

ВЫБОР СТРАТЕГИИ РАСПРЕДЕЛЕНИЯ ЛОГИЧЕСКИХ ФРАГМЕНТОВ БАЗ ДАННЫХ

к.т.н. Г.А. Кучук, к.т.н. Б.В. Остроумов, А.А. Пашнев
(представил проф. А.В. Королев)

Предложен подход к анализу различных стратегий размещения логических фрагментов баз данных в среде распределенной вычислительной сети.

Рассмотрим распределенную сеть U неоднородных ЭВМ, состоящую из m локальных вычислительных сетей (ЛВС) u_i ($U = \{u_i | i = \overline{1, m}\}$), связанных между собой каналами с относительно низкой пропускной способностью v_{i_1, i_2} ($V = \{v_{i_1, i_2} | i_1, i_2 = \overline{1, m}\}$). В среде сети U предполагается обеспечить функционирование базы данных B , состоящей из n непересекающихся логических фрагментов f_j ($B = \{f_j | j = \overline{1, n}\}$). Стратегии распределения фрагментов базы данных (БД) классифицируются в зависимости от количества задействованных ЛВС и наличия дублирования информации [1].

Для описания множества глобальных стратегий распределения баз данных $\Omega = \{w_l | l \in N\}$ введем отображение φ множества логических фрагментов БД на множество групп ЛВС ($\varphi : B \rightarrow \tilde{U}$, где $\tilde{U} = \{U_k \subset U\}$) [2].

При централизованной стратегии w_1 единственная копия БД располагается на сервере одной из ЛВС сети, которую обозначим как U_0 . Тогда:

$$\varphi(f_{j_1}) = \varphi(f_{j_2}) = U_0 \subset \tilde{U}, \quad \forall j_1, j_2 \in \overline{1, n};$$

$$\text{card}U_0 = 1.$$

При стратегии расчленения данных w_2 логические фрагменты единственной копии базы данных распределяются по разным ЛВС, следовательно:

$$\text{card}\varphi(f_j) = 1, \quad \forall j \in \overline{1, n};$$

$$\exists j_1, j_2 \in \overline{1, n} \mid \varphi(f_{j_1}) \neq \varphi(f_{j_2}).$$

Стратегия дублирования \mathbf{W}_3 характеризуется наличием нескольких полных копий БД на различных ЛВС, поэтому ее формализованное описание выглядит следующим образом:

$$\varphi(\mathbf{f}_{j_1}) = \varphi(\mathbf{f}_{j_2}) = U' \subset \tilde{U}, \forall j_1, j_2 \in \overline{1, n};$$

$$\text{card } U' > 1.$$

Для данной стратегии мощность подмножества U' определяет число полных копий базы данных.

Смешанная стратегия распределения логических фрагментов базы данных \mathbf{W}_4 позволяет иметь несколько копий фрагментов на различных ЛВС, причем каждая ЛВС может содержать произвольный фрагмент. При этом на отображение φ не накладываются дополнительные ограничения, а число копий фрагмента \mathbf{f}_i определяется как мощность подмножества $\varphi(\mathbf{f}_i)$.

Для оценки вышеописанных стратегий определим над множеством Ω множество показателей оценки качества функционирования $\mathbf{G} = \{g_l\}$ так, что $g_l: \Omega \rightarrow \mathbf{R}$ [3]. В частности, множество \mathbf{G} должно включать следующие элементы: \mathbf{g}_1 - оценка надежности функционирования БД; \mathbf{g}_2 - оценка степени локализации ссылок; \mathbf{g}_3 - оценка среднего времени доступа, учитывающая для рассматриваемой стратегии архитектуру сети и пропускные способности каналов $\forall i_1, i_2 \in \mathbf{V}$; \mathbf{g}_4 - оценка затрат на реализацию проекта; \mathbf{g}_5 - оценка затрат ведения базы данных. Для каждого $\mathbf{g}_i \in \mathbf{G}$ введем весовой коэффициент α_i , который определяется исходя из требований, предъявляемых к системе управления базой данных.

