

МЕТОД СИНТЕЗУ ЛОГІЧНОЇ СТРУКТУРИ МЕРЕЖЕВОЇ БАЗИ ДАНИХ

к.т.н. Г.А. Кучук
(подав проф. А.В. Корольов)

Запропонований метод синтезу логічної структури мережевої бази даних (БД), у якому вхідними даними математичної моделі є параметри канонічної моделі БД.

Бази даних за способом використання інформації, що міститься в них, класифікуються на однокористувальницькі та багатокористувальницькі, які поділяються на мережеві (МБД) і розподілені (РБД) [1]. МБД найчастіше функціонують у середовищі локальних обчислювальних мереж (ЛОМ) інформаційних систем (ІС). Для МБД суттєві як експлуатаційні характеристики (час доступу, час пошуку, об'єми інформаційних масивів), так і питання адміністрування (захист, достовірність, організація супроводження). Архітектура МБД суттєво залежить від топології базової ЛОМ. У найбільш поширених на даний час однорангових ЛОМ [2] МБД може розташовуватися на будь-якому комп'ютері мережі, який може виконувати як функції клієнта, так і функції серверу. В ЛОМ з виділеним сервером МБД розташовується на сервері і має архітектуру "клієнт - сервер", яка в загальному випадку подається у вигляді трьох моделей [3]: RDA - модель доступу до віддалених даних (Remote Data Access); DBS - модель серверу даних (Data Base Server); AS - модель серверу додатків (Application Server).

Основними критеріями при оцінюванні ефективності синтезованої логічної структури МБД є [4]: мінімум сумарного часу завантаження бази; мінімум сумарного часу обслуговування транзакцій користувачів; мінімум сумарної довжини шляхів доступу до шуканих інформаційних елементів; мінімум часу обслуговування оперативних запитів.

Логічна структура (ЛС) МБД, яка формується на базі заданої канонічної структури (КС) повинна [5]: забезпечувати зберігання семантичних властивостей інформаційних елементів предметної області та зв'язків між ними, визначених в КС; враховувати можливості як базової системи управління базами даних (СУБД), так і вимоги різних режимів використання ІС; забезпечувати зручність та простоту формування запитів до БД.

Враховуючи вищесказане, будемо розглядати синтез ЛС МБД як процес пошуку варіанта відображення КС МБД в ЛС, який є оптималь-

ним з точки зору заданого критерію ефективності та задовольняє основним вимогам, що пред'являються до ЛС.

Розглянемо множину відображень графа канонічної структури БД у можливі варіанти побудови відповідної логічної структури:

$$\Theta = \{ \theta \mid \theta : G_K(V, C_K) \rightarrow G_\Lambda(Z, C_\Lambda) \},$$

де $G_K(V, C)$ - граф канонічної структури БД; $G_\Lambda(Z, C_\Lambda)$ - граф одного з альтернативних варіантів логічної структури; $V = \{ v_l \mid l = \overline{1, L} \}$ - множина інформаційних елементів предметної області (груп даних канонічної моделі); $C_K = \{ (v_l, v_{l'}) \mid l, l' = \overline{1, L} \}$ - множина взаємозв'язків між елементами V ; $Z = \{ z_i \mid i = \overline{1, I} \}$, $C_\Lambda = \{ (z_i, z_{i'}) \mid i, i' = \overline{1, I} \}$ - множини логічних записів (ЛЗ) та взаємозв'язків між ними відповідно.

Для повного опису графу $G_K(V, C)$ визначимо булеву матрицю взаємозв'язків між групами даних $W = (w_{ll'})$, де $w_{ll'} = 1$ тоді, коли між групами v_l та $v_{l'}$ канонічної структури встановлений взаємозв'язок. Характеристики елементів графа описуються наступними параметрами: $N = (n_l \mid l = \overline{1, L})$ - вектор кількості екземплярів груп; $H = (h_l \mid l = \overline{1, L})$ - вектор довжин груп, де $h_l = \sum_{v \in v_l} h_v^y$ - довжина групи v_l , яка дорівнює

сумарній довжині її ключів та атрибутів даних; $R = (r_{ll'} \mid l, l' = \overline{1, L})$ - матриця усереднених коефіцієнтів зв'язку, які визначають кількість зв'язків між екземплярами груп v_l та $v_{l'}$; $B = (b_{ll'} \mid l, l' = \overline{1, L})$ - вектор потрібних екземплярів групи для кожного із зв'язків (l, l') .

