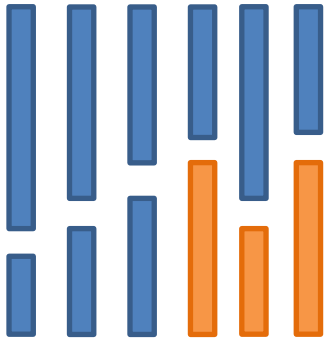


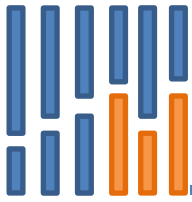


Параллельные вычислительные технологии 2014

Использование распределенных колоночных индексов для выполнения запросов к сверхбольшим базам данных

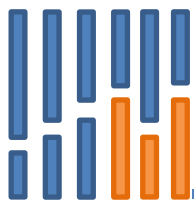
Е.В. Иванова, Л.Б. Соколинский

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

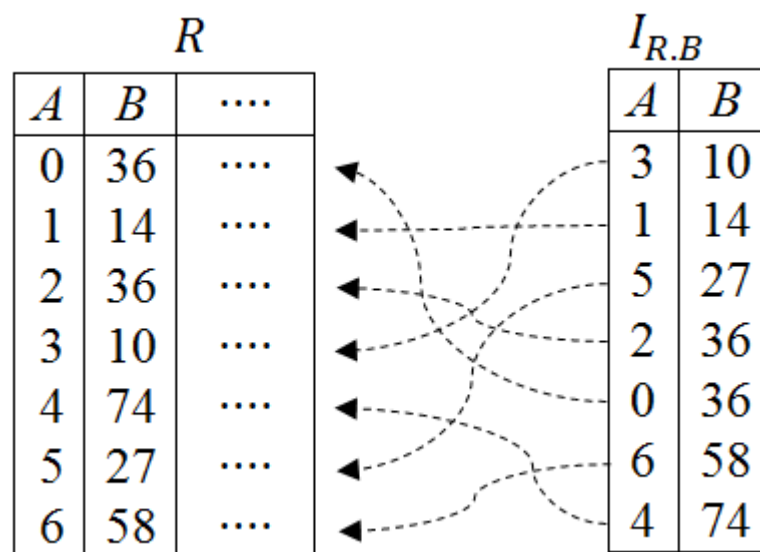
- $R(A^*, B, \dots)$ – отношение R с первичным ключом A и атрибутом B . A представляет собой *служебный первичный ключ* и состоит из целочисленных неотрицательных элементов. $\mathfrak{D}_B$ – *домен атрибута B* . На множестве $\mathfrak{D}_B$ задано отношение линейного порядка. $T(R) = n$ – количество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

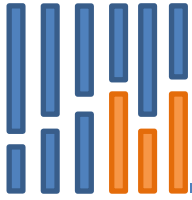
- **Колоночным индексом** $I_{R.B}$ атрибута B отношения R называется упорядоченное отношение, удовлетворяющее следующим требованиям:

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

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

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





Доменно-интервальная фрагментация

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

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

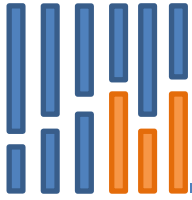
$$v_0 < v_1 < \dots < v_k;$$

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

- Доменная функция фрагментации:**

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

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



Доменно-интервальная фрагментация

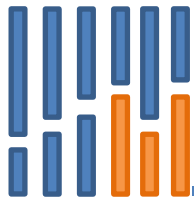
- **Функция фрагментации индекса:**

$$\begin{aligned}\varphi_{I_{R.B}}: I_{R.B} &\rightarrow \{0, \dots, k-1\} \\ \forall x \in I_{R.B} &(\varphi_{I_{R.B}}(x) = \varphi_{\mathcal{D}_B}(x.B))\end{aligned}$$

