

ПАРАЛЛЕЛЬНАЯ ДЕКОМПОЗИЦИЯ РЕЛЯЦИОННЫХ ОПЕРАЦИЙ НА ОСНОВЕ РАСПРЕДЕЛЕННЫХ КОЛОНОЧНЫХ ИНДЕКСОВ

Е.В. Иванова, Л.Б. Соколинский

Данная статья является продолжением и развитием более ранней работы авторов, в которой была рассмотрена декомпозиция операций пересечения и соединения колоночных индексов на основе доменно-интервальной фрагментации. Такая декомпозиция позволяет организовать параллельное выполнение реляционных операций над распределенными колоночными индексами без массовых обменов данными между процессорными узлами. В настоящей статье рассматривается декомпозиция операций проекции, выбора, удаления дубликатов и объединения. Кроме этого, вводится новый вид колоночных индексов, названных колоночными хеш-индексами. Колоночный хеш-индекс способен индексировать сразу несколько атрибутов отношения. Для распределенных колоночных хеш-индексов рассматривается декомпозиция операций пересечения, объединения и естественного соединения.

Ключевые слова: распределенные колоночные индексы, доменно-интервальная фрагментация, колоночные хеш-индексы, декомпозиция реляционных операций.

Введение

В последнее время большой интерес у исследователей вызывают системы баз данных с хранением баз данных по столбцам. Такое представление данных называется *колоночным* [1, 2]. Колоночное представление в отличие от традиционного строкового представления оказывается намного более эффективным при выполнении запросов класса OLAP [3]. Это объясняется тем, что при выполнении запросов колоночная СУБД считывает с диска только те атрибуты, которые необходимы для выполнения запроса, что сокращает объем операций ввода-вывода и, как следствие, уменьшает время выполнения запроса. Недостатком колоночного представления является низкая эффективность при выполнении строково-ориентированных операций, таких, например, как добавление или удаление кортежей. Вследствие этого колоночные СУБД могут проигрывать по производительности строковым при выполнении запросов класса OLTP. Одним из главных преимуществ строчных хранилищ является наличие в строковых СУБД мощных процедур оптимизации запросов, разработанных на базе реляционной модели. Строковые СУБД также имеют большое преимущество в скорости обработки запросов класса OLTP. В соответствии с этим в исследовательском сообществе баз данных были приняты интенсивные усилия по интеграции преимуществ столбцовой модели хранения данных в строковые СУБД. Анализ рассмотренных решений показывает [3], что нельзя получить выгоду от хранения данных по столбцам, воспользовавшись системой баз данных со строковым хранением с вертикально разделенной схемой, либо проиндексировав все столбцы, чтобы обеспечить к ним независимый доступ.

Авторами в работах [4–10] был предложен новый подход к интеграции преимуществ строчных и колоночных хранилищ, суть которого заключается во введении распреде-

ленных колоночных индексов, хранимых в оперативной памяти кластерной вычислительной системы. Эти колоночные индексы обрабатываются с помощью программной системы, получившей название «колоночный сопроцессор КСОП». Назначение КСОП — вычислять таблицы предварительных вычислений для ресурсоемких реляционных операций по запросу СУБД. Общая схема взаимодействия СУБД и КСОП изображена на рис. 1.

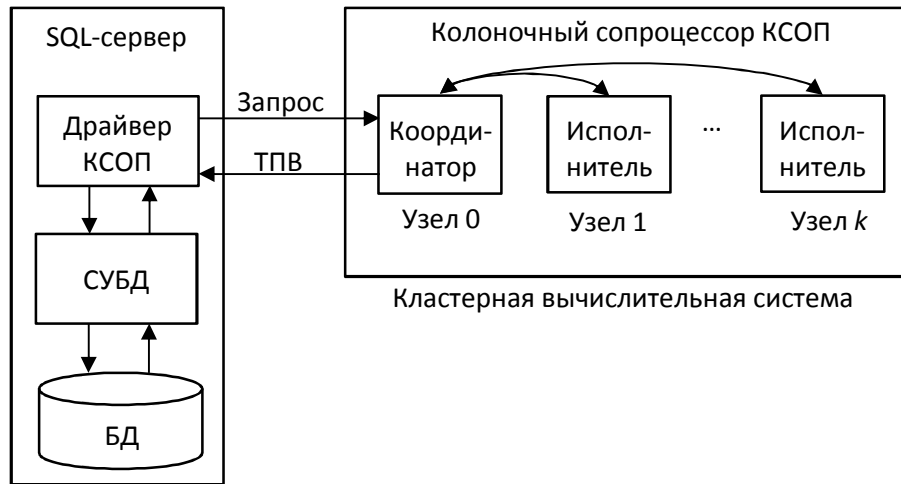


Рис. 1. Взаимодействие SQL-сервера с колоночным сопроцессором КСОП

КСОП включает в себя программу «Координатор», запускаемую на узле вычислительного кластера с номером 0, и программу «Исполнитель», запускаемую на всех остальных узлах, выделенных для работы КСОП. На SQL-сервере устанавливается специальная программа «Драйвер КСОП», обеспечивающая взаимодействие с координатором КСОП по протоколу TCP/IP.

Теоретические основы колоночного сопроцессора были изложены авторами в работе [4], где были описаны колоночные индексы и доменно-интервальная фрагментация. Суть математической модели заключается в следующем. Пусть задано отношение $R(A, B, \dots)$, $T(R) = n$. Атрибут A в отношении R играет роль суррогатного ключа. Пусть на множестве \mathcal{D}_B , являющемся доменом атрибута B , задано отношение линейного порядка. Колоночным индексом $I_{R,B}$ атрибута B отношения R будем называть упорядоченное отношение $I_{R,B}(A, B)$, удовлетворяющее следующим свойствам:

$$T(I_{R,B}) = n \text{ и } \pi_A(I_{R,B}) = \pi_A(R); \quad (1)$$

$$\forall x_1, x_2 \in I_{R,B} (x_1 \leq x_2 \Leftrightarrow x_1.B \leq x_2.B); \quad (2)$$

$$\forall r \in R (\forall x \in I_{R,B} (r.A = x.A \Rightarrow r.B = x.B)). \quad (3)$$

С содержательной точки зрения колоночный индекс $I_{R,B}$ представляет собой таблицу из двух колонок с именами A и B . Количество строк в колоночном индексе совпадает с количеством строк в индексируемой таблице. Колонка B индекса $I_{R,B}$ включает в себя все значения колонки B таблицы R , отсортированных в порядке возрастания. Каждая строка x индекса $I_{R,B}$ содержит в колонке A суррогатный ключ (адрес) строки r в таблице R , имеющей такое же значение в колонке B , что и строка x . На рис. 2 представлены примеры двух различных колоночных индексов для одного и того же отношения.

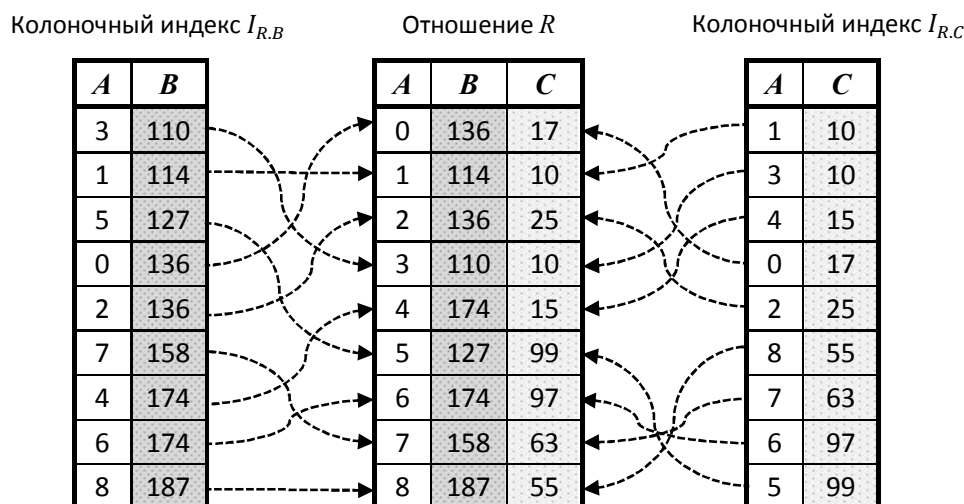


Рис. 2. Колоночные индексы

Для фрагментации колоночных индексов по узлам многопроцессорной системы авторами был предложен метод доменно-интервальной фрагментации [4], суть которого заключается в следующем. Пусть на множестве значений домена \mathcal{D}_B задано отношение линейного порядка. Разобьем множество \mathcal{D}_B на $k > 0$ непересекающихся интервалов:

$$\left. \begin{aligned} &V_0 = [v_0; v_1), V_1 = [v_1; v_2), \dots, V_{k-1} = [v_{k-1}; v_k); \\ &v_0 < v_1 < \dots < v_k; \\ &\mathcal{D}_B = \bigcup_{i=0}^{k-1} V_i. \end{aligned} \right\} \quad (4)$$

