УДК004.657

ВЫПОЛНЕНИЕ ОПЕРАЦИИ ТЕТА-СОЕДИНЕНИЯ НА ОСНОВЕ СЖАТЫХ КОЛОНОЧНЫХ ИНДЕКСОВ С ИСПОЛЬЗОВАНИЕМ МНОГОЯДЕРНЫХ УСКОРИТЕЛЕЙ

Е.В. Иванова

Статья посвящена вопросу декомпозиции реляционной операции тета-соединения путем использования распределенных колоночных индексов с доменно-интервальной фрагментацией. Такая декомпозиция позволяет организовать эффективное параллельное выполнение запросов к сверхбольшим базам данных на современных кластерных вычислительных системах, оснащенных многоядерными ускорителями. Все фрагменты колоночного индекса хранятся в оперативной памяти в сжатом виде. При параллельном выполнении реляционной операции упакованные фрагменты индексов входных отношений загружаются на различные процессорные ядра, где происходят их распаковка, выполнение реляционной операции над фрагментами и упаковка частичного результата, представляющего собой наборы ключей. Затем частичные результаты объединяются в результирующий набор ключей, с использованием которого СУБД собирает результирующее отношение.

Ключевые слова: сверхбольшие базы данных, параллельная обработка запросов, колоночные индексы, доменно-интервальная фрагментация, декомпозиция реляционных операций, тета-соединения.

Введение

В настоящее время фактически единственным эффективным решением проблемы хранения и обработки сверхбольших баз данных является использование параллельных систем баз данных, обеспечивающих распределенную обработку запросов на многопроцессорных вычислительных системах с распределенной памятью [1–5].

В последние годы основным способом наращивания производительности процессоров является увеличение количества ядер, а не тактовой частоты, и эта тенденция, вероятно, сохранится [6]. Сегодня GPU (Graphic Processing Units) и Intel MIC (Many Integrated Cores) значительно опережают традиционные процессоры в производительности по арифметическим операциям и пропускной способности памяти, позволяя использовать сотни процессорных ядер для выполнения десятков тысяч потоков. Последние исследования показывают, что многоядерные ускорители могут эффективно использоваться для обработки запросов к базам данных в оперативной памяти [7–9].

В соответствие с этим актуальной является задача разработки новых эффективных методов параллельной обработки база данных в оперативной памяти на современных многопроцессорных вычислительных системах с многоядерными ускорителями. Для решения этой задачи в работах [10, 11] были предложены индексные структуры специального вида, которые называются распределенными колоночными индексами. Распределенные колоночные индексы позволяют провести декомпозицию реляционных операций, допускающую их эффективное параллельное выполнение на кластерных вычислительных системах с многоядерными ускорителями. В данной работе рассмотрен вопрос декомпозиции реляционной операции тета-соединения. Для обозначения реляционных операций в статье используется нотация, заимствованная из монографии [12]. Символом $\circ$ обозначается конкатенация кортежей.

1. Колоночный индекс и доменно-интервальная фрагментация

Под $R\left(A^{*}, B_{1}, \ldots, B_{u}\right)$ будем понимать отношение R с первичным ключом A и атрибутами $B_{1}, \ldots, B_{u}$, представляющее собой множество кортежей длины $u+1$ вида $(a, b_{1}, \ldots, b_{u})$, где $a \in \mathbb{Z}_{\geq 0}$ и $\forall j \in \{1, \ldots, u\}\left(b_{j} \in \mathfrak{D}_{B_{j}}\right)$. Здесь $\mathfrak{D}_{B_{j}}$ - домен атрибута B_{j}. Через $r.B_{j}$ будем обозначать значение атрибута B_{j}, через $r.A$ – значение *первичного ключа* в кортеже r: $r = (r.A, r.B_{1}, \ldots, r.B_{u})$. *Первичный ключ* отношения R обладает свойством $\forall r', r'' \in R\left(r' \neq r'' \Leftrightarrow r'.A \neq r''.A\right)$. Под *адресом кортежа* r мы будем понимать значение первичного ключа этого кортежа.