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

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

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



Пример доменно-интервальной фрагментации

Степень фрагментации: 2

R		
A	B_1	B_2
0	10	a
1	10	z
2	5	s
3	1	s

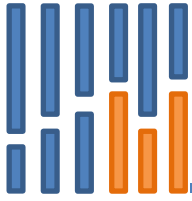
Узел 0

$I^0_{R.B_1}$		$I^0_{R.B_2}$	
A	B_1	A	B_2
3	1	0	a
2	5		

Узел 1

$I^1_{R.B_1}$		$I^1_{R.B_2}$	
A	B_1	A	B_2
0	10	2	s
1	10	3	s
		1	z

Узел 2



Операция пересечения

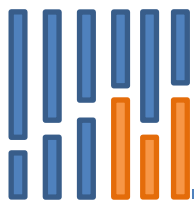
- Пусть заданы два отношения $R(A^*, B_1, \dots, B_u)$ и $S(A^*, B_1, \dots, B_u)$. Определим $Q = R \cap S$
- Пусть имеется два набора колоночных индексов по атрибутам $B_1, \dots, B_u$ и для всех этих индексов задана доменно-интервальная фрагментация степени k :

$$I_{R.B_j} = \bigcup_{i=0}^{k-1} I_{R.B_j}^i \quad I_{S.B_j} = \bigcup_{i=0}^{k-1} I_{S.B_j}^i .$$

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

$$P_j^i = \pi_{I_{R.B_j}^i.A \rightarrow A_R, I_{S.B_j}^i.A \rightarrow A_S} \left(I_{R.B_j}^i \bowtie_{R.B_j=S.B_j} I_{S.B_j}^i \right)$$

$$P_j = \bigcup_{i=0}^{k-1} P_j^i \quad P = \bigcap_{j=1}^u P_j \quad Q = \left\{ r \mid r \in R \wedge r.A \in \pi_{A_R}(P) \right\}$$



Пример операции пересечение

Определим $Q = R \cap S$

Степень фрагментации: 2

R

A	B_1	B_2
0	10	a
1	10	z
2	5	s
3	1	s

S

A	B_1	B_2
0	5	s
1	11	y
2	3	d
3	10	a
4	2	a

$I^0_{R.B_1}$ $I^0_{R.B_2}$ $I^0_{S.B_1}$ $I^0_{S.B_2}$

A	B_1
3	1
2	5

A	B_2
0	a

A	B_1
4	2
2	3
0	5

A	B_2
3	a
4	a
2	d

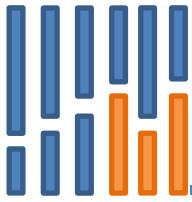
$I^1_{R.B_1}$ $I^1_{R.B_2}$ $I^1_{S.B_1}$ $I^1_{S.B_2}$

A	B_1
0	10
1	10

A	B_2
2	s
3	s
1	z

A	B_1
3	10
1	11

A	B_2
0	s
1	y



Пример операции пересечение

$I^0_{R.B_1}$	$I^0_{R.B_2}$	$I^0_{S.B_1}$	$I^0_{S.B_2}$																										
<table><tr><th>A</th><th>B_1</th></tr><tr><td>3</td><td>1</td></tr><tr><td>2</td><td>5</td></tr></table>	A	B_1	3	1	2	5	<table><tr><th>A</th><th>B_2</th></tr><tr><td>0</td><td>a</td></tr></table>	A	B_2	0	a	<table><tr><th>A</th><th>B_1</th></tr><tr><td>4</td><td>2</td></tr><tr><td>2</td><td>3</td></tr><tr><td>0</td><td>5</td></tr></table>	A	B_1	4	2	2	3	0	5	<table><tr><th>A</th><th>B_2</th></tr><tr><td>3</td><td>a</td></tr><tr><td>4</td><td>a</td></tr><tr><td>2</td><td>d</td></tr></table>	A	B_2	3	a	4	a	2	d
A	B_1																												
3	1																												
2	5																												
A	B_2																												
0	a																												
A	B_1																												
4	2																												
2	3																												
0	5																												
A	B_2																												
3	a																												
4	a																												
2	d																												

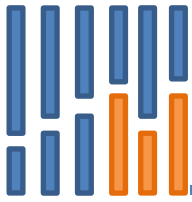


P^0_1	P^0_2										
<table border="1"> <tr><th>R.A</th><th>S.A</th></tr> <tr><td>2</td><td>0</td></tr> </table>	R.A	S.A	2	0	<table border="1"> <tr><th>R.A</th><th>S.A</th></tr> <tr><td>0</td><td>3</td></tr> <tr><td>0</td><td>4</td></tr> </table>	R.A	S.A	0	3	0	4
R.A	S.A										
2	0										
R.A	S.A										
0	3										
0	4										