Функция $\varphi_{\mathcal{D}_B} : \mathcal{D}_B \rightarrow \{0, \dots, k-1\}$ называется *доменной функцией фрагментации* для \mathcal{D}_B , если она удовлетворяет следующему условию:

$$\forall i \in \{0, \dots, k-1\} (\forall b \in \mathcal{D}_B (\varphi_{\mathcal{D}_B}(b) = i \Leftrightarrow b \in V_i)). \quad (5)$$

Пусть задан колоночный индекс $I_{R,B}$ для отношения $R(A, B, \dots)$ с атрибутом B над доменом \mathcal{D}_B и доменная функция фрагментации $\varphi_{\mathcal{D}_B}$. Функция

$$\varphi_{I_{R,B}} : I_{R,B} \rightarrow \{0, \dots, k-1\}, \quad (6)$$

определенная по правилу

$$\forall x \in I_{R,B} (\varphi_{I_{R,B}}(x) = \varphi_{\mathcal{D}_B}(x.B)), \quad (7)$$

называется *доменно-интервальной функцией фрагментации* для индекса $I_{R,B}$. Другими словами, функция фрагментации $\varphi_{I_{R,B}}$ сопоставляет каждому кортежу x из $I_{R,B}$ номер доменного интервала, которому принадлежит значение $x.B$.

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

$$I_{R,B}^i = \{x \mid x \in I_{R,B}; \varphi_{I_{R,B}}(x) = i\}. \quad (8)$$

Это означает, что в i -тый фрагмент попадают кортежи, у которых значение атрибута B принадлежит i -тому доменному интервалу. Фрагментация, построенная таким образом, называется *доменно-интервальной*. Количество фрагментов k называется *степенью*

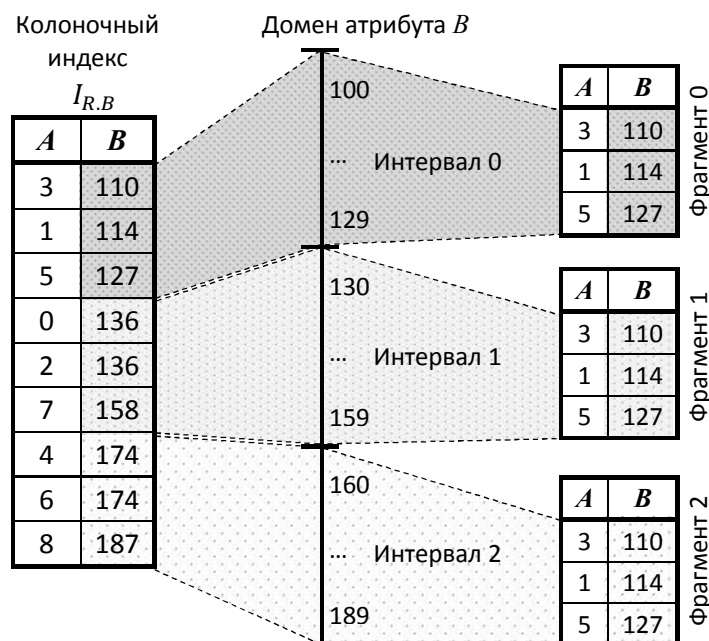


Рис. 3. Фрагментация колоночного индекса

фрагментации. На рис. 3 схематично изображена доменно-интервальная фрагментация колоночного индекса, имеющая степень $k = 3$.

Пусть для отношения $R(A, B, C, \dots)$ заданы колоночные индексы $I_{R,B}$ и $I_{R,C}$. Транзитивной фрагментацией индекса $I_{R,C}$ относительно индекса $I_{R,B}$ называется фрагментация, задаваемая функцией $\check{\varphi}_{I_{R,C}} : I_{R,C} \rightarrow \{0, \dots, k-1\}$, удовлетворяющей условию $\forall x \in I_{R,C} :$

$$\check{\varphi}_{I_{R,C}}(x) = \varphi_{I_{R,B}}(\sigma_{A=x.A}(I_{R,B})). \quad (9)$$

Транзитивная фрагментация позволяет разместить на одном и том же узле элементы колоночных индексов, соответствующие одному кортежу индексируемого отношения.

В работе в работе [4] авторами была выполнена декомпозиция операций пересечения и соединения, в [8] — операции группировки. В настоящей статье рассматривается декомпозиция операций проекции, выбора, удаления дубликатов и объединения. Кроме этого, вводится новый вид колоночных индексов, названных колоночными хеш-индексами. Колоночный хеш-индекс способен индексировать сразу несколько атрибутов отношения. Для распределенных колоночных хеш-индексов рассматривается декомпозиция операций пересечения, объединения и естественного соединения.

Статья имеет следующую структуру. В разделе 1 рассматривается декомпозиция операций проекции, выбора, удаления дубликатов и объединения с использованием фрагментированных колоночных индексов. В разделе 2 вводится понятие колоночного хеш-индекса и декомпозиция операций пересечения, объединения и естественного соединения с использованием распределенных хеш-индексов. В заключении суммируются полученные результаты, делаются итоговые выводы.

1. Декомпозиция реляционных операций с использованием фрагментированных колоночных индексов

В данном разделе рассматривается декомпозиция операций проекции, выбора, удаления дубликатов и объединения с использованием фрагментированных колоночных индексов.

1.1. Декомпозиция проекции

В этом разделе рассматривается декомпозиция операции проекции вида $\pi_{B,C_1,\dots,C_u}(R)$. Алгоритм выполнения проекции в строчных СУБД не содержит ресурсоемких вычислений, однако он требует чтения с диска всего отношения. В отличие от этого, предлагаемый ниже алгоритм использует только те столбцы (колоночные индексы) отношения, которые вовлекаются в проекцию. Кроме этого, он вообще не предполагает чтений с диска, так как колоночные индексы хранятся в оперативной памяти.

Пусть задано отношение $R(A,B,C_1,\dots,C_u,\dots)$ с суррогатным ключом A . Пусть имеется колоночный индекс $I_{R,B}(A,B)$. Пусть также имеются колоночные индексы $I_{R,C_1}(A,C_1), \dots, I_{R,C_u}(A,C_u)$. Пусть для индекса $I_{R,B}$ задана доменно-интервальная фрагментация степени k :

$$I_{R,B} = \bigcup_{i=0}^{k-1} I_{R,B}^i. \quad (10)$$

Пусть для индексов $I_{R,C_1}, \dots, I_{R,C_u}$ задана транзитивная относительно $I_{R,B}$ фрагментация:

$$\forall j \in \{1, \dots, u\} \left(I_{R,C_j} = \bigcup_{i=0}^{k-1} I_{R,C_j}^i \right). \quad (11)$$

Положим

$$P_i = I_{R,B}^i \bowtie I_{R,C_1}^i \bowtie \dots \bowtie I_{R,C_u}^i \quad (12)$$

для всех $i = 0, \dots, k-1$. Определим

$$P = \bigcup_{i=0}^{k-1} P_i. \quad (13)$$

Построим отношение $Q(B,C_1,\dots,C_u)$ следующим образом:

$$Q = \pi_{B,C_1,\dots,C_u}(P). \quad (14)$$

Теорема 1. $Q = \pi_{B,C_1,\dots,C_u}(R)$.