Пусть задано отношение $R\left(A^{*}, B, \ldots\right)$, $T\left(R\right) = n$. Пусть на множестве $\mathfrak{D}_{B}$ задано отношение линейного порядка. *Колоночным индексом* $I_{R.B}$ атрибута B отношения R называется упорядоченное отношение $I_{R.B}\left(A^{*}, B\right)$, удовлетворяющее следующим требованиям:

$$T(I_{R.B}) = n \text{ и } \pi_{A}\left(I_{R.B}\right) = \pi_{A}\left(R\right); \tag{1}$$

$$\forall x_{1}, x_{2} \in I_{R.B}\left(x_{1} \leq x_{2} \Leftrightarrow x_{1}.B \leq x_{2}.B\right); \tag{2}$$

$$\forall r \in R\left(\forall x \in I_{R.B}\left(r.A = x.A \Rightarrow r.B = x.B\right)\right). \tag{3}$$

Условие (1) означает, что множества значений первичных ключей (адресов) индекса и индексируемого отношения совпадают. Условие (2) означает, что элементы индекса упорядочены в порядке возрастания значений атрибута B. Условие (3) означает, что атрибут A элемента индекса содержит адрес кортежа отношения R, имеющего такое же значение атрибута B, как и у данного элемента колоночного индекса.

Теорема 1. Пусть задано отношение $R\left(A^*, B, \ldots\right)$. Пусть для отношения R задан колоночный индекс $I_{R.B}$. Тогда

$$\pi_B\left(I_{R.B}\right) = \pi_B\left(R\right). \tag{4}$$

Другими словами, колоночный индекс $I_{R.B}$ представляет все множество значений атрибута B отношения R с учетом повторяющихся значений.

Доказательство. Возьмем произвольное $b \in \mathfrak{D}_B$. Пусть $T\left(\sigma_{B=b}\left(R\right)\right) = k$. Без ограничения общности мы можем считать, что $\forall r \in R\left(r.A < k \Leftrightarrow r.B = b\right)$. Тогда из (1) и (3) следует, что $\forall x \in I_{R.B}\left(x.A < k \Leftrightarrow x.B = b\right)$. Откуда получаем $T\left(\sigma_{B=b}\left(I_{R.B}\right)\right) = k$. Таким образом (4) имеет место. *Теорема доказана*.

Пусть на множестве значений домена $\mathfrak{D}_B$ задано отношение линейного порядка. Пусть также задано разбиение множества $\mathfrak{D}_B$ на $k > 0$ непересекающихся интервалов:

$$\left.\begin{aligned} & V_0 = \left[v_0; v_1\right], V_1 = \left(v_1; v_2\right], \ldots, V_{k-1} = \left(v_{k-1}; v_k\right]; \\ & v_0 < v_1 < \ldots < v_k; \\ & \mathfrak{D}_B = \bigcup_{i=0}^{k-1} V_i. \end{aligned}\right\}. \tag{5}$$

Функция $\varphi_{\mathfrak{D}_B} : \mathfrak{D}_B \to \left\{0, \ldots, k-1\right\}$ называется *интервальной функцией фрагментации* для домена $\mathfrak{D}_B$, если она удовлетворяет следующему условию:

$$\forall i \in \{0, \ldots, k-1\}\left(\forall b \in \mathfrak{D}_B\left(\varphi_{\mathfrak{D}_B}(b) = i \Leftrightarrow b \in V_i\right)\right). \tag{6}$$

Пусть задан колоночный индекс $I_{R.B}$ для отношения $R\left(A^*, B, \ldots\right)$ с атрибутом B над доменом $\mathfrak{D}_B$ и интервальная функция фрагментации $\varphi_{\mathfrak{D}_B}$. Функция

$$\varphi_{I_{R.B}} : I_{R.B} \to \{0,\ldots,k-1\} \tag{7}$$

называется *доменно-интервальной функцией фрагментации* [2] для индекса $I_{R.B}$, если она удовлетворяет следующему условию:

$$\forall x \in I_{R.B} \left(\varphi_{I_{R.B}}(x) = \varphi_{\mathfrak{D}_B}(x.B) \right). \tag{8}$$

Определим i-тый фрагмент $(i = 0,\ldots,k-1)$ индекса $I_{R.B}$ следующим образом:

$$I^i_{R.B} = \left\{ x \mid x \in I_{R.B};\ \varphi_{I_{R.B}}(x) = i \right\}. \tag{9}$$

Это означает, что в i-тый фрагмент попадают кортежи, у которых значение атрибута B принадлежит i-тому доменному интервалу. Будем называть фрагментацию, построенную таким образом, *доменно-интервальной*. Количество фрагментов k будем называть *степенью фрагментации*.

