

Безкоровайний В. В., Колесник Л. В.

Харківський національний університет радіоелектроніки

МАТЕМАТИЧНА МОДЕЛЬ БАГАТОКРИТЕРІАЛЬНОЇ ЗАДАЧІ РЕІНЖИНІРИНГУ ФІЗИЧНИХ СТРУКТУР РОЗПОДІЛЕНИХ БАЗ ДАНИХ

A mathematical model is proposed for the multi criteria problem of reengineering the physical structures of distributed databases in terms of reduced costs, access time and network traffic. It allows, through the use of universal functions of general utility and utility of particular criteria, to more accurately take into account the preferences of the person making design decisions, and significantly reduce the time complexity of the procedures for calculating scalar estimates.

Одними з найважливішими елементами сучасних систем управління великомасштабними об'єктами, систем моніторингу, CADS-систем, інших об'єктів з різних сфер людської діяльності стають розподілені бази даних (РБД) [1]. Зміна технологій, елементної бази, умов функціонування або функціональних вимог до подібних об'єктів на певному етапі призводить до необхідності реінжинірингу їх інфраструктури, основу якої складають інформаційно-обчислювальні мережі та РБД. При цьому для РБД (як територіально розподілених об'єктів) характерною є залежність їх функціональних і вартісних характеристик від топології локальних баз (ЛБ) [2]. Це потребує розв'язання комплексу оптимізаційних комбінаторних задач з урахуванням множини функціональних і вартісних показників.

Завдання реінжинірингу фізичних структур РБД розглядається в такій постановці [3]. Задані: множина користувачів розподіленої бази $I = \{i\}$, $i = \overline{1, n}$, пов'язаних деякою комп'ютерною мережею $G = (I, R)$ (де $R = [r_{ik}]$, $i, k = \overline{1, n}$ – матриця суміжності, яка визначає множину безпосередніх зв'язків між вузлами комп'ютерної мережі); множина інформаційних ресурсів (ІР) у вигляді файлів, фондів тощо $J = \{j\}$, $j = \overline{1, m}$; обсяги використовуваних ІР $l = [l_j]$, $j = \overline{1, m}$; інтенсивності надходження запитів із кожного з вузлів комп'ютерної мережі до кожного з ІР $\lambda = [\lambda_{ij}]$, $i = \overline{1, n}$, $j = \overline{1, m}$; обсяги запитів до ІР з кожного з вузлів комп'ютерної мережі $a = [a_{ij}]$, $i = \overline{1, n}$, $j = \overline{1, m}$; витрати на зберігання ІР у вузлах комп'ютерної мережі $c(x) = [c_{ij}(x)]$, $i = \overline{1, n}$, $j = \overline{1, m}$; ($x = [x_{ij}]$ – матриця, яка відображає розміщення ІР у вузлах комп'ютерної

мережі; x_{ij} – булева змінна: $x_{ij} = 1$, якщо j -й ІР зберігається в i -му в вузлі мережі; $x_{ij} = 0$ – в іншому випадку); c_t – витрати на передачу одиниці інформації; $d = [d_{ij}]$ – обсяги інформації, що передається при оновленні ІР з вузлів мережі; $z = [z_{ij}]$ – матриця оновлень ІР ($z_{ij} = 1$, якщо j -й ІР оновлюється з i -го вузла комп'ютерної мережі; $z_{ij} = 0$ – в іншому випадку); $X = \{x\}$ – множина допустимих варіантів реалізації фізичних структур РБД.

Необхідно визначити найкращий варіант фізичної структури РБД з множини допустимих $x^o \in X$ (кількість ЛБ, розподіл ІР по ЛБ $x = [x_{ij}]$, розміщення ЛБ по вузлах мережі, обсяги запам'ятовуючих пристроїв для зберігання ЛБ, пропускні спроможності каналів між вузлами комп'ютерної мережі.

