

## СИНТЕЗ ЛОГИЧЕСКОЙ СТРУКТУРЫ РАСПРЕДЕЛЕННОЙ БАЗЫ ДАННЫХ

к.т.н. Г.А. Кучук  
(представил д.т.н. проф. В.С. Харченко)

Предложен подход к синтезу логической структуры распределенной базы данных (РБД), базирующийся на минимизации суммарного времени доступа к обрабатываемой информации.

При решении задачи оптимизации логической структуры баз данных (БД) общепринята следующая последовательность действий [1]:

- отображение множества приложений прикладных программ в концептуальной схеме;
- переход к канонической структуре БД без избыточных связей [2];
- разработка на базе канонической структуры оптимальной логической структуры БД (относительно выбранного критерия оптимальности).

В качестве критерия оптимальности в [3] предлагается выбрать минимум суммарного количества доступов по связям логической структуры в соответствии с механизмом работы системы управления базами данных (СУБД). Однако при синтезе РБД административной вычислительной сети (АВС) учреждения, расположенного на нескольких удаленных территориях, данный подход неприемлем вследствие разнородности как физических связей, так и используемой внешней памяти сети. Поэтому предлагается модификация данного критерия: минимизировать суммарное время доступа к внешней памяти БД в процессе функционирования АВС.

Пусть  $V = \{v_i \mid i = \overline{1, m}\}$  - множество реляционных таблиц (РТ) РБД АВС;  $W = \{\omega_j \mid j = \overline{1, n}\}$  - множество связей между элементами  $V$ , определяемое отображением  $\psi: \tilde{V} \rightarrow W$ ,  $\tilde{V} \subseteq V \times V$ . Тогда искомая каноническая структура представляется ориентированным графом  $G = (V, (W, \psi))$ . Для каждой  $k$ -ой транзакции прикладных программ РБД АВС из

© к.т.н. Г.А. Кучук, 1998

множества  $\mathbf{R} = \{r_k \mid k = \overline{1, l}\}$  определим частоту запуска за рассматриваемый период -  $f_k$ , а также кортеж задействованных РТ (вершин графа  $\mathbf{G}$ ) -  $\mathbf{Z}_k = \langle v_{i_k}^{(k)} \mid i_k \in \overline{1, m_k} \rangle$ , где  $m_k$  ( $m \geq m_k$ ) - количество вершин кортежа с номером  $k$ , которому соответствует путь на графе  $\mathbf{G}$  с началом в вершине  $v_1^{(k)}$  и концом в вершине  $v_{m_k}^{(k)}$ . При этом между соседними вершинами любого кортежа должна существовать логическая связь.

Рассмотрим два множества:  $\mathbf{A} = \{a_\beta \mid \beta = \overline{1, \beta_A}\}$  - множество всех возможных вариантов логических структур рассматриваемой РБД;  $\mathbf{B} = \{b_\gamma \mid \gamma = \overline{1, \gamma_B}\}$  - множество вариантов реализации связей между РТ в логической структуре. Конкретный вариант логической структуры  $a_{\beta'}$  для варианта реализации связи  $b_{\gamma'}$ , опишем булевой матрицей  $\mathbf{X}_{\beta'} = (x_{j\gamma'})$ , в которой  $x_{j\gamma'} = 1$ , если связь  $\omega_j$  реализована по варианту  $\gamma'$ .

Из множества  $\Theta_X = \{\mathbf{X}_{\beta'}\}$  выделим подмножество  $\Theta_{\tilde{X}} \subset \Theta_X$  матриц  $\mathbf{X}_{\beta'}$ , в которых физически реализуемы все непустые связи.

При реализации математической модели логической структуры РБД АВС потребуем соблюдения нижеперечисленных ограничений.

Ограничение 1. Каждая связь синтезируемой логической структуры реализуется только одним вариантом, следовательно,

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

Ограничение 2. Все выбранные связи должны быть физически реализуемы, т.е.

$$\mathbf{X}_{\beta'} = (x_{j\gamma'}) \in \Theta_{\tilde{X}}. \quad (2)$$

Ограничение 3. Суммарный объем задействованной внешней памяти при реализации переходов от кортежей  $\mathbf{Z}_k$  к логическим структурам РБД АВС не должен превышать предельного значения  $d_{lim}$ , т.е.

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

где  $\mathbf{d}_{j\gamma'}$  - объем внешней памяти АВС, необходимой для варианта  $\gamma'$ .

Так как выбранный критерий оптимальности логической структуры предполагает минимизацию суммарного времени доступа к требуемой информации РБД в процессе ее функционирования, то целевая функция записывается как

$$\sum_{k=1}^l \mathbf{f}_k \cdot \left( \sum_{j=1}^n \sum_{\gamma'=1}^{\gamma\beta} \mathbf{t}_{j\gamma'} \cdot l_{kj} \cdot \mathbf{x}_{j\gamma'} \right) \rightarrow \min, \quad (4)$$

если  $\mathbf{j} \in \left\{ \Psi \left( \mathbf{v}_{k'}^{(k)}, \mathbf{v}_{k'+1}^{(k)} \right) \mid k' \in 1, \overline{\mathbf{m}_k - 1} \right\}$ ,

$\mathbf{t}_{j\gamma'}$  - время пересылки одного логического блока по  $\mathbf{j}$  - ой связи в варианте  $\gamma'$ ,

$l_{kj}$  - средний размер одного запроса транзакций с номером  $\mathbf{k}$ , пересылаемого по пути  $\mathbf{j}$ .

Тогда для нахождения оптимальной логической структуры РБД необходимо решить задачу линейного булева программирования (1) - (4). Лучшие результаты по времени решения при больших размерностях  $\mathbf{V}$  и  $\mathbf{W}$  дал метод "ветвей и границ" с ветвлением "из последней вершины" [4].

Рассмотренный на примере логической структуры РБД АВС подход, целесообразно использовать для систем с большим количеством БД, каждая из которых на логическом уровне имеет сложную иерархическую структуру, а количество прикладных программ не менее, чем на порядок, превосходит количество БД.

## ЛИТЕРАТУРА

1. Хаббард Дж. Автоматизированное проектирование баз данных. - М.: Мир, 1984. - 292 с.
2. Мартин Дж. Организация баз данных в вычислительных системах. - М.: Мир, 1980. - 662 с.
3. 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.