Доказательство. Сначала докажем, что

$$Q \subset \pi_{B,C_1,\dots,C_u}(R). \quad (15)$$

Пусть $(b,c_1,\dots,c_u) \in Q$. В силу (14) и (13) существуют a и i такие, что $(a,b,c_1,\dots,c_u) \in P_i$.

В силу (12) имеем:

$$\begin{aligned} (a,b) &\in I_{R,B}; \\ (a,c_1) &\in I_{R,C_1}; \\ &\dots\dots\dots \\ (a,c_u) &\in I_{R,C_u}. \end{aligned}$$

По определению колоночного индекса (свойства (1) и (3)) отсюда следует, что существует $r \in R$ такой, что $r.A = a; r.B = b; r.C_1 = c_1; \dots; r.C_u = c_u$, откуда следует $(b, c_1, \dots, c_u) \in \pi_{B, C_1, \dots, C_u}(R)$, то есть (15) имеет место.

Теперь покажем, что

$$Q \supset \pi_{B, C_1, \dots, C_u}(R). \quad (16)$$

Пусть $(b, c_1, \dots, c_u) \in \pi_{B, C_1, \dots, C_u}(R)$. Тогда существует $r \in R$ такой, что $r.B = b; r.C_1 = c_1; \dots; r.C_u = c_u$. Положим $a = r.A$. По определению колоночного индекса (свойства (1) и (3)) отсюда следует что

$$\begin{aligned} (a, b) &\in I_{R.B}; \\ (a, c_1) &\in I_{R.C_1}; \\ &\dots\dots\dots \\ (a, c_u) &\in I_{R.C_u}. \end{aligned} \quad (17)$$

В силу свойств доменно-интервальной фрагментации существует i такой, что $(a, b) \in I_{R.B}^i$. По определению транзитивной относительно $I_{R.B}$ фрагментации отсюда, с учетом (17), следует

$$\begin{aligned} (a, c_1) &\in I_{R.C_1}^i; \\ &\dots\dots\dots \\ (a, c_u) &\in I_{R.C_u}^i. \end{aligned}$$

Тогда, в силу (12), $(a, b, c_1, \dots, c_u) \in P_i$. Учитывая (13), получаем $(a, b, c_1, \dots, c_u) \in P$, то есть $(b, c_1, \dots, c_u) \in \pi_{B, C_1, \dots, C_u}(P)$. Принимая во внимание (14), имеем $(b, c_1, \dots, c_u) \in Q$, что означает, что (16) имеет место. *Теорема доказана.*

1.2. Декомпозиция операции выбора

В данном разделе рассматривается декомпозиция операции выбора вида $\sigma_\theta(R)$, где θ — некоторое условие, накладываемое на атрибуты отношения R .

Пусть имеется отношение $R(A, B, C_1, \dots, C_u, D_1, \dots, D_w)$ с суррогатным ключом A . Рассмотрим операцию выбора вида $\sigma_{\theta(B, C_1, \dots, C_u)}(R)$, где $\theta(B, C_1, \dots, C_u)$ — некоторое условие, зависящее от значений атрибутов B, C_1, \dots, C_u отношения R . Пусть имеется колоночный индекс $I_{R.B}$. Пусть также имеются колоночные индексы:

$$I_{R.C_1}, \dots, I_{R.C_u}.$$

Пусть для индекса $I_{R.B}$ задана доменно-интервальная фрагментация степени k :

$$I_{R.B} = \bigcup_{i=0}^{k-1} I_{R.B}^i. \quad (18)$$

Пусть для индексов $I_{R.C_1}, \dots, I_{R.C_u}$ и $I_{R.D_1}, \dots, I_{R.D_w}$ задана транзитивная относительно $I_{R.B}$ фрагментация:

$$\forall j \in \{1, \dots, u\} \left(I_{R.C_j} = \bigcup_{i=0}^{k-1} I_{R.C_j}^i \right). \quad (19)$$

Положим

$$P_i = \pi_A \left(\sigma_{\theta(B, C_1, \dots, C_u)} \left(I_{R.B}^i \bowtie I_{R.C_1}^i \bowtie \dots \bowtie I_{R.C_u}^i \right) \right) \quad (20)$$

для всех $i = 0, \dots, k-1$. Определим

$$P = \bigcup_{i=0}^{k-1} P_i . \quad (21)$$

Построим отношение $Q(A, B, C_1, \dots, C_u, D_1, \dots, D_w)$ следующим образом:

$$Q = \{ \&_R(p.A) \mid p \in P \} . \quad (22)$$

Теорема 2. $Q = \sigma_{\theta(B, C_1, \dots, C_u)}(R)$.

Доказательство. Сначала докажем, что

$$Q \subset \sigma_{\theta(B, C_1, \dots, C_u)}(R) . \quad (23)$$

Пусть $(a, b, c_1, \dots, c_u, d_1, \dots, d_w) \in Q$. Тогда из (22) и (1) следует, что существует $r \in R$ такой, что

$$r.A = a; r.B = b; r.C_1 = c_1; \dots; r.C_u = c_u; r.D_1 = d_1; \dots; r.D_w = d_w . \quad (24)$$

С другой стороны, в силу (22) существует $p \in P$ такой, что $(a) \in P$. Тогда в силу (21) существует i такой, что $(a) \in P_i$. Отсюда, с учетом (20), существуют b', c'_1, \dots, c'_u такие, что

$$(a, b', c'_1, \dots, c'_u) \in \sigma_{\theta(B, C_1, \dots, C_u)}(I_{R.B}^i \bowtie I_{R.C_1}^i \bowtie \dots \bowtie I_{R.C_u}^i) . \quad (25)$$

Отсюда следует, что

$$\theta(b', c'_1, \dots, c'_u) = true \quad (26)$$

и

$$(a, b') \in I_{R.B};$$

$$(a, c'_1) \in I_{R.C_1};$$

.....

$$(a, c'_u) \in I_{R.C_u} .$$

По определению колоночного индекса (свойства (1) и (3)) отсюда следует, что существует $r' \in R$ такой, что $r'.A = a, r'.B = b', r'.C_1 = c'_1, \dots, r'.C_u = c'_u$. Принимая во внимание (26), тогда имеем $\theta(r'.B, r'.C_1, \dots, r'.C_u) = true$, то есть $r' \in \sigma_{\theta(B, C_1, \dots, C_u)}(R)$. Поскольку $r'.A = a = r.A$ и A — суррогатный ключ в R , с учетом (24), отсюда получаем $(a, b, c_1, \dots, c_u, d_1, \dots, d_w) \in \sigma_{\theta(B, C_1, \dots, C_u)}(R)$, то есть (23) имеет место.

Теперь покажем, что

$$Q \supset \sigma_{\theta(B, C_1, \dots, C_u)}(R) . \quad (27)$$

Пусть $(a, b, c_1, \dots, c_u, d_1, \dots, d_w) \in \sigma_{\theta(B, C_1, \dots, C_u)}(R)$. Тогда существует $r \in R$ такой, что

$$r.A = a; r.B = b; r.C_1 = c_1; \dots; r.C_u = c_u; r.D_1 = d_1; \dots; r.D_w = d_w \quad (28)$$

и

$$\theta(b, c_1, \dots, c_u) = true . \quad (29)$$

По определению колоночного индекса (свойства (1) и (3)) из (28) следует что

$$(a, b) \in I_{R.B};$$

$$(a, c_1) \in I_{R.C_1};$$

.....

$$(a, c_u) \in I_{R.C_u} .$$

В силу свойств доменно-интервальной фрагментации существует i такой, что $(a, b) \in I_{R.B}^i$. По определению транзитивной относительно $I_{R.B}$ фрагментации отсюда, с учетом (30), следует

$$(a, c_1) \in I_{R.C_1}^i;$$

.....

$$(a, c_u) \in I_{R.C_u}^i.$$

Тогда с учетом (29) из (20) получаем $(a) \in P_i$. Принимая во внимание (21), имеем $(a) \in P$. С учетом (22) и (1) отсюда следует, что существует $r' \in R$ такой, что

$$r'.A = a \tag{31}$$

и

$$(r'.A, r'.B, r'.C_1, \dots, r'.C_u, r'.D_1, \dots, r'.D_w) \in Q. \tag{32}$$

Так как A — суррогатный ключ, из (31) следует, что $r' = r$. Вместе с (32) и (28) это дает $(a, b, c_1, \dots, c_u, d_1, \dots, d_w) \in Q$, то есть (27) имеет место. Теорема доказана.