$I^1_{R.B_1}$	$I^1_{R.B_2}$	$I^1_{S.B_1}$	$I^1_{S.B_2}$																										
<table> <tr><td>A</td><td>B_1</td></tr> <tr><td>0</td><td>10</td></tr> <tr><td>1</td><td>10</td></tr> </table>	A	B_1	0	10	1	10	<table> <tr><td>A</td><td>B_2</td></tr> <tr><td>2</td><td>s</td></tr> <tr><td>3</td><td>s</td></tr> <tr><td>1</td><td>z</td></tr> </table>	A	B_2	2	s	3	s	1	z	<table> <tr><td>A</td><td>B_1</td></tr> <tr><td>3</td><td>10</td></tr> <tr><td>1</td><td>11</td></tr> </table>	A	B_1	3	10	1	11	<table> <tr><td>A</td><td>B_2</td></tr> <tr><td>0</td><td>s</td></tr> <tr><td>1</td><td>y</td></tr> </table>	A	B_2	0	s	1	y
A	B_1																												
0	10																												
1	10																												
A	B_2																												
2	s																												
3	s																												
1	z																												
A	B_1																												
3	10																												
1	11																												
A	B_2																												
0	s																												
1	y																												



P^1_1	P^1_2												
<table border="1"> <tr><th>R.A</th><th>S.A</th></tr> <tr><td>0</td><td>3</td></tr> <tr><td>1</td><td>3</td></tr> </table>	R.A	S.A	0	3	1	3	<table border="1"> <tr><th>R.A</th><th>S.A</th></tr> <tr><td>2</td><td>0</td></tr> <tr><td>3</td><td>0</td></tr> </table>	R.A	S.A	2	0	3	0
R.A	S.A												
0	3												
1	3												
R.A	S.A												
2	0												
3	0												



Пример операции пересечение

 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

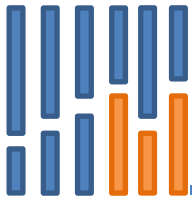
$R.A$	$S.A$
2	0
3	0

 P_1

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

 P_2

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



Пример операции пересечение

P_1

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

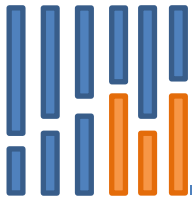
P_2

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



P

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



Пример операции пересечение

P

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

R

A	B_1	B_2
0	10	a
1	10	z
2	5	s
3	1	s

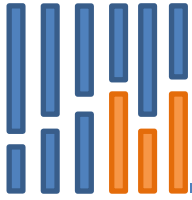
S

A	B_1	B_2
0	5	s
1	11	y
2	3	d
3	10	a
4	2	a



Q

B_1	B_2
10	a
10	z
5	s



Операция естественного соединения

- Пусть заданы два отношения R и S :

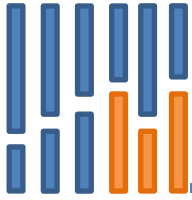
$$R(A^*, B_1, \dots, B_u, C_1, \dots, C_h) \quad S(A^*, B_1, \dots, B_u, D_1, \dots, D_w)$$

- Пусть имеется два набора колоночных индексов по атрибутам $B_1, \dots, B_u$ и для всех этих индексов задана доменно-интервальная фрагментация степени.
- Положим для всех $i = 0, \dots, k-1$ и $j = 1, \dots, u$

$$P_j^i = \pi_{I_{R.B_j}^i.A \rightarrow A_R, I_{S.B_j}^i.A \rightarrow A_S} \left(I_{R.B_j}^i \bowtie_{I_{R.B_j}^i.B_j = I_{S.B_j}^i.B_j} I_{S.B_j}^i \right)$$

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

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



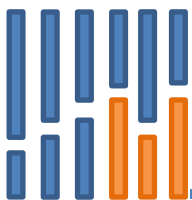
Операция тета-соединения

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

$$\tilde{P}_i = \pi_{I_{\bar{R}.F}^i \cdot A \rightarrow A_R, I_{\bar{S}.G}^i \cdot A \rightarrow A_S} \left(I_{\bar{R}.F}^i \bowtie_{F < G} I_{\bar{S}.G}^i \right);$$

