

Сомов С.К.

Эвристический подход к динамическому размещению фрагментов таблиц распределенных данных

Аннотация: В работе рассмотрен метод повышения эффективности работы распределенных баз данных, основанный на динамическом размещении фрагментов таблиц в зависимости от трафика запросов к данным. Предложен эвристический алгоритм размещения фрагментов таблиц по узлам системы. В качестве критерия оптимизации используется минимум затрат на обработку запросов к фрагментам.

Ключевые слова: распределенные базы данных, распределение фрагментов таблиц данных, эвристический алгоритм

Эффективность работы распределенных систем баз данных (РБД) в значительной степени зависит от качества ее проектирования и, в частности, от используемых подходов к фрагментации таблиц данных и от методов распределения фрагментов по узлам компьютерной сети, на базе которой работает РБД. Однако методы поиска оптимального распределения фрагментов по узлам сетей, особенно крупномасштабных, имеют большую вычислительную сложность и требуют больших ресурсов. Поэтому для снижения сложности задачи в данной работе представлен эвристический алгоритм распределения фрагментов таблиц РБД.

Как в теории, так и на практике для повышения надежности и производительности работы РБД используются различные методы горизонтального (horizontal partitioning) и вертикального (vertical partitioning) фрагментирования, а также и комбинация первых двух методов [1-4]. Еще большего повышения надежности и производительности работы РБД можно добиться при создании идентичных копий фрагментов таблиц (реплик) и их распределении по нескольким узлам компьютерной сети [4-6].

В данной работе рассматривается модель динамического не избыточного (без использования реплик) перераспределения фрагментов таблиц в узлах сети в зависимости от изменения

трафика запросов к данным РБД и величины затрат на миграцию фрагментов таблиц между узлами сети.

Модель предполагает, что фрагментирование таблиц выполнено, а размещение фрагментов по узлам сети, топология которой задана графом, происходит в два этапа. На первом этапе выполняется первоначальное распределение фрагментов по узлам, возможно с наличием дубликатов фрагментов. Затем, на втором этапе происходит не избыточное перераспределение полученного на первом этапе размещения фрагментов, результат которого зависит от текущего значения параметров сети, таких как трафик запросов к фрагментам данных и стоимость использования каналов связи сети. В качестве критерия оптимальности перераспределения фрагментов используется минимум затрат на обработку запросов к фрагментам данных.

Компьютерная сеть, на базе которой функционирует РБД, состоит из множества M узлов сети $N = \{N_1, \dots, N_M\}$, $M = |N|$. Для каждого узла задана его максимальная емкость C_m и ограничение на максимальное количество фрагментов FM_m , которые могут быть размещены в узле. Выполнено фрагментирование РБД, в результате которого определено множество $F = \{F = F_1, \dots, F_N, N = |F|\}$ из N фрагментов таблиц РБД. Для каждого n -го фрагмента подсчитан его объем FV_n , равный произведению длины записей фрагмента на его мощность (количество записей). Выполнено первоначальное размещение фрагментов по узлам сети, которое задается матрицей $Y = \{y_{nm}\}$, где $y_{nm} = 1$, если фрагмент F_1 размещен в узле N_m , и равен 0 в противном случае. При первоначальном размещении в различных узлах сети может быть размещено несколько экземпляров одного фрагмента. На любом узле сети может выполняться один или несколько из K различных процессов (программ) $P = \{P_1, \dots, P_K\}$. Частота выполнения процессов в узлах сети задана матрицей $PF = \{PF_{km}\}$, где PF_{km} - частота выполнения процесса P_k в m -м узле сети. Определены матрица $RF = \{RF_{kn}\}$ и матрица $UF = \{UF_{kn}\}$. Элементы которых для каждого процесса P_k определяют частоты RF_{kn} запросов на извлечение данных из фрагмента F_n и частоты UF_{kn} запросов на модификацию данных этого фрагмента. Размещение фрагментов в узлах сети задается матрицей $FP = \{FP_{nm}\}$, элемент FP_{nm} которой равен 1 , если фрагмент F_n размещен в узле N_m , и равен 0 в противном случае.

Для упрощения модели предполагается, что сеть является полносвязной и задана матрица $CC = \{C_{ij}\}$ стоимости C_{ij} передачи единицы данных между всеми парами узлов сети N_i, N_j ($i, j = \overline{1, M}$). На основе матрицы смежности (топология сети задана) и матрицы стоимости CC передачи данных, используя любой из известных алгоритмов поиска кратчайших путей, строится матрица $SWC = \{SWC_{ij}\}$, $i, j = \overline{1, M}$ затрат на передачу данных по кратчайшим путям между узлами сети.