1.3. Декомпозиция операции удаления дубликатов

Пусть задано отношение $R(A, B, C_1, \dots, C_u)$ с суррогатным ключом A . Пусть имеется колоночный индекс $I_{R.B}$. Пусть также имеются колоночные индексы $I_{R.C_1}, \dots, I_{R.C_u}$. Пусть для индекса $I_{R.B}$ задана доменно-интервальная фрагментация степени k :

$$I_{R.B} = \bigcup_{i=0}^{k-1} I_{R.B}^i. \tag{33}$$

Пусть для индексов $I_{R.C_1}, \dots, I_{R.C_u}$ задана транзитивная относительно $I_{R.B}$ фрагментация:

$$\forall j \in \{1, \dots, u\} \left(I_{R.C_j} = \bigcup_{i=0}^{k-1} I_{R.C_j}^i \right). \tag{34}$$

Положим

$$P_i = \pi_A \left(\gamma_{\min(A) \rightarrow A, B, C_1, \dots, C_u} \left(I_{R.B}^i \bowtie I_{R.C_1}^i \bowtie \dots \bowtie I_{R.C_u}^i \right) \right) \tag{35}$$

для всех $i = 0, \dots, k-1$. Определим

$$P = \bigcup_{i=0}^{k-1} P_i. \tag{36}$$

Построим отношение $Q(B, C_1, \dots, C_u)$ следующим образом:

$$Q = \{ (\&_R(p.A).B, \&_R(p.A).C_1, \dots, \&_R(p.A).C_u) \mid p \in P \}. \tag{37}$$

Теорема 3. $Q = \delta(\pi_{B, C_1, \dots, C_u}(R))$.

Доказательство. Сначала докажем, что

$$Q \subset \delta(\pi_{B, C_1, \dots, C_u}(R)). \tag{38}$$

Пусть $(b, c_1, \dots, c_u) \in Q$. Тогда из (37) следует что существует a такой, что $(a) \in P$, и существует $r \in R$ такой, что

$$r.A = a; r.B = b; r.C_1 = c_1; \dots; r.C_u = c_u. \tag{39}$$

Следовательно $(b, c_1, \dots, c_u) \in \delta(\pi_{B, C_1, \dots, C_u}(R))$.

Теперь покажем, что

$$Q \supset \delta(\pi_{B, C_1, \dots, C_u}(R)). \tag{40}$$

Пусть

$$(b, c_1, \dots, c_u) \in \delta(\pi_{B, C_1, \dots, C_u}(R)).$$

Положим

$$R'(A, B, C_1, \dots, C_u) = \{r \mid r \in R \wedge r.B = b \wedge r.C_1 = c_1 \wedge \dots \wedge r.C_u = c_u\}. \quad (41)$$

Вычислим кортеж

$$r' = \gamma_{\min(A) \rightarrow A, B, C_1, \dots, C_u}(R'). \quad (42)$$

Очевидно, что

$$r' \in R \quad (43)$$

и

$$r'.B = b; r'.C_1 = c_1; \dots; r'.C_u = c_u. \quad (44)$$

Обозначим

$$a = r'.A. \quad (45)$$

По определению колоночного индекса (свойства (1) и (3)) из (43), (44) и (45) следует что

$$\begin{aligned} (a, b) &\in I_{R.B}; \\ (a, c_1) &\in I_{R.C_1}; \\ &\dots\dots\dots \\ (a, c_u) &\in I_{R.C_u}. \end{aligned}$$

В силу свойств доменно-интервальной фрагментации существует i такой, что $(a, b) \in I_{R.B}^i$. По определению транзитивной относительно $I_{R.B}$ фрагментации отсюда следует

$$\begin{aligned} (a, c_1) &\in I_{R.C_1}^i; \\ &\dots\dots\dots \\ (a, c_u) &\in I_{R.C_u}^i. \end{aligned}$$

Тогда с учетом (42) из (35) получаем $(a) \in P_i$. Принимая во внимание (36), имеем $(a) \in P$. С учетом (37) и (1) отсюда следует, что существует $r'' \in R$ такой, что

$$r''.A = a \quad (46)$$

и

$$(r''.B, r''.C_1, \dots, r''.C_u) \in Q. \quad (47)$$

Так как A — суррогатный ключ, из (45) и (46) следует, что $r' = r''$. Вместе с (44) и (47) это дает $(b, c_1, \dots, c_u) \in Q$, то есть (40) имеет место. Теорема доказана.

1.4. Декомпозиция операции объединения

В данном разделе рассматривается декомпозиция операции объединения двух отношений вида $\pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$. При этом предполагается, что $\pi_{B_1, \dots, B_u}(R)$ и $\pi_{B_1, \dots, B_u}(S)$ не содержат дубликатов. Результирующее отношение также не должно содержать дубликатов.

Пусть заданы два отношения $R(A, B_1, \dots, B_u)$ и $S(A, B_1, \dots, B_u)$, имеющие одинаковый набор атрибутов. Пусть имеется два набора колоночных индексов по атрибутам B_1, \dots, B_u :

$$\begin{aligned} I_{R.B_1}, \dots, I_{R.B_u}; \\ I_{S.B_1}, \dots, I_{S.B_u}. \end{aligned}$$

Пусть для индексов $I_{R.B_1}$ и $I_{S.B_1}$ задана доменно-интервальная фрагментация степени k :

$$I_{R.B_1} = \bigcup_{i=0}^{k-1} I_{R.B_1}^i; \quad (48)$$

$$I_{S.B_1} = \bigcup_{i=0}^{k-1} I_{S.B_1}^i. \quad (49)$$

Пусть для индексов $I_{R.B_2}, \dots, I_{R.B_u}$ и $I_{S.B_2}, \dots, I_{S.B_u}$ задана транзитивная фрагментация относительно $I_{R.B_1}$ и $I_{S.B_1}$ соответственно:

$$\forall j \in \{2, \dots, u\} \left(I_{R.B_j} = \bigcup_{i=0}^{k-1} I_{R.B_j}^i \right); \quad (50)$$

$$\forall j \in \{2, \dots, u\} \left(I_{S.B_j} = \bigcup_{i=0}^{k-1} I_{S.B_j}^i \right). \quad (51)$$

Положим для всех $i = 0, \dots, k-1$ и $j = 1, \dots, u$

$$\tilde{P}_j^i = \pi_{I_{R.B_j}^i.A \rightarrow A_R, I_{S.B_j}^i.A \rightarrow A_S} \left(I_{R.B_j}^i \bowtie_{(I_{R.B_j}^i.B_j = I_{S.B_j}^i.B_j)} I_{S.B_j}^i \right). \quad (52)$$

Определим

$$\tilde{P}^i = \bigcap_{j=1}^u \tilde{P}_j^i. \quad (53)$$

Положим для всех $i = 0, \dots, k-1$

$$P_R^i = \pi_A \left(I_{R.B_1}^i \right) \quad (54)$$

и

$$P_S^i = \pi_A \left(I_{S.B_1}^i \bowtie_{(I_{S.B_1}^i.A \neq \tilde{P}^i.A_S)} \tilde{P}^i \right). \quad (55)$$

Определим

$$P_R = \bigcup_{i=0}^{k-1} P_R^i, \quad (56)$$

$$P_S = \bigcup_{i=0}^{k-1} P_S^i. \quad (57)$$

Построим отношение $Q(A, B_1, \dots, B_u)$ следующим образом:

$$Q = \{ \&_R(p.A) \mid p \in P_R \} \cup \{ \&_S(p.A) \mid p \in P_S \}. \quad (58)$$

Теорема 4. $\pi_{B_1, \dots, B_u}(Q) = \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$.

