним середовищем, регулюють рух інформації, аналізують її й на цій основі розробляють рекомендації для прийняття управ-

лінських рішень, удосконалювання оптимального управлінського впливу для досягнення цілей керування.

Інформаційно-аналітична підтримка необхідна при виконанні наступних завдань:

- моніторинг стану об'єкта управління;
- контроль за виконанням рішень і ефективністю виконавчих механізмів;
- аналіз зовнішніх і внутрішніх проблемних ситуацій і прогнозування їхнього розвитку;
- процес підготовки управлінських рішень і їх документальне оформлення;
- аналіз обґрунтованості і юридичної правомірності прийнятих рішень;
- функцій управління діяльністю самого органу управління.

Функції автоматизованої системи інформаційно-аналітичного забезпечення визначаються можливостями інформаційних технологій і засобів комп'ютерної техніки, основними з яких є:

- надійне зберігання й оперативний доступ до великих обсягів документальної й довідкової інформації;
- автоматизована підтримка складних процедур обробки інформації;
- ретроспективний аналіз, моделювання, прогнозування, експертна оцінка й т.п.;
- підтримка індивідуальної й колективної роботи з інформацією.

Сучасні інформаційно-аналітичні системи (далі – ІАС), які реалізують наведені функції, – це експертні системи реального часу, системи підтримки прийняття управлінських рішень, інтелектуальні аналітико-інструментальні пакети тощо. Ефективність цих систем залежить від наповненості та адекватності сховища даних, на основі яких вони організовані. Це відображено в появі комерційних продуктів, які включають три основні процеси: поповнення сховища даних з незалежних транзакційних систем; збереження даних і керування ними; аналіз даних з метою прийняття обґрунтованих бізнес-рішень.

Незважаючи на велику кількість програмного інструментарію, ці напрямки залишаються важливими для дослідження та пов'язані з основними компонентами системи підтримки прийняття рішень: оперативна аналітична обробка і пошук даних та базові інструменти для заповнення сховища даних.

Аналіз останніх досліджень і публікацій. При впровадженні ІАС на підприємствах з розвинутою інфраструктурою виникають проблеми, що пов'язані з територіальною розосередженістю таких підприємств. Перспективним є впровадження розподілених баз даних (РБД) – набору автономних обчислювальних вузлів, що сполучені мережею передач даних, вузли якої спроможні взаємодіяти між собою, забезпечуючи користувачу доступ до інформаційного простору [1].

Великий внесок у дослідження теоретичних і методологічних питань створення і впровадження ІАС внесли радянські і закордонні вчені: Бутинець Ф.Ф., Глушков В.М., Івахненко С.В., Подольський В.І., Рожнов В.С., Ходаков В.Є. та інші.

Особливо значні результати досягнуті у сфері створення локальних баз даних і застосування різноманітних систем керування базами даних. Проте низка проблем на цей час залишається невирішеною.

Останнім часом у світі оформився ряд нових концепцій зберігання та аналізу корпоративних даних:

- Сховища даних, або Склади даних (Data Warehouse);

- Оперативна аналітична обробка (On-Line Analytical Processing, OLAP);
- Інтелектуальний аналіз даних – ІАД (Data Mining).

ІАС – це сучасний високоефективний інструмент підтримки прийняття стратегічних, тактичних і оперативних управлінських рішень на основі наочного та оперативного надання всієї необхідної сукупності даних користувачам, відповідальним за аналіз стану справ та прийняття управлінських рішень. Основне призначення ІАС – динамічне подання та багатовимірний аналіз історичних та поточних даних, аналіз тенденцій, моделювання і прогнозування результатів різних управлінських рішень [4].

Основними функціями інформаційно-аналітичної системи є: витяг даних з різних джерел, їх перетворення і завантаження в сховище; зберігання даних; аналіз даних, в тому числі оперативний та інтелектуальний; підготовка результатів оперативного та інтелектуального аналізу для ефективного їх сприйняття споживачами.

ІАС є надбудовою над уже функціонуючими на підприємстві інформаційними додатками і не вимагають їх заміни; вони акумулюють дані по всіх видах діяльності підприємства. На корпоративних підприємствах найчастіше функціонує декілька оперативних інформаційних систем, що автоматизують щоденну діяльність підприємства. Ці системи виступають у ролі джерел даних для сховищ даних (СД) та можуть бути реалізовані різними способами: розміщуватися на центральній обчислювальній машині, бути додатком з архітектурою клієнт/сервер, належати до прикладних систем третіх фірм із власними сховищами інформації або бути комбінацією перерахованих варіантів. Завдання полягає в тім, щоб ідентифікувати джерела даних і доставляти з них інформацію в СД відповідно до продуманого розкладу. Задача збору інформації в СД із найрізноманітніших джерел даних називається придбанням даних.

Мета статті. Дослідження методології побудови багаторівневих територіально розподілених баз даних, а також узагальнення методології створення ефективних структур баз даних у сучасних інформаційно-аналітичних системах.