На начальной фазе и в процессе перераспределения фрагментов должны выполняться следующие ограничения:

$$\sum_{m=1}^M y_{nm} \geq 1, \quad 1 \leq n \leq N \quad (1)$$

Ограничение (1) определяет, что на начальной фазе каждый фрагмент F_n должен быть распределен, по меньшей мере, в одном узле сети N_m .

$$\sum_{n=1}^N FP_{nm} * FV_n \leq C_m, \quad 1 \leq m \leq M \quad (2)$$

Ограничение (2) означает, что суммарный объем фрагментов, размещенных в одном узле m , не может превышать емкость C_m этого узла.

$$\sum_{n=1}^N FP_{nm} \leq C_m, \quad 1 \leq m \leq M \quad (3)$$

Ограничение (3) гарантирует, что каждый узел m не будет хранить больше, чем определенное количество фрагментов FM_m .

$$\sum_{m=1}^M FP_{nm} = 1, \quad 1 \leq n \leq N \quad (4)$$

И ограничение (4) обеспечивает, что на этапе перераспределения каждый фрагмент будет распределен только в одном узле сети.

В предлагаемой модели используется параметр $FW = FW(N_j, F_i)$ - «вес» фрагмента F_i для узла N_j , который равен затратам на передачу по каналам сети запросов. Вес рассчитывается для каждого фрагмента F_i и каждого узла N_j , в

котором генерируются запросы к фрагменту F_i , в соответствии с формулой:

$$FW(N_j, F_i) = \sum_{m=1}^M QT_{mji} * SWC_{jm}, \quad 1 \leq j \leq M,$$
$$QT_{mji} = \sum_{k=1}^K PF_{km} * (RF_{km} + UF_{km}), \quad 1 \leq m \leq M$$

Здесь переменная QT_{mji} это общее количество запросов, генерируемых в каждом m -м узле сети к данным фрагмента F_i , размещенного в узле N_j .

Параметр FW используется в случае, когда из нескольких узлов поступают запросы к данным одного и того же фрагмента. Чтобы избежать дублирования информации, фрагмент F_i размещается в том узле, который имеет наибольшее значение этого параметра.

Величина затрат состоит из стоимости обработки информационных запросов и запросов на модификацию фрагментов.

В процессе работы алгоритма формируются несколько матриц, которые содержат следующую важную информацию:

- Частоты возникновения в сети запросов на обновление данных фрагментов.
- Частоты запросов выборку данных из фрагментов.
- Затраты на обновление данных фрагментов.
- Затраты на обработку запросов на доступ к данным фрагментов.

На основании данных этих матриц производится выбор узлов сети для перераспределения в них фрагментов из других узлов, подлежащих перераспределению.

С заданным определенным интервалом времени производится подсчет общего количества всех запросов, поступающих к фрагментам таблиц данных, расположенных в отдельных узлах системы, за фиксированный промежуток времени. Если количество запросов превышает определенный для каждого узла верхний предел, то производится перераспределение (миграция) одного или нескольких фрагментов, размещенных в данном узле, на один или несколько других узлов сети.

Данная работа подготовлена в рамках программы Президиума РАН № 30 (7) «Теория и технологии многоуровневого децентрализованного группового управления в условиях конфликта и кооперации»

Литература:

1. *Ozsu M. T., Valduriez P.* Principles of distributed database systems: Third Edition // Springer, NY, USA: Springer Science+Business Media. – 2011. — 866 p.
 2. *Чернышев Г.А.* Обзор подходов к организации физического уровня в СУБД // Труды СПИИРАН. 2013. – 2013. – Вып. 1(24). – С. 222 – 275.
 3. *Микрин Е.А., Сомов С.К.* Обзор моделей и методов обеспечения сохранности данных в распределенных системах обработки данных // Информационные технологии и вычислительные системы. – 2017. – № 4. – С. 5–28.
 4. *Сомов С.К.* Сохранность информации в распределенных системах обработки данных. – М.: ИПУ РАН, 2019. – 254 с.
 5. *Сомов С.К.* Репликация как инструмент повышения надежности функционирования распределенных систем // Информационные технологии и вычислительные системы. – 2018. – С. 69–79.
 6. *Hassan I.A.* A New Data Re-Allocation Model for Distributed Database Systems // International Journal of Database Theory and Application. – 2012. – №2. –Vol.5. – P. 45–60.
-