Доказательство. Сначала докажем, что

$$\pi_{B_1, \dots, B_u}(Q) \subset \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S). \quad (59)$$

Пусть

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(Q). \quad (60)$$

Из (58) следует, что как минимум одно из следующих условий истинно:

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u} \left(\{ \&_R(p.A) \mid p \in P_R \} \right); \quad (61)$$

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u} \left(\{ \&_S(p.A) \mid p \in P_S \} \right). \quad (62)$$

Предположим сначала, что условие (61) истинно. Тогда существует $p \in P_R$ такой, что

$$\&_R(p.A) = r \in R$$

и

$$r = (a, b_1, \dots, b_u),$$

где $a = p.A$. Отсюда получаем $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R)$. Следовательно, $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$, и (59) имеет место. Для случая, когда истинным явля-

ется условие (62), это утверждение доказывается аналогично (в рассуждениях выше надо просто заменить R на S и r на s).

Теперь докажем, что

$$\pi_{B_1, \dots, B_u}(Q) \supset \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S). \quad (63)$$

По условию, объединение в правой части не должно содержать дубликатов (в предположении, что $\pi_{B_1, \dots, B_u}(R)$ и $\pi_{B_1, \dots, B_u}(S)$ не содержат дубликатов). Такое объединение без дубликатов может быть вычислено следующим образом:

$$\tilde{S} = \pi_{S,*} \left(R \underset{R.B_1=S.B_1 \wedge \dots \wedge R.B_u=S.B_u}{\bowtie} S \right); \quad (64)$$

$$\bar{S} = \pi_{S,*} \left(\tilde{S} \underset{\bar{S}.A \neq S.A}{\bowtie} S \right); \quad (65)$$

$$\pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S) = \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(\bar{S}). \quad (66)$$

Поэтому условие (63) эквивалентно условию

$$\pi_{B_1, \dots, B_u}(Q) \supset \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(\bar{S}). \quad (67)$$

Покажем, что последнее истинно. Пусть $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(\bar{S})$. Тогда $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R)$, либо $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(\bar{S})$. Предположим сначала, что $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R)$. Это означает, что существует $r \in R$ такой, что

$$r.B_1 = b_1, \dots, r.B_u = b_u. \quad (68)$$

Положим

$$r.A = a. \quad (69)$$

По определению колоночного индекса (свойства (1) и (3)) отсюда следует что $(a, b_1) \in I_{R.B_1}$. В силу свойств доменно-интервальной фрагментации существует $i \in \mathbb{N}$ такое, что $(a, b_1) \in I_{R.B_1}^i$. Тогда с учетом (54) и (56) получаем

$$p = (a) \in P_R. \quad (70)$$

Положим

$$r' = \&_R(p.A). \quad (71)$$

По построению

$$r'.A = a. \quad (72)$$

Принимая во внимание, что A — суррогатный ключ, из (72) и (69) следует

$$r = r'. \quad (73)$$

Из (58), (70), (71), (73) следует, что $r \in Q$. Вместе с (68) это дает $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(Q)$, то есть (63) имеет место для рассмотренного случая.

Предположим теперь, что

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(\bar{S}). \quad (74)$$

Это означает, что существует $\bar{s} \in \bar{S}$ такой, что

$$\bar{s}.B_1 = b_1, \dots, \bar{s}.B_u = b_u. \quad (75)$$

Положим

$$\bar{s}.A = a. \quad (76)$$

Из (65) следует, что $\bar{s} \in S$. По определению колоночного индекса из (75) следует что

$$\begin{aligned} (a, b_1) &\in I_{S.B_1}; \\ &\dots\dots\dots \\ (a, b_u) &\in I_{S.B_u}. \end{aligned} \quad (77)$$

В силу свойств доменно-интервальной фрагментации существует i такой, что $(a, b_1) \in I_{S, B}^i$. По определению транзитивной относительно I_{S, B_1} фрагментации отсюда, с учетом (77), имеем

$$\begin{aligned} (a, b_2) &\in I_{S, B_2}^i; \\ &\dots\dots\dots \\ (a, b_u) &\in I_{S, B_u}^i. \end{aligned}$$

Покажем, что

$$a \notin \tilde{P}^i \tag{78}$$

Предположим противное, то есть $a \in \tilde{P}^i$. Тогда существует $(a', b'_1, \dots, b'_u) \in R$ такой, что $(a', a) \in \tilde{P}^i$. По определению колоночного индекса $(a', b'_1) \in I_{R, B_1}$. В силу свойств доменно-интервальной фрагментации $(a', b'_1) \in I_{R, B_1}^i$. По определению транзитивной относительно I_{R, B_1} фрагментации отсюда следует

$$\begin{aligned} (a', b'_2) &\in I_{R, B_2}^i; \\ &\dots\dots\dots \\ (a', b'_u) &\in I_{R, B_u}^i. \end{aligned}$$

Но тогда с учетом (53) и (52) получаем

$$\begin{aligned} b'_1 &= b_1; \\ &\dots \\ b'_u &= b_u. \end{aligned}$$

Отсюда в силу (64) следует $(a, b_1, \dots, b_u) \in \tilde{S}$. Принимая во внимание (65), получаем $(a, b_1, \dots, b_u) \notin \bar{S}$. Так как мы предположили, что $\pi_{B_1, \dots, B_u}(S)$ не содержит дубликатов, отсюда следует $(b_1, \dots, b_u) \notin \pi_{B_1, \dots, B_u}(\bar{S})$. Получили противоречие с (74). Таким образом, (78) истинно. Тогда из (55) следует, что $(a) \in P_S^i$. С учетом (57) получаем $(a) \in P_S$. Теперь из (58), (76), (75) и (65) следует, что $(a, b_1, \dots, b_u) \in Q$, то есть $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(Q)$. Теорема доказана.

2. Декомпозиция реляционных операций с использованием фрагментированных колоночных хеш-индексов

В этом разделе вводится понятие колоночного хеш-индекса. Колоночный хеш-индекс позволяет использовать один колоночный индекс для индексирования нескольких атрибутов одного отношения. Для распределенных колоночных хеш-индексов рассматривается декомпозиция операций пересечения, объединения и естественного соединения.

2.1. Колоночный хеш-индекс

Определение 1. Пусть задано отношение $R(A, B_1, \dots, B_u, \dots)$. Пусть задана инъективная хеш-функция $h: \mathfrak{D}_{B_1} \times \dots \times \mathfrak{D}_{B_u} \rightarrow \mathbb{Z}_{\geq 0}$. Колоночным хеш-индексом $I_h(A, H)$ атрибу-

тов B_1, \dots, B_u отношения R будем называть упорядоченное отношение, удовлетворяющее тождеству:

$$I_h = \tau_H \left(\pi_{A, h(B_1, \dots, B_u)}(R) \right). \quad (79)$$

Колоночный хеш-индекс обладает следующим основным свойством:

$$\forall r', r'' \in R \left(r'.B_1 = r''.B_1 \wedge \dots \wedge r'.B_u = r''.B_u \Leftrightarrow h(r'.B_1, \dots, r'.B_u) = h(r''.B_1, \dots, r''.B_u) \right). \quad (80)$$

Заметим, что обратная импликация следует из инъективности хеш-функции h . Из (80) непосредственно вытекает следующее свойство колоночного хеш-индекса:

$$\forall r', r'' \in R \left(h(r'.B_1, \dots, r'.B_u) \neq h(r''.B_1, \dots, r''.B_u) \Leftrightarrow r'.B_1 \neq r''.B_1 \vee \dots \vee r'.B_u \neq r''.B_u \right).$$

Фрагментация колоночного хеш-индекса осуществляется на основе доменно-интервального принципа с помощью функции фрагментации $\varphi_{I_h} : I_h \rightarrow \{0, \dots, k-1\}$, определенной следующим образом:

$$\forall x \in I_h \left(\varphi_{I_h}(x) = \varphi_{\mathbb{Z}_{\geq 0}}(x.H) \right), \quad (81)$$

где $\varphi_{\mathbb{Z}_{\geq 0}} : \mathbb{Z}_{\geq 0} \rightarrow \{0, \dots, k-1\}$ — доменная функция фрагментации для домена $\mathcal{D}_H = \mathbb{Z}_{\geq 0}$.