Критерії оптимізації:

– витрати на x -реалізацію фізичної структури РБД:

$$c(x) = \sum_{i=1}^n \sum_{j=1}^m c_{ij}(x) x_{ij} + c_t \sum_{i=1}^n \sum_{j=1}^m (\alpha_{ij} + \beta_{ij}) x_{ij} + c_t \sum_{i=1}^n \sum_{j=1}^m d_{ij} z_{ij} \rightarrow \min_{x \in X}, \quad (1)$$

де $\alpha_{ij} = \lambda_{ij} a_{ij} T$ – сумарний обсяг запитів до j -го ІР з i -го вузла мережі (λ_{ij} і a_{ij} інтенсивність виникнення запитів і обсяг запиту до j -го ІР з i -го вузла мережі; T – інтервал часу, для якого здійснюється оцінка витрат); β_{ij} – сумарний обсяг відповідей на запити до j -го ІР з i -го вузла мережі.

– час доступу до інформаційних ресурсів при x -реалізації фізичної структури бази даних:

$$t(x) = \frac{1}{n} \sum_{i=1}^n [t_{ij}^{tr}(x) + t_{ij}^{pr}(x) + t_{ij}^{qp}(x) + t_{ij}^{rp}(x)] x_{ij} \rightarrow \min_{x \in X}, \quad (2)$$

де $t_{ij}^{tr}(x)$ – час передачі запиту з i -го вузла комп'ютерної мережі до j -го ІР; $t_{ij}^{pr}(x)$ – час очікування в черзі запиту з i -го вузла по j -му ІР; $t_{ij}^{qp}(x)$ – час обробки запиту з i -го вузла мережі по j -му ІР; $t_{ij}^{rp}(x)$ – час передачі відповіді на запит з i -го вузла по j -му ІР;

– обсяг інформації, що передається при x -реалізації варіанта фізичної структури бази даних:

$$v(x) = \sum_{i=1}^n \sum_{j=1}^m (\alpha_{ij} + \beta_{ij}) x_{ij} + \sum_{i=1}^n \sum_{j=1}^m l_{ij} z_{ij} \rightarrow \min_{x \in X}. \quad (3)$$

При цьому необхідно забезпечити повноту РБД за рахунок розподілу всіх ІР (з

можливим дублюванням) по локальних базах $\sum_{i=1}^n x_{ij} \geq 1 \quad \forall j=1, n$ і виконання обмеження на час доступу до інформаційних ресурсів бази $t(x) \leq t^*$ (де t^* – максимально допустиме значення часу доступу до ІР).

Комбінаторний характер задач реінжинірингу фізичних структур РБД передбачає використання автоматичної кількісної оцінки варіантів з множини допустимих $x \in X$. Як методологічну основу для побудови оцінок використаємо теорію корисності, відповідно до якої для кожної з альтернатив $x \in X$ може бути визначено значення її корисності $P(x)$ [4]. При цьому для всіх $x, y \in X$:

$$x \sim y \leftrightarrow P(x) = P(y); \quad x \succ y \leftrightarrow P(x) > P(y); \quad x \succeq y \leftrightarrow P(x) \geq P(y). \quad (4)$$

Формально задача вибору найкращого варіанту фізичної структури РБД $x^o \in X$ може бути зведена до задачі оптимізації:

$$x^o = \arg \max_{x \in X} P(x). \quad (5)$$

Введемо позначення введених критеріїв оптимізації:

$$k_1(x) = c(x) \rightarrow \min_{x \in X}, \quad k_2(x) = t(x) \rightarrow \min_{x \in X}, \quad k_3(x) = v(x) \rightarrow \min_{x \in X}.$$

Для скалярного оцінювання варіантів пропонується використати універсальну функцію корисності, побудовану на основі полінома Колмогорова-Габора [5], яка з урахування вибраної множини часткових критеріїв, матиме такий вигляд:

$$P(x) = \sum_{i=1}^3 \lambda_i \xi_i(x) + \sum_{i=1}^3 \sum_{j=i}^3 \lambda_{ij} \xi_i(x) \xi_j(x) + \sum_{i=1}^3 \sum_{j=i}^3 \sum_{l=j}^3 \lambda_{ijl} \xi_i(x) \xi_j(x) \xi_l(x), \quad (6)$$