Доменно-интервальная фрагментация обладает следующими фундаментальными свойствами, вытекающими непосредственно из ее определения:

$$I_{R.B} = \bigcup_{i=0}^{k-1} I^i_{R.B}; \tag{10}$$

$$\forall i, j \in \{0,\ldots,\mathrm{k}-1\} \left(i \neq j \Rightarrow I^i_{R.B} \cap I^j_{R.B} = \varnothing \right). \tag{11}$$

Теорема 2. Пусть для колоночного индекса $I_{R.B}$ отношения $R\left(A^*, B, \ldots\right)$ задана доменно-интервальная фрагментация степени k. Тогда

$$\forall i \in \{0,\ldots,\mathrm{k}-1\} \left(\forall x \in I_{R.B} \left(x \in I^i_{R.B} \Leftrightarrow x.B \in V_i \right) \right). \tag{12}$$

Доказательство. Сначала докажем, что

$$\forall i \in \{0,\ldots,\mathrm{k}-1\} \left(\forall x \in I_{R.B} \left(x \in I^i_{R.B} \Rightarrow x.B \in V_i \right) \right). \tag{13}$$

Пусть $x \in I^i_{R.B}$. Тогда из (9) следует $\varphi_{I_{R.B}}(x) = i$. С учетом (8) получаем $\varphi_{\mathfrak{D}_B}(x.B) = i$. Отсюда и из (6) следует $x.B \in V_i$, то есть (13) имеет место. Теперь докажем, что

$$\forall i \in \{0,\ldots,\mathrm{k}-1\} \left(\forall x \in I_{R.B} \left(x.B \in V_i \Rightarrow x \in I^i_{R.B} \right) \right). \tag{14}$$

Пусть $x \in I_{R.B}$ и $x.B \in V_i$. Тогда из (6) следует, что $\varphi_{\mathfrak{D}_B}(x.B) = i$. С учетом (8) получаем $\varphi_{\mathfrak{D}_B}(x.B) = \varphi_{I_{R.B}}(x) = i$. Отсюда и из (9) следует, что $x \in I^i_{R.B}$, то есть (14) имеет место. *Теорема доказана*.

2. Декомпозиция операции тета-соединения

Пусть заданы два отношения $R\left(A^*, B_1, \ldots, B_u\right)$ и $S\left(A^*, C_1, \ldots, C_v\right)$. Пусть также заданы функции $f : \pi_{B_1,\ldots,B_u}(R) \to \mathbb{R}$ и $g : \pi_{C_1,\ldots,C_v}(S) \to \mathbb{R}$. Определим $\bar{R} = \pi_{A, f(r.B_1,\ldots,r.B_u) \to F}(R)$ и $\bar{S} = \pi_{A, g(s.C_1,\ldots,s.C_v) \to G}(S)$. Пусть имеются колоночные индексы $I_{\bar{R}.F}$ и $I_{\bar{S}.G}$. Пусть для этих индексов задана доменно-интервальная фрагментация степени k:

$$I_{\bar{R}.F} = \bigcup_{i=0}^{k-1} I_{\bar{R}.F}^{i} \; ; I_{\bar{S}.G} = \bigcup_{i=0}^{k-1} I_{\bar{S}.G}^{i} \, . \tag{15}$$

Положим

$$\tilde{P}_i = \pi_{I_{\bar{R}.F}^{i}.A \to A_R,\, I_{\bar{S}.G}^{i}.A \to A_S}\left(I_{\bar{R}.F}^{i} \underset{F < G}{\bowtie} I_{\bar{S}.G}^{i}\right) \tag{16}$$

Определим

$$M_i = \pi_{A \to A_R}\left(I_{\bar{R}.F}^{i}\right); \tag{17}$$

$$N_i = \pi_{A \to A_S}\left(I_{\bar{S}.G}^{i}\right) \tag{18}$$

для всех $i = 0, \ldots, k-1$. и

$$\bar{P}_i = M_i \times \bigcup_{l=i+1}^{k-1} N_l \, . \tag{19}$$

Положим

$$P_i = \tilde{P}_i \cup \bar{P}_i; \tag{20}$$

$$P = \bigcup_{i=0}^{k-1} P_i \, . \tag{21}$$

Определим

$$Q = \left\{ r \circ s \mid r \in R \wedge s \in S \wedge (r.A, s.A) \in P \right\}. \tag{22}$$

Теорема 3. $\pi_{*\backslash R.A,S.A}(Q) = \pi_{*\backslash A}(R) \underset{f(r.B_1,\ldots,r.B_u) < g(s.C_1,\ldots,s.C_v)}{\bowtie} \pi_{*\backslash A}(S)$

Доказательство. Сначала докажем, что

$$\pi_{*\backslash R.A,S.A}(Q) \subset \pi_{*\backslash A}(R) \underset{f(r.B_1,\ldots,r.B_u) < g(s.C_1,\ldots,s.C_v)}{\bowtie} \pi_{*\backslash A}(S). \tag{23}$$

