УДК 621.324

ФОРМУВАННЯ ЛОГІЧНОЇ СТРУКТУРИ РОЗПОДІЛЕНОГО ГОМОГЕННОГО СХОВИЩА ДАНИХ

А.А. Пашнєв[1], О.В. Курко[2], О.В. Мігура[2]
([1]Харківський університет Повітряних Сил ім. І. Кожедуба,
[2]Обєднаний науково-дослідний інститут ЗС України, Харків)

Запропоновано підхід до формування логічної структури розподіленого гомогенного сховища даних (РГСД), що базується на мінімізації сумарного часу доступу до оброблюваної інформації при виконанні ряду фіксованих обмежень для корпоративної обчислювальної мережі.

логічна структура, розподілене гомогенне сховище даних

Вступ. Концепція розподіленого гомогенного сховища даних вимагає наявності нереляційної системи файлів даних, які мають наступні властивості [1]: однакову предметну орієнтованість (не є оперативно-прикладними), є інтегрованими (розподілені не тільки файли сховища, а й джерела поповнення інформації) та прив'язаними до часу, є незмінними (тобто не можуть бути зміненими в оперативному режимі, а процес поповнення ніяк не пов'язаний із зміною вже накопиченої інформації) та призначеними для підтримки прийняття рішень. Кінцевою метою створення такого сховища даних є інтеграція корпоративних даних у єдиному надійному репозитарії [2] для подальшого управління ними та проведення постійного аналізу. При створенні РГСД серед багатьох проблемних задач є задача формування логічної структури, підхід до рішення якої розглянутий у даній статті.

Формування структури сховища. При оптимізації логічної структури взаємопов'язаних файлів даних (баз або сховищ даних) загальноприйнятою є наступна послідовність дій [3]: відображення множини додатків прикладних програм у концептуальній схемі; перехід до канонічної структури даних без надлишкових зв'язків [4]; розробка на базі канонічної структури оптимальної логічної структури (щодо обраного критерію оптимальності). Як критерій оптимальності в [5] пропонується вибрати мінімум сумарної кількості доступів по зв'язках логічної структури відповідно до механізму роботи транзакцій обробки даних. Однак при синтезі РГСД корпоративної обчислювальної мережі (КОМ) установи, розташованого на декількох віддалених територіях, даний підхід є неприйнятним внаслідок різнорідності як фізичних зв'язків, так і вико-

ристовуваної зовнішньої пам'яті мережі. Тому пропонується модифікація даного критерію: мінімізувати сумарний час доступу до зовнішньої пам'яті сховища у процесі функціонування КОМ.

Нехай $V = \left\{ v_i \middle| i = \overline{1,m} \right\}$ – множина файлів даних (ФД) РГСД КОМ; $W = \left\{ \omega_j \middle| j = \overline{1,n} \right\}$ – множина зв'язків між елементами **V**, обумовлена відображенням $\psi : \tilde{V} \to W$, $\tilde{V} \subseteq V \times V$. Тоді шукана канонічна структура представляється орієнтованим графом G =(V, (W,ψ)).

Для кожної k-ої транзакції прикладних програм РГСД КОМ із множини $R = \left\{ r_k \middle| k = \overline{1,l} \right\}$ визначимо частоту запуску за розглянутий період – f_k, а також кортеж задіяних ФД (вершин графа G) – $Z_k = \left\langle v_{i_k}^{(k)} \middle| i_k \in \overline{1,m_k} \right\rangle$, де m_k ($m \geq m_k$) – кількість вершин кортежу з номером k, якому відповідає шлях на графі G з початком у вершині $v_1^{(k)}$ і кінцем у вершині $v_{m_k}^{(k)}$. При цьому між сусідніми вершинами будь-якого кортежу повинен існувати логічний зв'язок.

Розглянемо дві множини: $A = \left\{ a_\beta \middle| \beta = \overline{1,\beta_A} \right\}$ – множину усіх можливих варіантів логічних структур розглянутого РГСД; $e = \left\{ b_\gamma \middle| \gamma = \overline{1,\gamma_B} \right\}$ – множину варіантів реалізації зв'язків між ФД у логічній структурі. Конкретний варіант логічної структури $a_{\beta'}$ для варіанта реалізації зв'язку $b_{\gamma'}$ опишемо булевою матрицею $X_{\beta'} = \left(x_{j\gamma'} \right)$, у якій $x_{j\gamma'} = 1$, якщо зв'язок ω_j реалізований за варіантом γ'.