Виклад основного матеріалу досліджень. Проектування архітектури розподіленої БД проводиться на стадії логічного проектування, на якому здійснюється перетворення інтегрованої концептуальної схеми ІАС у схему БД СУБД визначеного типу (реляційну, об'єктно-орієнтовану). Вхідною інформацією для цього служить специфікація даних (глобальна схема БД) і прикладних задач. Аналіз необхідного характеру розподіленості БД здійснюється на основі таблиць фрагментації. У результаті фрагментації одержуємо можливі розміщувані одиниці БД [3].

Ефективність доступу користувачів у значній мірі визначається організацією розподіленої бази даних (РБД). Однією з істотних проблем, що виникають при створенні й експлуатації РБД, є оптимальний, з погляду обробки запитів користувачів, розподіл файлів даних по вузлах обчислювальної мережі. Розв'язанню цієї проблеми присвячений ряд робіт, що відрізняються як постановкою задач, так і підходами до їх рішення [1]. Однак вони не дозволяють будувати моделі оптимального розподілу БД по вузлах обчислювальної мережі з довільною топологією. В основі більшості робіт лежить математична теорія черг, що не дозволяє будувати реальні моделі оптимального розподілу файлів по вузлах обчислювальної мережі зі складною топологією.

Однією з найбільш важливих під час проектування та адміністрування розподілених комп'ютерних інформаційних систем є задача оптимального розподілу інформаційних ресурсів, яку ми розглядаємо як задачу оптимального розподілу файлів серед вузлів обчислювальних мереж [3].

У процесі функціонування ІАС запити користувачів розділяють спільні ресурси мережі: обчислювальні потужності, канали зв'язку, бази даних тощо. Як наслідок, під час виконання запиту певного користувача час повного виконання запиту перевищує мінімальний час повного виконання запиту у випадку, коли б цей запит міг використовувати всі ресурси системи в монопольному режимі, тобто виникають затримки, спричинені спільним використанням ресурсів мережі. Крім того, при переповненні буферів каналів зв'язку чи надмірній завантаженості обчислювальних потужностей деякі запити можуть бути взагалі не оброблені. Середній час виконання запитів користувачів визначає якість обслуговування мережі. При фіксованій інтенсивності запитів від користувачів та обчислювальних потужностях мережі якість обслуговування залежить переважно від способу розташування файлів даних та програм серед вузлів мережі. При оптимізації інформаційних ресурсів мережі входними даними є запити прикладних програм на отримання інформації, що розташована в файлах (сайтах) вузлів мережі. Отже, варіанти розміщення файлів можуть бути вибрані як керовані змінні.

Розглянемо методику розподілу інформаційних ресурсів серед вузлів розподілених обчислювальних мереж. Постає задача оптимально розмістити файли серед вузлів мережі, щоб отримати найвищу якість обслуговування, для цього необхідно створити програмний додаток, назовемо його програмним агентом, який буде виконувати функції моніторингу (контролювати середовище, в якому він діє) та адміністрування (резервне копіювання, пошук та обмін інформацією з користувачем).

Якість обслуговування мережі є інтегральним показником, складові якого можуть змінюватись залежно від вимог, що ставляться до мережі. Одним з основних його складників є середній час виконання запитів. Отже, необхідно на основі побудованих математичних моделей розробити методику оптимального розподілу інформаційних ресурсів серед вузлів розподілених обчислювальних мереж, яка не залежить від схеми обробки запитів у мережі.

При розробці інформаційної РБД підприємства вирішальне значення має поетапний аналіз і оцінка ефективності оптимального розподілу інформаційних ресурсів серед вузлів обчислювальної мережі. Ці оцінки формуються під час проектування та адміністрування розподілених комп'ютерних інформаційних систем.

Розглянемо основні етапи моніторингу й адміністрування РБД.

1 етап: формування комплексу вихідних даних, які одержує програмний додаток при роботі ІАС. Для проведення дослідження необхідно визначити такі характеристики системи:

T – середній час реакції системи на запити користувачів,

$\mu_i(t)$ – інтенсивність передачі даних по каналах зв'язку;

Q_i – загальна кількість запитів, які надходять за одиницю часу;

n – кількість сайтів, що беруть участь в обробці запиту.

Сукупність параметрів $\{n, \mu_i(t), Q_i, T\}$ є початковою інформацією для послідовності алгоритмів, використовуваних ПА при адмініструванні РБД. На

цьому етапі для прогнозу стану РБД використовується інформація, отримана при проектуванні інтегрованої концептуальної схеми бази даних. Ці параметри є граничними і відображають максимально необхідний час на обробку запитів. У такий спосіб виконується перша стадія адміністрування РБД – визначення сукупності вхідних характеристик.

2 етап: обчислення отриманих в процесі експлуатації ІАС параметрів. Для проведення дослідження необхідно мати статистичні оцінки числових характеристик дискретних розподілів звертань до сайтів та їх обслуговування:

Q_i – кількість запитів до файлів цього сайту;

S_i – кількість звертань, обслуговані власними файлами сайту;

M_i – звертання, не було обслуговане сайтом та переадресоване на інший сайт.