$$J_i = \pi_{A \rightarrow A_R} \left(I_{\bar{R}.F}^i \right);$$

$$H_i = \pi_{A \rightarrow A_S} \left(I_{\bar{S}.G}^i \right)$$



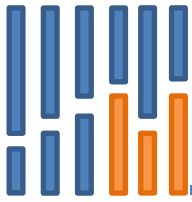
Операция тета-соединения

$$\bar{P}_i = J_i \times \bigcup_{j=i+1}^{k-1} H_j .$$

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

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

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



Пример операции тета-соединения

Определим $Q = R \bowtie S$
 $(R.B_1 + R.B_3) < S.B_1$

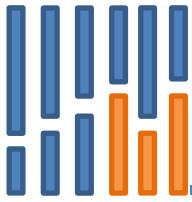
Степень фрагментации: 2

R				S		
A	B_1	B_2	B_3	A	B_1	B_2
0	10	a	0	0	5	s
1	10	z	2	1	11	y
2	5	s	1	2	3	d
3	1	s	3	3	10	a
				4	2	a



$\bar{R}$	
A	F
0	10
1	12
2	6
3	4

$\bar{S}$	
A	G
0	5
1	11
2	3
3	10
4	2



Пример операции тета-соединения

Определим $Q = R \bowtie S$

$$(R.B_1 + R.B_3) < S.B_1$$

Степень фрагментации: 2

$\bar{R}$

A	F
0	10
1	12
2	6
3	4

$\bar{S}$

A	G
0	5
1	11
2	3
3	10
4	2

$I^0_{\bar{R}.F}$

A	F
3	4
2	6

$I^0_{\bar{S}.G}$

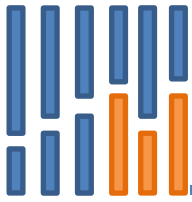
A	G
4	2
2	3
0	5

$I^1_{\bar{R}.F}$

A	F
0	10
1	12

$I^1_{\bar{S}.G}$

A	B ₁
3	10
1	11



Пример операции тета-соединения

$I^0_{\bar{R}.F}$

A	F
3	4
2	6

$I^0_{\bar{S}.G}$

A	G
4	2
2	3
0	5



$\tilde{P}_0$

$\bar{R}.A$	$\bar{S}.A$
3	0

J_0

$\bar{R}.A$
3
2

H_0

$\bar{S}.A$
4
2
0

$I^1_{\bar{R}.F}$

A	F
0	10
1	12

$I^1_{\bar{S}.G}$

A	B_1
3	10
1	11



$\tilde{P}_1$

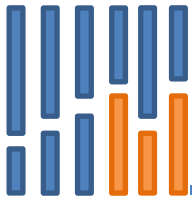
$\bar{R}.A$	$\bar{S}.A$
0	1

J_1

$\bar{R}.A$
0
1

H_1

$\bar{S}.A$
3
1



Пример операции тета-соединения

$\tilde{P}_0$		J_0	H_0	H_1
$\bar{R}.A$	$\bar{S}.A$	$\bar{R}.A$	$\bar{S}.A$	$\bar{S}.A$
3	0	3	4	3
		2	2	1
			0	



$\bar{P}_0$

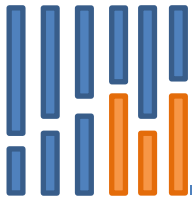
$\bar{R}.A$	$\bar{S}.A$
3	3
3	1
2	3
2	1

$\tilde{P}_1$		J_1	H_1
$\bar{R}.A$	$\bar{S}.A$	$\bar{R}.A$	$\bar{S}.A$
0	1	0	3
		1	1