З множини $\Theta_X = \left\{ X_{\beta'} \right\}$ виділимо підмножину $\Theta_{\tilde{X}} \subset \Theta_X$ матриць $X_{\beta'}$, у яких фізично реалізовані всі непорожні зв'язки.

При реалізації математичної моделі логічної структури РГСД КОМ будемо вимагати дотримання нижчеперелічених обмежень.

Обмеження 1. Кожний зв'язок синтезованої логічної структури реалізується тільки одним варіантом, отже,

$$\sum_{\gamma'=1}^{\gamma_\beta} x_{j\gamma'} = 1, \quad j = \overline{1,n}. \tag{1}$$

Обмеження 2. Всі обрані зв'язки повинні бути фізично реалізованими:

$$X_{\beta'} = \left(x_{j\gamma'} \right) \subset \Theta_{\tilde{X}}. \tag{2}$$

Обмеження 3. Сумарний об'єм задіяної зовнішньої пам'яті при реалізації переходів від кортежів Z_k до логічних структур РГСД КОМ не повинен перевищувати граничного значення d_{lim}, тобто

$$\sum_{j=1}^{n}\sum_{\gamma'=1}^{\gamma_\beta} d_{j\gamma'} \cdot x_{j\gamma'} \le d_{lim} , \qquad (3)$$

де $d_{j\gamma'}$ – об'єм зовнішньої пам'яті КОМ, необхідної для варіанта γ'.

Виходячи з того, що обраний критерій оптимальності логічної структури припускає мінімізацію сумарного часу доступу до необхідної інформації РГСД у процесі її функціонування, то цільова фукнція записується як

$$\sum_{k=1}^{l} f_k \cdot \left(\sum_{j=1}^{n}\sum_{\gamma'=1}^{\gamma_\beta} t_{j\gamma'} \cdot \boldsymbol{l}_{kj} \cdot x_{j\gamma'} \right) \to \min , \qquad (4)$$

якщо $j \in \left\{ \psi\left(v_{k'}^{(k)}, v_{k'+1}^{(k)}\right) \middle| k' \in \overline{1, m_k - 1} \right\}$; $t_{j\gamma'}$ – час пересилання одного логічного блоку за j-м зв'язком у варіанті γ'; l_{kj} – середній розмір одного запиту транзакцій з номером k, що пересилається шляхом j.

Тоді для знаходження оптимальної логічної структури РБД необхідно вирішити задачу лінійного булева програмування (1) – (4). Кращі результати за часом рішення при великих розмінностях V і W дав метод "віток та меж" з розгалуженням "з останньої вершини".

Висновок. Розглянутий підхід до формування логічної структури РГСД КОМ доцільно використовувати для систем з великою кількістю ФД із складною ієрархічною структурою, а кількість прикладних програм не менш, ніж на порядок, перевершує кількість ФД. Напрямком подальших досліджень є оптимізація інформаційних потоків РГСД.

ЛІТЕРАТУРА

1. *Berson A., Smith S. Data Warehousing. – N.Y.: McGraw Hill C., 1997. – 532 p.*
2. *Конолли Т., Бетт К., Стращан А. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. – М.: Вильямс, 2000. – 1120 с.*
3. *Хаббард Дж. Автоматизированное проектирование баз данных. – М.: Мир, 1984. – 292 с.*
4. *Королёв А.В., Кучук Г.А., Пашнев А.А. Управление сетевыми ресурсами. – Х.: ХВУ, 2004. – 224 с.*
5. *Wood H.M., Kimblton S.R. Access Control Mechanisms for a Network Operations Systems // AFIFS Conf. Proc. – New York: AFIFS Press. – 1989. – Vol. 48. – P. 821-829.*

Надійшла 6.02.2006

Рецензент: доктор технічних наук, професор В.А. Краснобаєв,
Харківський національний технічний університет сільського господарства..