

МЕТОДИ ОРГАНІЗАЦІЇ І ПОШУКУ 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 методику розв'язку задачі оптимального розподілу файлів серед вузлів обчислювальної мережі територіально розосередженої системи підприємства.

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

OLEG ISHCHUK, OLEKSANDR OGNEVOY
Khmelnytskyi National University

DEVELOPMENT OF FRAME AND TECHNIQUE OF CONSTRUCTION OF THEMATIC REPOSITORIES

In the article the frame decision support systems (DSS) is considered, which one maps the basic correlations functional and subsystems at appropriate levels of management. The basic parameters of business-processes of the enterprise are determined. On the basis of it definite reference directions, which one decides DSS. The problems of creation of thematic repositories lighted, which one are used in DSS for preservation of the information. Designed conceptual model of repository.

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

Вступ

Зусилля комерційних компаній і наукових кіл привели до серйозного технологічного прогресу в рішенні задач збереження даних. Це знайшло відображення в безлічі комерційних продуктів, що доступні для кожної з трьох основних операцій: поповнення сховища даних з незалежних транзакційних систем; збереження даних і керування ними; аналіз даних з метою прийняття обґрунтованих бізнес-рішень. Однак, незважаючи на достатню кількість комерційного інструментарію, ці напрямки залишаються важливими для дослідження та пов'язані з двома основними компонентами системи підтримки прийняття рішень: інструментарій оперативної аналітичної обробки та пошуку даних та базові інструменти для заповнення сховища даних.

Аналіз останніх досліджень і публікацій

Еволюція засобів обчислювальної техніки, застосовуваних в управлінні економікою, спрямована на створення систем, які реалізують функції штучного інтелекту: експертні системи реального часу, системи підтримки прийняття управлінських рішень, інтелектуальні аналітико-інструментальні пакети, тощо. Ефективність цих систем залежить від наповненості та адекватності баз знань на основі яких вони організовані.

При впровадженні комп'ютерної інформаційно-аналітичної системи (ІАС) на корпоративних підприємствах з розвинутою інфраструктурою виникають проблеми, що пов'язані з територіальною розосередженістю таких підприємств - віддалені об'єкти, включені в єдину систему, повинні мати можливість корегувати дані і одержувати зміни, внесені на інших об'єктах. Перспективним варіантом є впровадження розподілених баз даних (РБД) - набору автономних обчислювальних вузлів, що сполучені мережею передачі даних, вузли якої спроможні взаємодіяти між собою, забезпечуючи користувачу послуги системи керування БД. [1]

Серед проблем проектування, РБД можна виділити такі:

- формування загальної методології синтезу структури РБД (визначення глобального критерію синтезу, етапів і задач проектування і методології їх рішення);
- формування ефективної інформаційної структури РБД (синтез загальної логічної структури бази даних, розміщення бази, вибір дисципліни опрацювання запитів користувачів);
- створення програмного забезпечення РБД (розробка системи керування розподіленою базою даних, програм опрацювання запитів користувачів, програм ефективної маршрутизації повідомлень);
- розробка технічного забезпечення (визначення структури мережі ЕОМ, вибір типу базової ЕОМ, визначення числа розміщення обчислювальних комплексів у мережі ЕОМ, числа і типів каналів зв'язку, методів передачі даних).

Постановка задачі. Великий внесок у дослідження і розробку проблем створення і впровадження баз даних і систем керування базами даних внесли радянські і закордонні вчені. Особливо значні результати досягнуті в області створення локальних баз даних і застосування різноманітних систем керування базами даних. Проте ряд проблем у даний час залишається невирішеним [2]:

- розробка загальної методології побудови багаторівневих територіально розподілених баз даних,
- створення систем керування розподіленими базами даних, розробка загальної методології і комплексу математичних моделей створення ефективних (оптимальних) структур баз даних,
- створення діалогових систем автоматизованого проектування складних багаторівневих

територіально розподілених баз даних корпоративних підприємств та інформаційних систем різноманітних класів.

У сучасних корпоративних підприємствах функціонують декілька оперативних інформаційних систем, що автоматизують щоденну діяльність підприємства. Ці системи виступають у ролі джерел даних для ТСД, вони можуть бути реалізовані самими різними способами: розміщатися на центральній обчислювальній машині, або ж являти собою додаток з архітектурою клієнт/сервер, або відноситися до прикладних систем третіх фірм із власними сховищами інформації, а також являти собою комбінацію перерахованих варіантів реалізації. Завдання полягає в тім, щоб ідентифікувати джерела даних і доставляти з них інформацію в ТСД відповідно до продуманого розкладу. Задача збору інформації в сховище із самих різних джерел даних називається придбанням даних.

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 здійснюється на основі таблиць фрагментації. У результаті фрагментації одержуємо можливі розміщувані одиниці БД. [2]

Ефективність доступу користувачів у значній мірі визначається організацією розподіленої бази даних (РБД). Однієї 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 [3,4,5], що відрізняються як постановкою задач, так і підходами до їх рішення. 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 [6, 7]. В 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, як звичайно, 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 вхідними 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 в мережі.

При розробці інформаційної РБД підприємства (ІБДП) 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, що використовуються при моніторингу й адмініструванні ІБДП.

1 етап: формування комплексу вихідних даних, які одержує програмний додаток при роботі ІАС.

Для проведення дослідження необхідно визначити наступні 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 ІБДП. На цьому етапі для прогнозу стану ІБДП

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 даного 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 $\{X_n\} = \{X(t_n)\}$.