Пусть

$$(a, b_1, \ldots, b_u, a', c_1, \ldots, c_v) \in Q. \tag{24}$$

Из (22) следует, что существуют кортежи *r* и *s* такие, что

$$(a, b_1, \ldots, b_u) = r \in R, \tag{25}$$

$$(a', c_1, \ldots, c_v) = s \in S \tag{26}$$

и

$$(r.A, s.A) \in P. \tag{27}$$

Отсюда следует, что $\exists p \in P\left(p.A_R = a \wedge p.A_S = a'\right)$. Отсюда и из (21) получаем, что существует $i \in \{0, \ldots, k-1\}$, такой, что

$$p \in P_i \wedge p.A_R = a \wedge p.A_S = a'. \tag{28}$$

В силу (20) имеем $p \in \tilde{P}_i \vee p \in \overline{P}_i$. Рассмотрим сначала случай, когда $p \in \tilde{P}_i$. Тогда из (28) и (15)–(16) следует, что

$$\exists x \in I_{\bar{R}.F}\left(\exists y \in I_{\bar{S}.G}\left(x.A = a \wedge x.F < y.G \wedge y.A = a'\right)\right).$$

По определению колоночного индекса отсюда получаем

$$\exists \tilde{r} \in R\left(\exists \tilde{s} \in S\left(\tilde{r}.A = a \wedge \tilde{s}.A = a' \wedge f(\tilde{r}.B_1, \ldots, \tilde{r}.B_u) < g(\tilde{s}.C_1, \ldots, \tilde{s}.C_v)\right)\right). \tag{29}$$

Поскольку A является первичным ключом в R и S, с учетом (25) и (26) отсюда следует, что $\tilde{r} = (a, b_1, \ldots, b_u)$, $\tilde{s} = (a', c_1, \ldots, c_v)$ и $(b_1, \ldots, b_u, c_1, \ldots, c_v) \in \pi_{*\backslash A}(R) \underset{f(r.B_1, \ldots, r.B_u) < g(s.C_1, \ldots, s.C_v)}{\bowtie} \pi_{*\backslash A}(S)$. Таким образом, (23) имеет место.

Пусть теперь $p \in \overline{P}_i$. Тогда из (28), и (17)–(19) следует, что существуют $i \in \{0, \ldots, k-1\}$ и $l \in \{i+1, \ldots, k-1\}$, такие, что

$$\exists x \in I^i_{\bar{R}.F}\left(x.A = a\right), \tag{30}$$

$$\exists y \in I^l_{\bar{S}.G}\left(y.A = a'\right). \tag{31}$$

Так как $l > i$, то по определению колоночного индекса и доменно-интервальной фрагментации из (30) и (31) получаем, что

$$\exists x \in I_{\bar{R}.F}\left(\exists y \in I_{\bar{S}.G}\left(x.A = a \wedge y.A = a' \wedge x.F < y.G\right)\right),$$

откуда следует (29), и, таким образом, (23) также имеет место.

Теперь докажем, что

$$\pi_{*\backslash A}(R) \underset{f(r.B_1,\ldots,r.B_u) < g(s.C_1,\ldots,s.C_v)}{\bowtie} \pi_{*\backslash A}(S) \subset \pi_{*\backslash R.A,S.A}(Q). \quad (32)$$

Пусть

$$(a,b_1,\ldots,b_u) = r \in R, \quad (33)$$

$$(a',c_1,\ldots,c_v) = s \in S, \quad (34)$$

и

$$f(b_1,\ldots,b_u) < g(c_1,\ldots,c_v). \quad (35)$$

Тогда по определению колоночного индекса с учетом (4) имеем

$$\exists x \in I_{\bar{R}.F}\left(\exists y \in I_{\bar{S}.G}\left(x.A = a \wedge x.F = f(b_1,\ldots,b_u) < g(c_1,\ldots,c_v) = y.G \wedge y.A = a'\right)\right).$$

На основе (5)–(9) и (12) отсюда следует, что возможны только два следующих случая:

1) $\exists i \in \{0,\ldots,\mathrm{k}-1\}\left(x \in I^{i}_{\bar{R}.F} \wedge y \in I^{i}_{\bar{S}.G}\right);$ (36)

2) $\exists i \in \{0,\ldots,\mathrm{k}-1\}\left(\exists l \in \{i+1,\ldots,\mathrm{k}-1\}\left(x \in I^{i}_{\bar{R}.F} \wedge y \in I^{l}_{\bar{S}.G}\right)\right).$ (37)

Допустим сначала, что имеет место первый случай. Тогда с учетом (16) из (36) следует

$$\exists i \in \{0,\ldots,\mathrm{k}-1\}\left(\exists p \in \tilde{P}_i\left(p.A_R = a \wedge p.A_S = a'\right)\right).$$

Учитывая (20) и (21) отсюда имеем $(a,a') \in P$. Вместе с (22), (33) и (34) это дает

$$(a,b_1,\ldots,b_u,a',c_1,\ldots,c_v) \in Q,$$

то есть (32) имеет место.