2.2. Декомпозиция операции пересечения

В данном разделе рассматривается декомпозиция операции пересечения вида $\pi_{B_1, \dots, B_u}(R) \cap \pi_{B_1, \dots, B_u}(S)$ с использованием распределенных колоночных хеш-индексов. При этом предполагается, что $\pi_{B_1, \dots, B_u}(R)$ и $\pi_{B_1, \dots, B_u}(S)$ не содержат дубликатов.

Пусть заданы два отношения $R(A, B_1, \dots, B_u)$ и $S(A, B_1, \dots, B_u)$, имеющие одинаковый набор атрибутов. Пусть имеются два колоночных хеш-индекса $I_{R,h}$ и $I_{S,h}$ для атрибутов B_1, \dots, B_u отношений R и S , построенные с помощью одной и той же инъективной хеш-функции $h : \mathcal{D}_{B_1} \times \dots \times \mathcal{D}_{B_u} \rightarrow \mathbb{Z}_{\geq 0}$. Пусть для этих индексов задана доменно-интервальная фрагментация степени k :

$$I_{R,h} = \bigcup_{i=0}^{k-1} I_{R,h}^i; \quad (82)$$

$$I_{S,h} = \bigcup_{i=0}^{k-1} I_{S,h}^i. \quad (83)$$

Положим

$$P^i = \pi_{I_{R,h}^i.A \rightarrow A_R, I_{S,h}^i.A \rightarrow A_S} \left(I_{R,h}^i \bowtie_{(I_{R,h}^i.H = I_{S,h}^i.H)} I_{S,h}^i \right) \quad (84)$$

для всех $i = 0, \dots, k-1$. Определим

$$P = \bigcup_{i=0}^{k-1} P^i. \quad (85)$$

Положим

$$Q = \{ \&_R(p.A_R) \mid p \in P \}. \quad (86)$$

Теорема 5. $\pi_{B_1, \dots, B_u}(Q) = \pi_{B_1, \dots, B_u}(R) \cap \pi_{B_1, \dots, B_u}(S)$.

Доказательство. Сначала докажем, что

$$\pi_{B_1, \dots, B_u}(Q) \subset \pi_{B_1, \dots, B_u}(R) \cap \pi_{B_1, \dots, B_u}(S). \quad (87)$$

Пусть

$$(a, b_1, \dots, b_u) \in Q. \quad (88)$$

Из (86) следует, что

$$(a, b_1, \dots, b_u) = r \in R \quad (89)$$

и

$$a = r.A \in \pi_{A_R}(P). \quad (90)$$

Отсюда следует, что существует $p \in P$ такой, что $p.A_R = a \wedge p.A_S = a'$. С учетом (85) получаем, что существует i , для которого $p \in P^i$. С учетом (84) отсюда получаем, что при некотором $\chi \in \mathbb{Z}_{\geq 0}$:

$$\begin{aligned} (a, \chi) &\in I_{R,h}^i \subset I_{R,h}; \\ (a', \chi) &\in I_{S,h}^i \subset I_{S,h}. \end{aligned}$$

Поскольку хеш-функция h является инъективной, то для нее существует обратная функция. Положим $(b'_1, \dots, b'_u) = h^{-1}(\chi)$. Тогда по определению колоночного хеш-индекса имеем

$$\begin{aligned} (a, b'_1, \dots, b'_u) &\in R; \\ (a', b'_1, \dots, b'_u) &\in S. \end{aligned}$$

Поскольку A является суррогатным ключом в R , с учетом (89) получаем

$$\begin{aligned} (a, b_1, \dots, b_u) &\in R; \\ (a', b_1, \dots, b_u) &\in S. \end{aligned}$$

Следовательно $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cap \pi_{B_1, \dots, B_u}(S)$, и (87) имеет место.

Теперь докажем, что

$$\pi_{B_1, \dots, B_u}(R) \cap \pi_{B_1, \dots, B_u}(S) \subset \pi_{B_1, \dots, B_u}(Q). \quad (91)$$

Пусть

$$(a, b_1, \dots, b_u) = r \in R \quad (92)$$

и

$$(a', b_1, \dots, b_u) = s \in S. \quad (93)$$

Положим $\chi = h(b_1, \dots, b_u)$. Тогда по определению колоночного хеш-индекса

$$\begin{aligned} (a, \chi) &\in I_{R,h}; \\ (a', \chi) &\in I_{S,h}. \end{aligned}$$

В силу свойств доменно-интервальной фрагментации существует i такой, что

$$\begin{aligned} (a, \chi) &\in I_{R,h}^i; \\ (a', \chi) &\in I_{S,h}^i. \end{aligned}$$

На основе (84) и (85) отсюда получаем $(a, a') \in P^i \subset P$. Поскольку A — суррогатный ключ в R , с учетом (86) имеем $(a, b_1, \dots, b_u) \in Q$ и следовательно $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(Q)$.

Таким образом (91) имеет место. *Теорема доказана.*

2.3. Декомпозиция операции объединения

Пусть заданы два отношения $R(A, B_1, \dots, B_u)$ и $S(A, B_1, \dots, B_u)$. Выполним декомпозицию операции объединения вида $\pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$. Пусть имеются два колоночных хеш-индекса $I_{R,h}$ и $I_{S,h}$ для атрибутов B_1, \dots, B_u отношений R и S , построенные с помощью одной и той же инъективной хеш-функции $h: \mathfrak{D}_{B_1} \times \dots \times \mathfrak{D}_{B_u} \rightarrow \mathbb{Z}_{\geq 0}$. Пусть для этих индексов задана доменно-интервальная фрагментация степени k :

$$I_{R,h} = \bigcup_{i=0}^{k-1} I_{R,h}^i ; \quad (94)$$

$$I_{S,h} = \bigcup_{i=0}^{k-1} I_{S,h}^i . \quad (95)$$

Положим

$$P_i = \pi_A(I_{S,h}^i) \setminus \pi_{I_{S,h}^i, A} \left(I_{R,h}^i \underset{(I_{R,h}^i, H=I_{S,h}^i, H)}{\bowtie} I_{S,h}^i \right) \quad (96)$$

для всех $i = 0, \dots, k-1$.

Определим

$$P = \bigcup_{i=0}^{k-1} P_i . \quad (97)$$

Положим

$$Q = \{ \&_S(p.A) \mid p \in P \} . \quad (98)$$

Теорема 6. $\pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(Q) = \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$.

Доказательство. Сначала докажем, что

$$\pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(Q) \subset \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S) . \quad (99)$$

Пусть

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(Q) . \quad (100)$$

Тогда как минимум одно из следующих условий истинно:

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) ; \quad (101)$$

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(Q) . \quad (102)$$

Если имеет место (101), то очевидно $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$ и (99) имеет место.

Пусть истинно условие (102). Тогда из (98) следует, что существует $p \in P$ такой, что

$$\&_S(p.A) = s \in S$$

и

$$s = (a, b_1, \dots, b_u) ,$$

где $a = p.A$. Отсюда получаем $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(S)$, то есть

$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S)$ и (99) также имеет место.

Теперь докажем, что

$$\pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(Q) \supset \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S) . \quad (103)$$

Пусть

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(S) . \quad (104)$$

Тогда как минимум одно из следующих условий истинно:

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) ; \quad (105)$$

$$(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(S) . \quad (106)$$

Если имеет место (105), то очевидно $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R) \cup \pi_{B_1, \dots, B_u}(Q)$ и (103) имеет место.

