66-я научная конференция профессорско-преподавательского состава ЮУрГУ 2014

Методы обработки запросов с использованием распределенных колоночных индексов

Е.В. Иванова

Национальный исследовательский

Южно-Уральский государственный университет

Цель работы

- В настоящее время научная и практическая деятельность человека выдвигает все новые масштабные задачи, требующие обработки сверхбольших баз данных.
- Фактически единственным эффективным решением проблемы хранения и обработки сверхбольших баз данных является использование параллельных систем баз данных на многопроцессорных вычислительных системах.
- Мы предлагаем решение на основе индексных структур специального вида, которые называются ***распределенными колоночными индексами***.

Колоночный индекс

- Пусть $R(A^*, B, \dots)$– отношение R с первичным ключом A и атрибутом B. A представляет собой *служебный первичный ключ* и состоит из целочисленных неотрицательных элементов.
- $\mathfrak{D}_B$ – *домен атрибута* B. На множестве $\mathfrak{D}_B$задано отношение линейного порядка. $T(R) = n$ – количество элементов в R.
- ***Колоночным индексом*** $I_{R.B}$ атрибута B отношения R называется упорядоченное отношение, удовлетворяющее следующим требованиям:

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

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

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

R

A	B	....
0	36	....
1	14	....
2	36	....
3	10	....
4	74	....
5	27	....
6	58	....

$I_{R.B}$

A	B
3	10
1	14
5	27
2	36
0	36
6	58
4	74

Доменно-интервальная фрагментация

- Разобьем множество значений домена $\mathfrak{D}_B$ на k непересекающихся интервалов:

$$V_0 = [v_0; v_1]; V_1 = (v_1; v_2]; \dots; V_{k-1} = (v_{k-1}; v_k];$$

$$v_0 < v_1 < \dots < v_k;$$

$$\mathfrak{D}_B = \bigcup_{i=0}^{k-1} V_i$$

- ***Доменная функция фрагментации:***

$$\varphi_{\mathfrak{D}_B} : \mathfrak{D}_B \to \{0, \dots, k-1\}$$

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

Доменно-интервальная фрагментация

- ***Функция фрагментации индекса***:

$$\varphi_{I_{R.B}} : I_{R.B} \to \{0, \dots, k-1\}$$
$$\forall x \in I_{R.B} \left(\varphi_{I_{R.B}}(x) = \varphi_{\mathfrak{D}_B}(x.B) \right)$$

- Определим ***i-тый фрагмент индекса*** следующим образом

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

- Будем называть фрагментацию, построенную таким образом, *доменно-интервальной*.
- Количество фрагментов k будем называть *степенью фрагментации*.

Пример доменно-интервальной фрагментации колоночного индекса

Степень фрагментации: 2

Устройство хранения

R

A^*	B	C
0	**10**	a
1	**10**	z
2	**5**	s
3	**1**	s

$I^0{}_{R.B_1}$

A^*	**B**
3	**1**
2	**5**
0	**10**
1	**10**

Кластерная вычислительная система

Узел 0

$I^0{}_{R.B_1}$

A^*	B
3	**1**
2	**5**

Узел 1

$I^1{}_{R.B_1}$

A^*	B
0	**10**
1	**10**

Операция естественного соединения

- Пусть $R(A^*, B_1, \dots, B_u, C_1, \dots, C_h)$ и $S(A^*, B_1, \dots, B_u, D_1, \dots, D_w)$.
- Определим Q как результат операции $R \bowtie S$.
- Пусть имеется два набора колоночных индексов по атрибутам $B_1, \dots, B_u$, для которых задана доменно-интервальная фрагментация степени k:
$$I_{R.B_j} = \bigcup_{i=0}^{k-1} I^i_{R.B_j}; \quad I_{S.B_j} = \bigcup_{i=0}^{k-1} I^i_{S.B_j}.$$
- Положим для всех $i = 0, \dots, k-1$ и $j = 1, \dots, u$