Структуру транзакцій $t_j \in T = \{ t_j \mid j = \overline{1, J} \}$ зручно задавати за допомогою дерев пошуку [6], які визначаються на графі канонічної структури. При цьому необхідно визначити: булеві матриці потрібних груп $A = (a_{lj} \mid l = \overline{1, L}; j = \overline{1, J})$ та зв'язків $F = (f_{ll'}^j \mid l, l' = \overline{1, L}; j = \overline{1, J})$, які використовуються при пошуку необхідних інформаційних елементів; вектор частоти використання транзакцій $\Gamma = (\gamma_j \mid j = \overline{1, J})$; матрицю середніх значень сумарної кількості переглядаємих при пошуку показників зв'язку та екземплярів груп $M = (\mu_{ll'}^j \mid l, l' = \overline{1, L}; j = \overline{1, J})$; множину матриць середніх значень сумарної кількості потрібних екземплярів груп $P = (p_{ll'}^j \mid j = \overline{1, J}; l, l' = \overline{1, L}; f_{ll'}^j = 1)$. Враховуючи те, що при зміні точки

входу транзакції змінюються чисельні характеристики параметрів, для кожної транзакції t_j розглянемо вектор можливих точок входу $U_j = \left(u_{mj} \mid m = \overline{1, m_j} \right)$ та відповідно такі множини: $A^{(m)} = \left(a_{lj}^m \right)$; $F^{(m)} = \left(f_{ll'}^{jm} \right)$; $\Gamma^{(m)} = \left(\gamma_j^m \right)$; $M^{(m)} = \left(\mu_{ll'}^{jm} \right)$; $P^{(m)} = \left(p_{ll'}^{jm} \right)$.

Для МБД при побудові логічної структури необхідно також враховувати завдання на коректування $k_s \in K = \left\{ k_s \mid s = \overline{1, S} \right\}$, які виконуються в монопольному режимі обробки інформації. Для цього необхідно задати булеві матриці використання груп $W = \left(\omega_{sl}^{(k)} \mid s = \overline{1, S}; l = \overline{1, L} \right)$; вектор частот запуску завдань на коректування $\Gamma^{(k)} = \left(\gamma_s^{(k)} \mid s = \overline{1, S} \right)$; матрицю середніх значень кількості коректованих екземплярів вибраних груп $\Delta^{(k)} = \left(\delta_{sl} \mid s = \overline{1, S}; l = \overline{1, L} \right)$ із середнім часом коректування t_k .

Будемо враховувати також середній час виконання таких дій: t_z - запису ЛЗ до бази; t_u - формування покажчика; t_{ru} - аналізу покажчика; t_l - пошуку та вибору одного екземпляра групи v_l ; t_o - об'єднання груп в ЛЗ.

Введемо булеві змінні x_{li} , \tilde{x}_i , $y_{ll'}^{ll'}$, $\tilde{y}_{ll'}$, \tilde{y}_{jm} таким чином, що $x_{li} = 1$, якщо група v_l входить до складу ЛЗ z_i ; $\tilde{x}_i = 1$, якщо ЛЗ z_i входить до логічної структури БД; $y_{ll'}^{ll'} = 1$, якщо дуга (l, l') графа канонічної структури входить в зв'язок (i, i') логічної структури; $\tilde{y}_{ll'} = 1$, якщо зв'язок (i, i') суттєвий, тобто $\sum_{l, l'=1}^L y_{ll'}^{ll'} = 1$; $\tilde{y}_{jm} = 1$, якщо для транзакції t_j обира-

ється m - та точка входу. Для визначення того, чи є ЛЗ z_i коректованим,

введемо булеву змінну \tilde{x}_{is} , причому $\left(\tilde{x}_{is} = 1 \Leftrightarrow \sum_{l=1}^L x_{li} \cdot \omega_{sl}^{(k)} \geq 1 \right)$.

У вищенаведених позначеннях задача синтезу оптимальної логічної структури МБД за критерієм мінімуму сумарного часу загрузки інформації до бази даних, обробки транзакцій користувачів і коректування БД адміністратором, формулюється таким чином:

$$\left(\min_{\tilde{x}_{is}, \tilde{y}_{jm}} \sum_{i=1}^I \left(\left(\sum_{l=1}^L x_{li} - 1 \right) \cdot t_o + \tilde{x}_i \cdot t_z + \sum_{s=1}^S \gamma_s^{(k)} \cdot \sum_{l=1}^L \delta_{sl}^{(k)} \cdot \tilde{x}_{is} \cdot t_k + \sum_{i \neq i'} \sum_{l \neq l'}^L y_{ll'}^{ll'} \cdot \left(\beta_{ll'} \cdot t_u + \sum_{j=1}^J \gamma_j \cdot \sum_{m=1}^M y_{jm} \cdot \left(p_{ll'}^{jm} \cdot t_l + \mu_{ll'}^{jm} \cdot t_{ru} \right) \right) \right) \right) \quad (1)$$

Цільова функція (1) розглядається при таких обмеженнях:

- на допустиму довжину N_l^{\max} формуемого логічного запису

$$\sum_{l=1}^L x_{fi} \cdot \left(h_l + \sum_{l' \neq l}^L x_{fi'} (r_{ll'} - 1) \cdot h_{l'} \right) \leq N_l^{\max}; \quad (2)$$

- на відсутність дублювання даних

$$\sum_{i=1}^I x_{fi} = 1, \quad l = \overline{1, L}; \quad (3)$$

- на склад логічного запису:

$$\sum_{l=1}^L x_{fi} \leq N_d^{(\max)}, \quad i = \overline{1, I}; \quad (4)$$

$$\sum_{l' \in L_f} x_{fi} \leq 1, \quad i = \overline{1, I}, \quad (5)$$

де $N_d^{(\max)}$ - обмеження числа груп у складі ЛЗ; $L_f \in L_F$; L_F - множина наборів груп канонічної структури БД, які повинні попасти до складу одного ЛЗ;