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

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

ІДБП 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_r , відповідно вхідні та параметри одержані при експлуатації системи, то запас можливостей визначається $Z = B_i - B_r$.

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 ІДБП як $k = B_r/B_i$. Тоді, чим ближче значення k до одиниці, тим більше повніше 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 в ланцюжку $\{\mu_i(t), Q_i\} \rightarrow \{T\}$.

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

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

Таким чином основні етапи (рис.1), що використовуються при проведенні моніторингу й адмініструванні ІБДП, 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 інформаційної системи, що дозволить у будь-який момент часу провести ефективну реорганізацію розподіленої бази даних по сайтах обчислювальної мережі. Подібні зміни надають інформацію про найбільш ймовірну еволюцію системи і ресурси, що можуть знадобитися в майбутньому.

Розроблена методика оптимального розміщення файлів серед вузлів мережі. В якості критерію 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 в локальній мережі інформаційного зв'язку між структурними підрозділами підприємства та результати, отримані з використанням розглянутих підходів, що застосовуються при адмініструванні ІБДП.

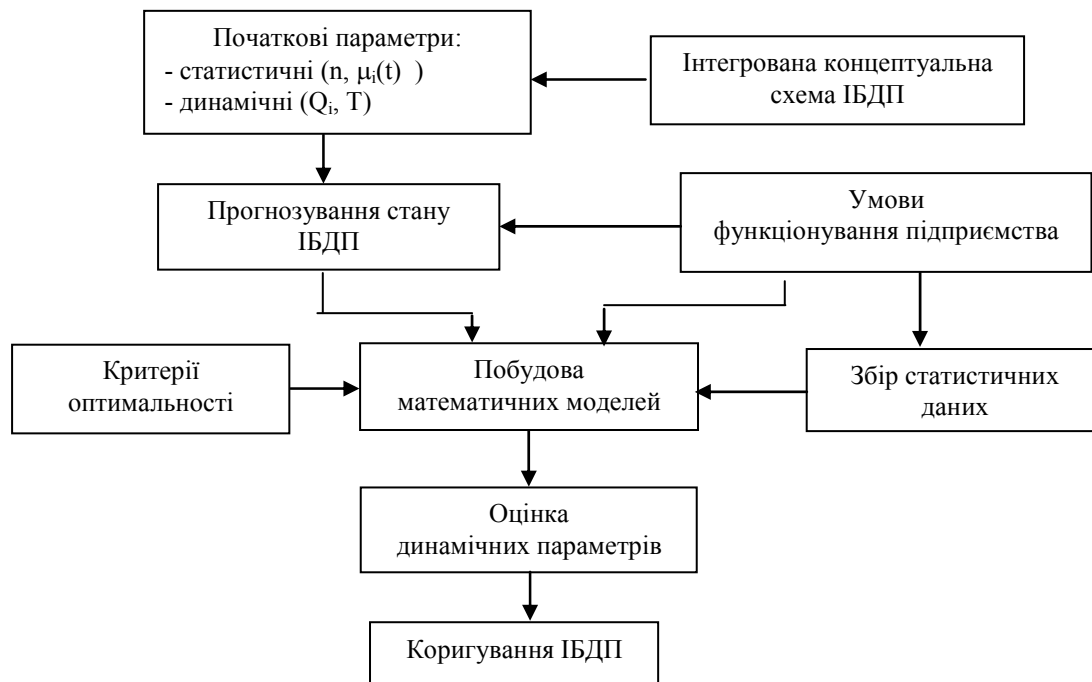


Рис.1. Послідовність етапів адміністрування РБД комп'ютерних інформаційних систем

Література

1. Данілін А. В. Огляд технологій інтеграції інформаційних систем/ Данілін А. В. // [Електронний ресурс]. – Режим доступу: [http://www.microsoft.com/Ukraine/Government/ Analytics/ IntegrationTechnologies / Overview.aspx](http://www.microsoft.com/Ukraine/Government/Analytics/IntegrationTechnologies/Overview.aspx).
2. Нешадим Н.О. Розробка схеми консолідації інформації на великому підприємстві/ Н.О. Нешадим// Наукові праці НУХТ. – 2011. – №40. – С. 180-188.
3. Огневий О.В. Проблеми автоматизації проектування структур розподілених баз даних/ О.В.Огневий О.В., 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

References

1. Danilin A. V. Ohliad tekhnolohii intehtatsii informatsiinykh system/ Danilin A. V. // [Elektronnyi resurs]. – Rezhym dostupu: [http://www.microsoft.com/Ukraine/Government/Analytics/IntegrationTechnologies /Overview.aspx](http://www.microsoft.com/Ukraine/Government/Analytics/IntegrationTechnologies/Overview.aspx).
2. 6 Neshchadym N.O. Rozrobka skhemy konsolidatsii informatsii na velykomu pidpriemstvi/ N.O. Neshchadym// Naukovi pratsi NUKhT. – 2011. – #40. – S. 180-188.
3. Ohnievyi O.V., Problemy avtomatyzatsii proektuvannia struktur rozpodilenykh baz danykh / O.V.Ohnievyi , A.M. Ohneva A.M. // Vymiriuvalna ta obchysliuvalna tekhnika v tekhnolohichnykh protsesakh-2001-#3.
4. Subbotin S. O. Podannia y obrobka znan u systemakh shtuchnoho intelektu ta pidtrymky pryiniattia rishen: Navchalnyi posibnyk / S.O. Subbotin. – Zaporizhzhia: ZNTU, 2008. – 341 s

Рецензія/Peer review : 17.11.2014 р.

Надрукована/Printed :1.1.2015 р.