Пусть истинно условие (106) и

$$(b_1, \dots, b_u) \notin \pi_{B_1, \dots, B_u}(R) . \quad (107)$$

Тогда существует

$$s = (a, b_1, \dots, b_u) \in S. \quad (108)$$

Положим

$$\chi = h(b_1, \dots, b_u). \quad (109)$$

Тогда по определению колоночного хеш-индекса

$$(a, \chi) \in I_{S,h}.$$

В силу свойств доменно-интервальной фрагментации существует $i \in \{0, \dots, k-1\}$ такой, что

$$(a, \chi) \in I_{S,h}^i.$$

Покажем, что

$$(a) \notin \pi_{I_{S,h}^i \cdot A} \left(I_{R,h}^i \underset{(I_{R,h}^i \cdot H = I_{S,h}^i \cdot H)}{\bowtie} I_{S,h}^i \right) \quad (110)$$

Предположим противное, то есть

$$(a) \in \pi_{I_{S,h}^i \cdot A} \left(I_{R,h}^i \underset{(I_{R,h}^i \cdot H = I_{S,h}^i \cdot H)}{\bowtie} I_{S,h}^i \right) \quad (111)$$

Тогда существует $(a') \in \pi_{I_{R,h}^i \cdot A} (I_{R,h}^i)$ такой, что $(a', \chi) \in I_{R,h}^i$. С учетом (109) в силу инъективности хеш-функции h получаем $(a', b_1, \dots, b_u) \in R$, то есть $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(R)$. Получили противоречие с (107). Значит (110) имеет место. Из (110) и (96) следует, что $(a) \in P_i$, и из (97) следует, что $(a) \in P$. Тогда из (98) следует, что существует $s' = (a, b'_1, \dots, b'_u) \in Q \subseteq S$. Так как A — суррогатный ключ, сопоставляя это с (108), получаем $(a, b_1, \dots, b_u) = (a, b'_1, \dots, b'_u) \in Q$, откуда $(b_1, \dots, b_u) \in \pi_{B_1, \dots, B_u}(Q)$. Таким образом (103) имеет место. *Теорема доказана.*

2.4. Декомпозиция операции естественного соединения

Пусть заданы два отношения

$$R(A, B_1, \dots, B_u, C_1, \dots, C_v)$$

и

$$S(A, B_1, \dots, B_u, D_1, \dots, D_w).$$

Пусть имеются два колоночных хеш-индекса $I_{R,h}$ и $I_{S,h}$ для атрибутов B_1, \dots, B_u отношений R и S , построенные с помощью одной и той же инъективной хеш-функции $h: \mathfrak{D}_{B_1} \times \dots \times \mathfrak{D}_{B_u} \rightarrow \mathbb{Z}_{\geq 0}$. Пусть для этих индексов задана доменно-интервальная фрагментация степени k :

$$I_{R,h} = \bigcup_{i=0}^{k-1} I_{R,h}^i; \quad (112)$$

$$I_{S,h} = \bigcup_{i=0}^{k-1} I_{S,h}^i. \quad (113)$$

Положим

$$P^i = \pi_{I_{R,h}^i \cdot A \rightarrow A_R, I_{S,h}^i \cdot A \rightarrow A_S} \left(I_{R,h}^i \underset{(I_{R,h}^i \cdot H = I_{S,h}^i \cdot H)}{\bowtie} I_{S,h}^i \right) \quad (114)$$

для всех $i = 0, \dots, k-1$. Определим

$$P = \bigcup_{i=0}^{k-1} P^i. \quad (115)$$

Построим отношение $Q(B_1, \dots, B_u, C_1, \dots, C_v, D_1, \dots, D_w)$ следующим образом:

$$Q = \left\{ (\&_R(p.A_R).B_1, \dots, \&_R(p.A_R).B_u, \right. \\ \&_B(p.A_B).C_1, \dots, \&_B(p.A_B).C_v, \\ \left. \&_S(p.A_S).D_1, \dots, \&_S(p.A_S).D_w) \mid p \in P \right\}. \quad (116)$$

Теорема 7. $Q = \pi_{*\setminus A}(R) \bowtie \pi_{*\setminus A}(S)$.

Доказательство. Сначала докажем, что

$$Q \subset \pi_{*\setminus A}(R) \bowtie \pi_{*\setminus A}(S). \quad (117)$$

Пусть

$$(b_1, \dots, b_u, c_1, \dots, c_v, d_1, \dots, d_w) \in Q. \quad (118)$$

Из (116) следует, что существуют кортежи r и s такие, что

$$(a, b_1, \dots, b_u, c_1, \dots, c_v) = r \in R, \quad (119)$$

$$(a', b'_1, \dots, b'_u, d_1, \dots, d_w) = s \in S \quad (120)$$

и

$$(r.A, s.A) = p \in P. \quad (121)$$

С учетом (115) получаем, что существует i , для которого $p \in P^i$. С учетом (114) отсюда получаем, что при некотором $\chi \in \mathbb{Z}_{\geq 0}$:

$$(a, \chi) \in I_{R,h}^i \subset I_{R,h};$$

$$(a', \chi) \in I_{S,h}^i \subset I_{S,h}.$$

Поскольку хеш-функция h является инъективной, то для нее существует обратная функция. Положим $(b_1'', \dots, b_u'') = h^{-1}(\chi)$. Тогда по определению колоночного хеш-индекса имеем

$$(a, b_1'', \dots, b_u'') \in \pi_{A, B_1, \dots, B_u}(R); \quad (122)$$

$$(a', b_1'', \dots, b_u'') \in \pi_{A, B_1, \dots, B_u}(S). \quad (123)$$

Поскольку A является суррогатным ключом в R и S , из (119), (120), (122) и (123) следует $b_1 = b_1'' = b'_1, \dots, b_u = b_u'' = b'_u$, то есть

$$(a, b_1, \dots, b_u, c_1, \dots, c_v) \in R,$$

$$(a', b_1, \dots, b_u, d_1, \dots, d_w) \in S.$$

Следовательно $(b_1, \dots, b_u, c_1, \dots, c_v, d_1, \dots, d_w) \in \pi_{*\setminus A}(R) \bowtie \pi_{*\setminus A}(S)$, и (117) имеет место.

Теперь докажем, что

$$Q \supset \pi_{*\setminus A}(R) \bowtie \pi_{*\setminus A}(S). \quad (124)$$

Пусть

$$(a, b_1, \dots, b_u, c_1, \dots, c_v) \in R \quad (125)$$

и

$$(a', b_1, \dots, b_u, d_1, \dots, d_w) \in S. \quad (126)$$

Положим $\chi = h(b_1, \dots, b_u)$. Тогда по определению колоночного хеш-индекса

$$(a, \chi) \in I_{R,h};$$

$$(a', \chi) \in I_{S,h}.$$

В силу свойств доменно-интервальной фрагментации существует $i \in \{0, \dots, k-1\}$ такой, что

$$(a, \chi) \in I_{R,h}^i;$$

$$(a', \chi) \in I_{S,h}^i.$$

На основе (114) и (115) отсюда получаем $(a, a') \in P^i \subset P$. Поскольку A — суррогатный ключ в R и S , с учетом (116), (125) и (126) имеем $(b_1, \dots, b_u, c_1, \dots, c_v, d_1, \dots, d_w) \in Q$. Таким образом (124) имеет место. *Теорема доказана.*

Заключение

В статье были рассмотрены вопросы декомпозиции операций проекции, выбора, удаления дубликатов и объединения с использованием фрагментированных колоночных индексов. Предложен новый вид колоночных индексов, названных колоночными хеш-индексами. Колоночный хеш-индекс способен индексировать сразу несколько атрибутов отношения. Для распределенных колоночных хеш-индексов рассмотрена декомпозиция операций пересечения, объединения и естественного соединения. Предложенные модели и методы являются теоретической основой для разработки колоночного сопроцессора, позволяющего во взаимодействии с реляционной СУБД эффективно выполнять запросы класса OLAP.

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).