де $\lambda_i, \lambda_{ij}, \lambda_{ijl}$ – вагові коефіцієнти часткових критеріїв $k_i(x)$, $i = \overline{1,3}$ та їхніх добутоків $\lambda_i \geq 0, \lambda_{ij} \geq 0, \lambda_{ijl} \geq 0$; $\xi_i(x)$ – функції корисності часткових критеріїв $k_i(s)$, $i = \overline{1,3}$.

Кількісна оцінка вагових коефіцієнтів часткових критеріїв, може бути проведена класичними експертними методами або з використання технології компараторної ідентифікації. З метою скорочення множини пошуку рішень задачі доцільним є попереднє виділення або паралельне з генерацією формування підмножини ефективних (Парето-оптимальних) проєктних рішень.

Для оцінки корисності окремих значень часткових критеріїв використаємо функцію-склейку, яка має переваги за показниками точності та кількості комп'ютерних операцій для її обчислення у порівнянні з відомими функціями Гаусса, Харрінгтона і логістичною функцією [5]:

$$\xi_i(x) = \begin{cases} \bar{a}_i \cdot (b_{i1} + 1) \cdot \left\{ 1 - \left[b_{i1} / \left(b_{i1} + \frac{\bar{k}_i(x)}{\bar{k}_{ia}} \right) \right] \right\}, & 0 \leq \bar{k}_i(x) \leq \bar{k}_{ia}; \\ \bar{a}_i + (1 - \bar{a}_i) \cdot (b_{i2} + 1) \cdot \left\{ 1 - \left[b_{i2} / \left(b_{i2} + \frac{\bar{k}_i(x) - \bar{k}_{ia}}{1 - \bar{k}_{ia}} \right) \right] \right\}, & \bar{k}_{ia} < \bar{k}_i(x) \leq 1, \end{cases} \quad (7)$$

де \bar{k}_{ia}, \bar{a}_i – нормовані значення координат точки склеювання складових функцій, $0 \leq \bar{k}_{ia} \leq 1, 0 \leq \bar{a}_i \leq 1$; b_{i1}, b_{i2} – параметри, нелінійності початкової і кінцевої складових функцій.

Запропонована математична модель трикритеріальної задачі реінжинірингу фізичних структур РБД дозволяє за рахунок використання універсальних функцій загальної корисності та функції корисності часткових критеріїв точніше враховувати переваги особи, що приймає проєктні рішення. Використання універсальної функції корисності часткових критеріїв за рахунок виключення операцій піднесення до степені дозволяє істотно скоротити часову складність процедур обчислення скалярних оцінок. Напрямок подальших досліджень може бути врахування в моделі невизначеності функціональних і вартісних характеристик РБД з використанням апарату інтервального аналізу.

Література

1. Ёсу М. Т., Вальдуріес П. Принципы организации распределенных баз данных / пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2021. 672 с.
2. Бескоровайнй В. В., Подоляка К. Е. Разработка системологической модели проблемы структурно-топологического реинжиниринга систем крупномасштабного мониторинга // Восточно-Европейский журнал передовых технологий. 2015. №3(75). С. 37-42.
3. Бескоровайнй В. В., Ульянова О. С. Математические модели многокритериального синтеза физических структур распределенных баз данных // Восточно-европейский журнал передовых технологий. 2010. № 4/4 (46). С. 44-48.
4. Интроспективный анализ. Методы и средства экспертного оценивания / В.В. Крючковский, Э.Г. Петров, Н.А. Соколова, В.Е. Ходаков. Херсон: Гринь ДС, 2011. 284 с.
5. Beskorovainyi V., Berezovskyi H. Identification of preferences in decision support systems // ECONTechMOD. 2017. Vol. 06, No. 4. P. 15-20.