Зміни параметрів можна описати сімейством матриць

$$X(t_n) = \begin{matrix} & \begin{matrix} Q & S & M \end{matrix} \\ \begin{matrix} \vdots \\ x_{i1}(t_n) \\ \vdots \\ x_{n1}(t_n) \end{matrix} & \begin{bmatrix} x_{11}(t_n) & x_{12}(t_n) & x_{13}(t_n) \\ \dots & \dots & \dots \\ x_{i1}(t_n) & x_{i2}(t_n) & x_{i3}(t_n) \\ \dots & \dots & \dots \\ x_{n1}(t_n) & x_{n2}(t_n) & x_{n3}(t_n) \end{bmatrix} \end{matrix}$$

Матриця $X(t_n)$ є функцією часу t_n , де n змінюється від 1 до N (кількість сайтів). Для будь-якого n розмір матриці є рівним $t_n \times 3$, таким чином кількість рядків залежить від часу, а стовпцями є значення q_i, n_i, m_i .

3 етап: аналіз траєкторії змін параметрів: порівняння обох варіантів отриманих даних. Розрахована траєкторія змін порівнюється з параметрами, визначеними попередньо, і використовується при прогнозуванні стану РБД.

РБД має певні можливості для ефективної обробки запитів. Надлишковий запас можливостей приведе до втрат і зниження економічної ефективності системи. Якщо прийняти, що B_i і B_T відповідно вхідні та параметри одержані при експлуатації системи, то запас можливостей визначається $Z = B_s - B_i$.

При цьому:

$Z = 0$, можливості ІДБП використовуються цілком;

$Z > 0$, запас можливостей надмірний;

$Z < 0$, можливостей ІДБП недостатньо для ефективної обробки запитів.

Визначимо коефіцієнт використання можливостей РБД як $k = B_i / B_T$. Чим ближче значення k до одиниці, тим більш повно використовуються можливості ІДБП. Якщо $k=1$, можливості ІДБП використовуються цілком. У зв'язку з цим, необхідно постійно аналізувати і приймати відповідні рішення для підвищення ефективності виконання запитів у ланцюжку $\{\mu_i(t), Q_i\} \rightarrow \{T\}$.

Мінімізація загального часу виконання запитів обумовлюється критерієм $\tau = \max\{1/\{\mu_i(t)\} \rightarrow \min$, особливістю якого є можливість визначити область допустимих значень величини ϕ .

4 етап. На основі розроблених математичних моделей оптимального розміщення інформації в РБД здійснюється наступна стадія – власне переміщення фрагментів даних з урахуванням їх взаємодії.

Таким чином, основні етапи, що використовуються при проведенні моніторингу й адмініструванні РБД, мають таку послідовність:

- визначення сайтів, що беруть участь в обробці запитів;
- виділення зв'язків на основі матриці зв'язків обробки запитів з номером сайту, з використанням двосторонніх графів;
- оптимізаційне проектування зв'язків на основі критеріїв: максимум часу обробки запитів; правил фрагментації даних; оптимального числа копій даних;
- дослідження стабільності ІРБП і визначення траєкторії змін динамічних параметрів.

Висновки. Процес моніторингу повинен підтримуватися протягом усього життя ІАС, що дозволить у будь-який момент часу провести ефективну реорганізацію РБД. Подібні зміни надають інформацію про найбільш ймовірну еволюцію системи і ресурси, що можуть знадобитися в майбутньому.

Розроблена методика оптимального розміщення файлів серед вузлів мережі. Як критерій оцінки обслуговування запитів використовуються: загальний середній час відповіді на запит по всій розподіленій обчислювальній системі та обсяг даних, що будуть переслані в результаті функціонування системи протягом одиниці часу.

Представлені етапи обслуговування запитів оптимізації обробки розподілених запитів у локальній мережі інформаційного зв'язку між структурними підрозділами підприємства та результати, отримані з використанням розглянутих підходів, що застосовуються при адмініструванні РБД.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Данілін А.В. Огляд технологій інтеграції інформаційних систем / А.В. Данілін [Електронний ресурс]. – Режим доступу : <http://www.microsoft.com/Ukraine/Government/Analytics/IntegrationTechnologies/Overview.aspx>.
2. Нецадим Н.О. Розробка схеми консолідації інформації на великому підприємстві / Н.О. Нецадим // Наукові праці НУХТ. – 2011. – № 40. – С. 180–188.
3. Огневий О.В. Проблеми автоматизації проектування структур розподілених баз даних / О.В. Огневий, А.М. Огнева // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2001. – № 3.
4. Субботін С.О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навчальний посібник / С.О. Субботін. – Запоріжжя : ЗНТУ, 2008. – 341 с.

Отримано 14.06.2016

Рецензент Рибальський О.В., д.т.н.