- на складність зв'язків між парами логічних записів

$$\sum_{(l, l')}^I y_{ii'}^{ll'} \cdot \beta_{ll'} \leq N_c^{(\max)}, \quad (6)$$

де $N_c^{(\max)}$ - максимально допустима кількість екземплярів зв'язків між ЛЗ;

- на допустиме число суттєвих логічних зв'язків ($N_z^{(\max)}$):

$$\sum_{i' \neq i}^I \tilde{y}_{ii'} \leq N_z^{(\max)}, \quad i = \overline{1, I}; \quad (7)$$

- на час виконання оперативних транзакцій (ОТ):

$$\sum_{i, i'=1}^I \sum_{l, l'=1}^L y_{ii'}^{ll'} \cdot \sum_{m=1}^M \tilde{y}_{jm} \cdot (p_{ll'}^{jm} \cdot t_l + \mu_{ll'}^{jm} \cdot t_{ru}) \leq T_j^{(\max)}, \quad j \in J_0, \quad (8)$$

де J_0 - список ОТ; а $T_j^{(\max)}$ - верхня межа часу обробки транзакції t_j ;

- на однозначність визначення точки входу транзакції

$$\sum_{m=1}^{M_j} \tilde{y}_{jm} = 1, \quad j = \overline{1, J}; \quad (9)$$

- на загальну кількість можливих точок входу:

$$\sum_{j=1}^J \sum_{m=1}^{M_j} \tilde{y}_{jm} \leq N_{inp}^{(\max)}; \quad (10) \quad \sum_{j=1}^J \sum_{m=1}^{M_j} \tilde{y}_{jm} \cdot z_j^{(inp)} \leq V^{(\max)}, \quad (11)$$

де $N_{\text{inp}}^{(\text{max})}$ - максимально можлива кількість точок входу; $V^{(\text{max})}$ - об'єм оперативної пам'яті (ОП), виділений для реалізації вторинних індексів та інверсних показників; $z_j^{(\text{inp})}$ - витрати ОП на реалізацію додаткової точки входу для транзакції t_j ;

- на максимально допустимий час блокування доступу користувачів до даних ЛЗ z_i ($T_i^{(\text{max})}$) з метою коректування

$$\sum_{i' \neq i}^I \left(\sum_{l, l'}^L y_{ii'}^{ll'} \cdot \sum_{m=1}^{M_j} \tilde{y}_{jm} \cdot \left(p_{ll'}^{jm} \cdot t_l + \mu_{ll'}^{jm} \cdot t_{ru} \right) + \sum_{s=l'=1}^S \sum_{is} \tilde{x}_{is} \cdot \delta_{sl}^{(k)} \cdot t_k \right) \leq T_i; \quad (12)$$

- на максимальну кількість записів БД для коректування ($N_{\text{кор}}^{(\text{max})}$):

$$\sum_{i=1}^I \tilde{x}_i \cdot \sum_{s=1}^S \tilde{x}_{is} \leq N_{\text{кор}}^{(\text{max})}. \quad (13)$$

Задача синтезу логічної структури бази даних з цільовою функцією (1) і обмеженнями (2) - (13) є нелінійною задачею дискретного цілочисельного програмування з булевими змінними. Для рішення задач такого класу при допустимих значеннях розмірності розроблений ряд точних і приблизних алгоритмів [7]. Становить інтерес отримання приблизного рішення цієї задачі з мінімально можливою похибкою, якщо задача має велику розмірність.

Результати, що будуть отримані при рішенні задачі (1) - (13), є вихідними для проектування оптимальної фізичної структури МБД, логічної та фізичної структури базових СУБД і обчислювальної мережі з ефективним мережевим протоколом.

ЛІТЕРАТУРА

1. Тамер О.М., Валдуриз П. Распределенные и параллельные базы данных // СУБД. – 1996. – № 4. – С. 4 - 26.
2. Зильбершатц А., Стоунбрейкер М., Ульман Д. Базы данных: достижения и перспективы // СУБД. – 1996. – № 3. – С. 103 - 117.
3. Loomis M. Client - server architecture // Object Oriented Programming. – 1999. – № 4 (99). – Р. 40 - 44.
4. Зиндер Е.З. Проектирование баз данных: новые требования, новые подходы // СУБД. – 1996. – № 3. – С. 10 - 22.
5. Mattinson R. Data Warehousing: Strategies, Technologies and Techniques. – New York: McGraw Hill, 1998. – 912 p.
6. Саймак А. Обработка транзакций // СУБД. – 1997. – № 2. – С. 70 - 82.

7. Сергиенко И.В., Каспицкая М.Ф. Модели и методы решения на ЭВМ комбинаторных задач оптимизации. – К.: Наук. думка, 1981. – 287 с.

Подана до редколегії 16.01.2001