Тогда выбор стратегии распределения логических фрагментов БД можно осуществить при помощи

$$\xi(\mathbf{w}_i) = \sum_{l \in \mathbf{L}} \alpha_l g_l(\mathbf{w}_i),$$

где $\mathbf{L} \subset \mathbf{G}$ – подмножество показателей, актуальных для данной БД.

Предложенный подход был использован для анализа функционирования различных баз данных административной вычислительной сети (АВС). Однозначный выбор глобальной стратегии размещения логических фрагментов без предварительного анализа был невозможен, так как в каждом конкретном случае каждая стратегия обладала как преимуществами, так и недостатками перед остальными.

Основным преимуществом стратегий \mathbf{W}_1 , безусловно, является простота реализации, хотя по остальным показателям она уступает стратегиям $\mathbf{W}_2 \div \mathbf{W}_4$. Стратегия \mathbf{W}_3 превосходит остальные по надежности функциони-

рования системы, по доступности и эффективности выборки данных, однако имеет место очень большая избыточность данных и тенденция к нарушению согласованности копий. Стратегия **W₂** позволяет более равномерно распределить нагрузку между ЛВС по сравнению со стратегиями **W₁** и **W₃**, однако для БД с малой локализацией ссылок резко увеличивается время обработки транзакций и уменьшается надежность системы.

Стратегия **W₄** объединяет подходы, связанные с расчленением и дублированием данных с целью приобретения преимуществ, которыми они обладают. Но к сожалению, эта стратегия приобретает сложности каждого из объединяемых подходов. Она является общей в том, что любая часть БД может быть дублирована произвольное количество раз и в каждом узле может содержаться желаемая часть БД. Недостатком стратегии является необходимость хранить информацию о том, где находятся данные в сети, и согласовать произвольное количество хранимых документов, связанных с каждым логическим фрагментом (ЛФ). Обработка и оптимизация запросов являются при использовании смешанной стратегии нетривиальными задачами. Однако, ключевым преимуществом смешанной стратегии является гибкость. Например, можно установить компромисс между объемом памяти, используемой в целом и в каждом отдельном узле, обеспечиваемым уровнем надежности и различными мерами эффективности. Так, архивные данные необходимо запоминать только в одном месте, напротив, более критические данные могут быть дублированы, если требуется достичь требуемого уровня надежности. При дублировании ЛФ, запоминании более одного фрагмента, стоимость согласования, включая стоимость связи, возрастает, однако большее количество данных становится локально доступным, что ведет к снижению качества пересылок и стоимости связи при выполнении запросов.

Хотя распределенная СУБД, реализующая стратегию **W₄**, и является предельно гибкой, остается проблема взаимозависимости различных факторов, влияющих на производительность системы, ее надежность и требования к памяти. Изолировать один фактор от другого весьма трудно. Таким образом, учитывая свойства рассмотренных стратегий распределения фрагментов, можно сделать вывод, что стратегия **w₁**, допускающая лишь централизованное распределение, является простейшей и применяется для небольших БД с невысокими требованиями к надежности системы и небольшими объемами информации. Стратегия **w₂** наиболее подходит для случая, когда либо внешняя память ограничена по сравнению с объемом БД, либо недостаточна надежность централизованной БД, либо когда должна быть повышена эффективность функционирования программных средств обработки транзакций. Стратегия **w₃** применяется для тех ситуаций, когда фактор надежности является критическим, база небольшая, а интенсивность

обновления может быть невысокой, например, БД с интенсивными запросами справочного типа. И наконец, стратегия **W4** является приемлемой тогда, когда ни одна из более простых стратегий не является удовлетворительной.

ЛИТЕРАТУРА

1. Тиоми Т., Фрай Дж. Проектирование структур баз данных // М.: Мир, 1985. – 348 с.
 2. Rothnie I.B. Introduction to a System for Distributed Databases // ACM Trans. Database Syst. – 1980. – № 5. – Р. 1 – 17.
 3. Костюк В.И., Дешко А.И., Игнатенко Б.В. Проектирование информационных моделей в гибких системах // К. – Вища школа, 1987. – 194 с.
-