$$P_j^i = \pi_{I^i_{R.B_j}.A \to A_R,\, I^i_{S.B_j}.A \to A_S} \left(I^i_{R.B_j} \underset{I^i_{R.B_j}.B_j = I^i_{S.B_j}.B_j}{\bowtie} I^i_{S.B_j} \right)$$

$$P_j = \bigcup_{i=0}^{k-1} P_j^i \qquad P = \bigcap_{j=1}^{u} P_j$$

$$Q = \left\{ r \circ (s.D_1, \dots, s.D_w) \mid r \in R \wedge s \in S \wedge (r.A, s.A) \in P \right\}$$

Пример операции естественного соединения

Определим $R \bowtie S$

Степень фрагментации: 2

R

A	B_1	B_2	C
0	10	a	Ni
1	10	z	Au
2	5	s	Pb
3	1	s	Ag

S

A	B_1	B_2	D
0	5	s	Pb
1	11	y	Pb
2	3	d	Ni
3	10	a	Fr
4	2	a	Ag

Пусть имеются распределенные колоночные индексы для атрибутов соединения B_1 и B_2 в таблицах R и S

Узел 0

${I^0}_{R.B_1}$

A	B_1
3	1
2	5

${I^0}_{R.B_2}$

A	B_2
0	a

${I^0}_{S.B_1}$

A	B_1
4	2
2	3
0	5

${I^0}_{S.B_2}$

A	B_2
3	a
4	a
2	d

Узел 1

${I^1}_{R.B_1}$

A	B_1
0	10
1	10

${I^1}_{R.B_2}$

A	B_2
2	s
3	s
1	z

${I^1}_{S.B_1}$

A	B_1
3	10
1	11

${I^1}_{S.B_2}$

A	B_2
0	s
1	y

Пример операции естественного соединения

Узел 0

$I^0{}_{R.B_1}$

A	B_1
3	1
2	5

$I^0{}_{R.B_2}$

A	B_2
0	a

$I^0{}_{S.B_1}$

A	B_1
4	2
2	3
0	5

$I^0{}_{S.B_2}$

A	B_2
3	a
4	a
2	d

$P^0{}_1$

$R.A$	$S.A$
2	0

$P^0{}_2$

$R.A$	$S.A$
0	3
0	4

Узел 1

$I^1{}_{R.B_1}$

A	B_1
0	10
1	10

$I^1{}_{R.B_2}$

A	B_2
2	s
3	s
1	z

$I^1{}_{S.B_1}$

A	B_1
3	10
1	11

$I^1{}_{S.B_2}$

A	B_2
0	s
1	y

$P^1{}_1$

$R.A$	$S.A$
0	3
1	3

$P^1{}_2$

$R.A$	$S.A$
2	0
3	0

Пример операции естественного соединения

Узел 0

$P^0{}_1$

$R.A$	$S.A$
2	0

$P^0{}_2$

$R.A$	$S.A$
0	3
0	4

$P^1{}_1$

$R.A$	$S.A$
0	3
1	3

$P^1{}_2$

$R.A$	$S.A$
2	0
3	0

Узел 1

$P^1{}_1$

$R.A$	$S.A$
0	3
1	3

$P^1{}_2$

$R.A$	$S.A$
2	0
3	0

Пример операции естественного соединения

Узел 0

$P^0{}_1$

$R.A$	$S.A$
2	0

$P^0{}_2$

$R.A$	$S.A$
0	3
0	4

$P^1{}_1$

$R.A$	$S.A$
0	3
1	3

$P^1{}_2$

$R.A$	$S.A$
2	0
3	0

P_1

$R.A$	$S.A$
2	0
0	3
1	3

P_2

$R.A$	$S.A$
0	3
0	4
2	0
3	0

$$P_j = \bigcup_{i=0}^{k-1} P_j^i$$

Пример операции естественного соединения

Узел 0

P_1

$R.A$	$S.A$
2	0
0	3
1	3

P_2

$R.A$	$S.A$
0	3
0	4
2	0
3	0

P

$R.A$	$S.A$
2	0
0	3

$$P = \bigcap_{j=1}^{u} P_j$$

Пример операции естественного соединения

R

A	B_1	B_2	C
0	10	a	Ni
1	10	z	Au
2	5	s	Pb
3	1	s	Ag

S

A	B_1	B_2	D
0	5	s	Pb
1	11	y	Pb