Литература

1. Чернышев, Г.А. Организация физического уровня колоночных СУБД / Г.А Чернышев // Труды СПИИРАН. — 2013. — № 7. Вып. 30. — С. 204–222.
2. Abadi, D.J. The Design and Implementation of Modern Column-Oriented Database Systems / D.J. Abadi, P.A. Boncz, S. Harizopoulos, S. Idreos, S. Madden // Foundations and Trends in Databases. — 2013. — Vol. 5, No. 3. — P. 197–280. DOI: 10.1561/1900000024.
3. Abadi, D.J. Column-Stores vs. Row-Stores: How Different Are They Really? / D.J. Abadi, S.R. Madden, N. Hachem // Proceedings of the 2008 ACM SIGMOD international conference on Management of data, June 9–12, 2008, Vancouver, BC, Canada. — ACM, 2008. — P. 967–980. DOI: 10.1145/1376616.1376712.
4. Иванова, Е.В. Декомпозиция операций пересечения и соединения на основе доменно-интервальной фрагментации колоночных индексов / Е.В. Иванова, Л.Б. Соколинский // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. — 2015. — Т. 4, № 1. — С. 44–56. DOI: 10.14529/cmse150104.
5. Ivanova, E. Decomposition of Natural Join Based on Domain-Interval Fragmented Column Indices / E. Ivanova, L. Sokolinsky // Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO, May 25–29, 2015, Opatija, Croatia. — IEEE, 2015. — P. 223–226. DOI: 10.1109/mipro.2015.7160266.
6. Иванова, Е.В. Использование сопроцессоров Intel Xeon Phi для выполнения естественного соединения над сжатыми данными / Е.В. Иванова, Л.Б. Соколинский // Суперкомпьютерные дни в России: Труды международной конференции (28–29 сентября 2015 г., Москва). — М.: Изд-во МГУ, 2015. — С. 190–198.

7. 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.
8. Иванова, Е.В. Декомпозиция операции группировки на базе распределенных колоночных индексов / Е.В. Иванова, Л.Б. Соколинский // Наука ЮУрГУ. — Челябинск: Издательский центр ЮУрГУ, 2015. — С. 15–23.
9. Иванова, Е.В. Исследование эффективности использования фрагментированных колоночных индексов при выполнении операции естественного соединения с использованием многоядерных ускорителей / Е.В. Иванова // Параллельные вычислительные технологии (ПАВТ'2015): труды международной научной конференции (30 марта — 3 апреля 2015 г., Екатеринбург). — Челябинск: Издательский центр ЮУрГУ, 2015. — С. 393–398.
10. 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 (22–27 сентября 2014 г., Новороссийск). — М.: Изд-во МГУ, 2014. — С. 102–104.

Иванова Елена Владимировна, программист отдела поддержки и обучения пользователей Лаборатории суперкомпьютерного моделирования, Южно-Уральский государственный университет (Челябинск, Российская Федерация), Elena.Ivanova@susu.ru.

Соколинский Леонид Борисович, д. ф.-м. н., профессор, проректор по информатизации, Южно-Уральский государственный университет (Челябинск, Российская Федерация), Leonid.Sokolinsky@susu.ru.

Поступила в редакцию 5 сентября 2015 г.

PARALLEL DECOMPOSITION OF RELATIONAL OPERATIONS BASED ON FRAGMENTED COLUMN INDICES

E.V. Ivanova, South Ural State University (Chelyabinsk, Russian Federation)
Elena.Ivanova@susu.ru,

L.B. Sokolinsky, South Ural State University (Chelyabinsk, Russian Federation)
Leonid.Sokolinsky@susu.ru

This paper is a continuation and development of the previous our work, where the decompositions of intersection and join operations for columnar indices on the basis of domain-interval fragmentation were proposed. This decomposition allows to organize the parallel execution of relational operations on the distributed columnar indices without massive data exchange between the processor nodes. In this paper the decompositions of the projection, selection, duplicate elimination and union operations are considered. Also, a new kind of columnar indices called columnar hash-indices is introduced. The columnar hash-index can index multiple attributes of a relation. For distributed columnar hash-indices, the decompositions of intersection, union, and natural join operations are considered.

Keywords: distributed column indices, domain-interval fragmentation, column hash indices, decomposition of relational operations.

References

1. Chernyshev G.A Organizaciya fizicheskogo urovnya kolonochnyh SUBD [Physical Layer Organization of Columnar DBMS] // Trudy SPIIRAN [Proceedings of the SPIIRAS]. 2013. Vol. 7. No. 30. P. 204–222.
2. Abadi D.J., Boncz P.A., Harizopoulos S., Idreos S., Madden S. The Design and Implementation of Modern Column-Oriented Database Systems // Foundations and Trends in Databases. 2013. Vol. 5, No. 3. P. 197–280. DOI: 10.1561/1900000024.
3. Abadi D.J., Madden S.R, Hachem N. Column-Stores vs. Row-Stores: How Different Are They Really? // Proceedings of the 2008 ACM SIGMOD international conference on Management of data, June 9–12, 2008, Vancouver, BC, Canada. ACM, 2008. P. 967–980. DOI: 10.1145/1376616.1376712.
4. Ivanova E.V., Sokolinsky L.B. Dekompoziciya operacij peresecheniya i soedineniya na osnove domenno-interval'noj fragmentacii kolonochnyh indeksov [Decomposition of Intersection and Join Operations Based on Domain-Interval Fragmented Column Indices] // Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya «Vychislitel'naya matematika i informatika» [Bulletin of South Ural State University. Series: Computational Mathematics and Software Engineering]. 2015. Vol. 4, No. 1. P. 44–56. DOI: 10.14529/cmse150104.
5. Ivanova E., Sokolinsky L. Decomposition of Natural Join Based on Domain-Interval Fragmented Column Indices // Proceedings of the 38th International Convention on In-

- formation and Communication Technology, Electronics and Microelectronics, MIPRO, May 25-29, 2015, Opatija, Croatia. IEEE, 2015. P. 223–226. DOI: 10.1109/mipro.2015.7160266.
6. Ivanova E.V., Sokolinsky L.B. Ispol'zovanie soprocesorov Intel Xeon Phi dlya vypolneniya estestvennogo soedineniya nad szhatymi dannymi [Using Intel Xeon Phi coprocessor for execution of natural join on compressed data] // Superkomp'yuternye dni v Rossii: Trudy mezhdunarodnoj konferencii [Russian Supercomputer Days: Proceedings of the International Conference], September 28-29, 2015, Moscow, Russia. Moscow: MSU publishing center, 2015. P. 190–198.
 7. Ivanova E.V., Sokolinsky L.B. Ispol'zovanie raspredelennykh kolonochnykh indeksov dlya vypolneniya zaprosov k sverhbol'shim bazam dannyx [Using Distributed Column Indices for Query Execution for Very Large Databases] // Parallel'nye vychislitel'nye tekhnologii (PaVT'2014). Trudy mezhdunarodnoy nauchnoy konferentsii [Proceedings of the International Conference Parallel Computational Technologies (PCT'2014)]. Chelyabinsk: SUSU publishing center, 2014. P. 270–275.
 8. Ivanova E.V., Sokolinsky L.B. Dekompoziciya operacii gruppirovki na baze raspredelennykh kolonochnykh indeksov [Decomposition of Grouping Operation Based on Fragmented Column Indices] // Nauka YUUrGU [Science of SUSU]. Chelyabinsk: SUSU publishing center, 2015. P. 15–23.
 9. Ivanova E.V. Issledovanie ehffektivnosti ispol'zovaniya fragmentirovannykh kolonochnykh indeksov pri vypolnenii operacii estestvennogo soedineniya s ispol'zovaniem mnogo-yadernykh uskoritelej [Research of efficiency fragmented columnar indices for the natural join operation using multi-core accelerators.] // Parallel'nye vychislitel'nye tekhnologii (PaVT'2015). Trudy mezhdunarodnoy nauchnoy konferentsii [Proceedings of the International Conference Parallel Computational Technologies (PCT'2015)], March 30 — April 3, 2015, Ekaterinburg, Russia. Chelyabinsk: SUSU publishing center, 2015. P. 393–398.
 10. Ivanova E.V. Ispol'zovanie raspredelennykh kolonochnykh hesh-indeksov dlya obrabotki zaprosov k sverhbol'shim bazam dannyx [Using Distributed Column Hash Indices for Query Execution for Very Large Databases] // Nauchnyj servis v seti Internet: mnogoobrazie superkomp'yuternykh mirov: Trudy Mezhdunarodnoj superkomp'yuternoj konferencii [Scientific service in Internet: The variety of supercomputing worlds: Proceedings of the International Supercomputer Conference], September 22–27, 2014, Novorossiysk, Russia. Moscow: MSU publishing center, 2014. P. 102–104.

Received September 5, 2015.