Теперь допустим, что имеет место второй случай. Тогда с учетом (19) из (37) следует

$$\exists i \in \{0,\ldots,\mathrm{k}-1\}\left(\exists p \in \bar{P}_i\left(p.A_R = a \wedge p.A_S = a'\right)\right).$$

Учитывая (20) и (21) отсюда имеем $(a,a') \in P$. Вместе с (22), (33) и (34) это дает

$$(a,b_1,\ldots,b_u,a',c_1,\ldots,c_v) \in Q,$$

то есть (32) имеет место. *Теорема доказана*.

Заключение

В статье было представлено формальное описание метода декомпозиции операции тета-соединения по неравенству на основе доменно-интервальной фрагментации колоночных индексов. Доказана корректность метода. Практическая ценность предложенного метода декомпозиции заключается в том, что ресурсоемкие вычисления могут производиться независимо над соответствующими фрагментами. Для тета-соединения это – вычисление $\tilde{P}_i$ по формуле (16).

Работа выполнена при финансовой поддержке Минобрнауки РФ в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014—2020 годы» (Госконтракт № 14.574.21.0035).

Библиографический список

1. Соколинский, Л.Б. Параллельные машины баз данных / Л.Б. Соколинский // Природа. — 2001. — № 8. — С. 10–17.

2. Соколинский, Л.Б. Параллельные системы баз данных / Л.Б. Соколинский — Москва: Издательство Московского государственного университета, 2013. — 184 с.

3. Sokolinsky, L.B. Design and Evaluation of Database Multiprocessor Architecture with High Data Availability / L.B. Sokolinsky // Proceedings of the 12th International workshop on database and expert systems applications. — IEEE Computer Society, 2001. — P. 115–120.

4. Pan, C.S. Taming Elephants, or How to Embed Parallelism into PostgreSQL / C.S. Pan, M.L. Zymbler // Lecture Notes in Computer Science. — 2013. — Vol. 8055, Part 1. — P. 153–164.

5. Костенецкий, П.С. Моделирование иерархических многопроцессорных систем баз данных / П.С. Костенецкий, Л.Б. Соколинский // Программирование. — 2013. — Т. 39, № 1. — С. 3–22.

6. Fang, J. Sesame: A User-Transparent Optimizing Framework for Many-Core Processors / J. Fang, A.L. Varbanescu, H. Sips // Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid2013), May 13–16, 2013, Delft, Netherlands. — IEEE, 2013. — P. 70–73.

7. Breß, S. Efficient Co-Processor Utilization in Database Query Processing / S. Breß, F. Beier, H. Rauhe, et al. // Information Systems. — 2013. — Vol. 38, No. 8. — P. 1084–1096.

8. Scherger, M. Design of an In-Memory Database Engine Using Intel Xeon Phi Coprocessors / M. Scherger // Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA’14), July 21–24, 2014, Las Vegas, USA. — CSREA Press, 2014. — P. 21–27.

9. Беседин, К.Ю. Моделирование обработки запросов на гибридных вычислительных системах с многоядерными сопроцессорами и графическими ускорителями / К.Ю. Беседин, П.С. Костенецкий // Программные системы: теория и приложения. — 2014. — Т. 5, № 1-1 (19). — С. 91–110.

10. Иванова, Е.В. Использование распределенных колоночных индексов для выполнения запросов к сверхбольшим базам данных / Е.В. Иванова, Л.Б. Соколинский // Параллельные вычислительные технологии (ПАВТ'2014).

Труды международной научной конференции. — Челябинск: Издательский центр ЮУрГУ, 2014. — С. 270–275.

11. Иванова, Е.В. Использование распределенных колоночных хеш-индексов для обработки запросов к сверхбольшим базам данных / Е.В. Иванова // Научный сервис в сети Интернет: многообразие суперкомпьютерных миров. Труды Международной суперкомпьютерной конференции. — М.: Изд-во МГУ, 2014. — С. 102–104.

12. Гарсиа-Молина, Г. Системы баз данных. Полный курс. / Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом — М.: Издательский дом «Вильямс». — 2004. — 1088 с.