2	3	d	Ni
3	10	a	Fr
4	2	a	Ag

Узел 0

P

$R.A$	$S.A$
2	0
0	3

$R.A$	$R.B_1$	$R.B_2$	$R.C$	$S.D$
2	5	s	Pb	Pb
0	10	a	Ni	Fr

Хэш-индекс

- Хеш-индекс позволяет использовать один колоночный индекс для индексирования нескольких атрибутов одного отношения.
- Пусть задано отношение $R(A^*, B_1, \dots, B_u, C, \dots)$. Пусть задана хеш-функция $h : \mathfrak{D}_{B_1} \times \cdots \times \mathfrak{D}_{B_u} \to \mathbb{Z}_{\geq 0}$. *Хеш-индексом* $I_h(A^*, H)$ атрибутов $B_1, \dots, B_u$ отношения R называется упорядоченное отношение, удовлетворяющее тождеству:

$$I_h = \tau_H\left(\pi_{A,h(B_1,\dots,B_u)\to H}\left(R\right)\right)$$

- *Фрагментация хеш-индекса* осуществляется на основе доменно-интервального принципа.

Операция естественного соединения с использованием хэш-индекса

- Пусть $R(A^*, B_1, \dots, B_u, C_1, \dots, C_v)$ и $S(A^*, B_1, \dots, B_u, D_1, \dots, D_w)$.
- Определим Q как результат операции $R \bowtie S$.
- Пусть имеется два хэш-индекса для атрибутов $B_1, \dots, B_u$, построенные с помощью одной и той же хеш-функции h для которых задана доменно-интервальная фрагментация степени k:

$$I_{R,h} = \bigcup_{i=0}^{k-1} I_{R,h}^i \qquad I_{S,h} = \bigcup_{i=0}^{k-1} I_{S,h}^i$$

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

$$P^i = \pi_{I_{R,h}^i.A \to A_R,\ I_{S,h}^i.A \to A_S} \left(I_{R,h}^i \underset{(I_{R,h}^i.H = I_{S,h}^i.H)}{\bowtie} I_{S,h}^i \right) \qquad P = \bigcup_{i=0}^{k-1} P^i$$

$$Q = \{ \left(\&_R(p.A_R).B_1, \dots, \&_R(p.A_R).B_u, \&_S(p.A_S).D_1, \dots, \&_S(p.A_S).D_w \right) |$$
$$p \in P, \left(\&_R(p.A_R).B_1, \dots, \&_R(p.A_R).B_u \right) = \left(\&_S(p.A_S).B_1, \dots, \&_S(p.A_S).B_u \right) \}$$

Пример операции естественного соединения с хэш-индексом

Определим $R \bowtie S$

Степень фрагментации: 2

$$\boldsymbol{h}(\boldsymbol{b_1}, \boldsymbol{b_2}) = \boldsymbol{b_1} + \boldsymbol{ACSII}\_\text{код}(\boldsymbol{b_2})$$

R

A	B_1	B_2	C
0	10	a	Ni
1	10	z	Au
2	5	s	Pb
3	1	s	Ag

S

A	B_1	B_2	D
0	5	s	Pb
1	11	y	Pb
2	3	d	Ni
3	10	a	Fr
4	2	a	Ag

Узел 0

$I^0{}_{R.h}$

A	H
0	107
3	116

$I^0{}_{S.h}$

A	H
4	99
2	103
3	107

Узел 1

$I^1{}_{R.h}$

A	H
2	120
1	132

$I^1{}_{S.h}$

A	H
0	120
1	132

Спасибо за внимание!

Вопросы?