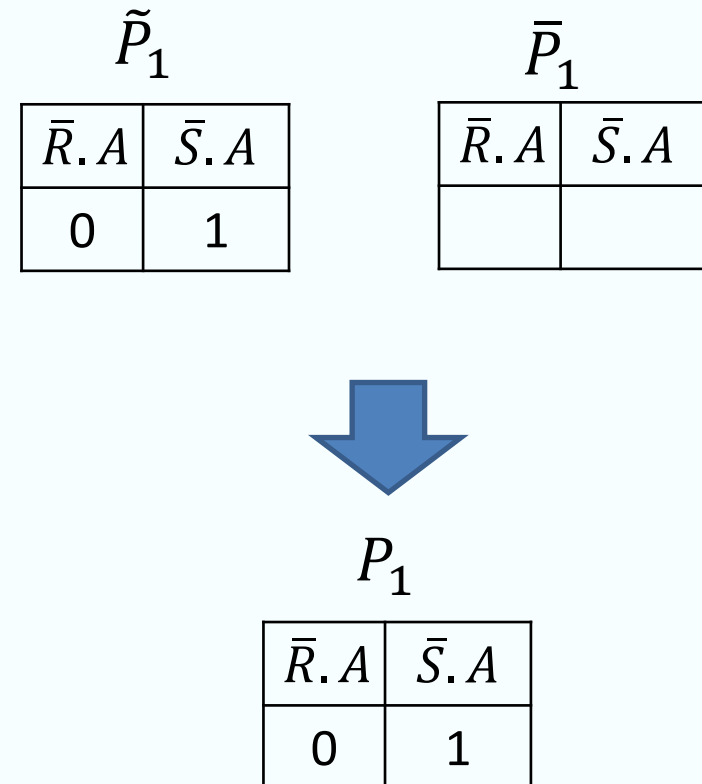
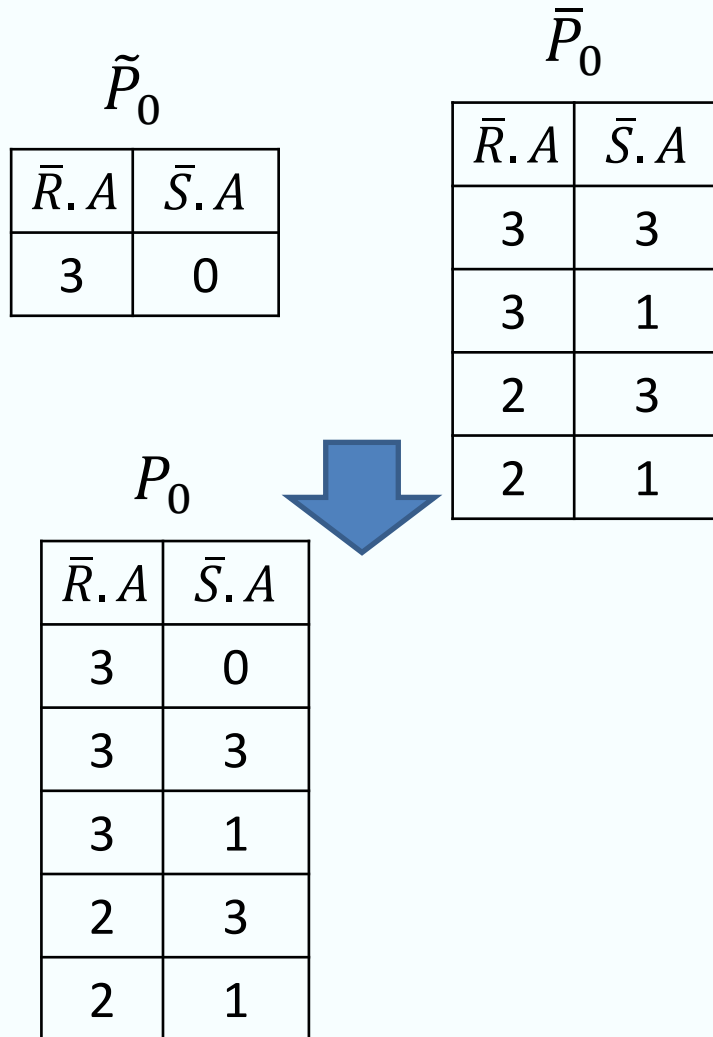


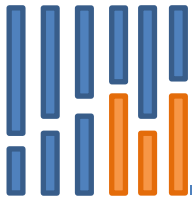
$\bar{P}_1$

$\bar{R}.A$	$\bar{S}.A$



Пример операции тета-соединения





Пример операции тета-соединения

P_0

$\bar{R}.A$	$\bar{S}.A$
3	0
3	3
3	1
2	3
2	1

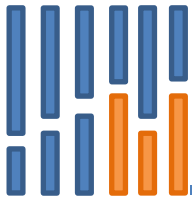
P_1

$\bar{R}.A$	$\bar{S}.A$
0	1



P

$\bar{R}.A$	$\bar{S}.A$
3	0
3	3
3	1
2	3
2	1
0	1



Пример операции тета-соединения

P

$\bar{R}.A$	$\bar{S}.A$
3	0
3	3
3	1
2	3
2	1
0	1

R

A	B_1	B_2	B_3
0	10	a	0
1	10	z	2
2	5	s	1
3	1	s	3

S

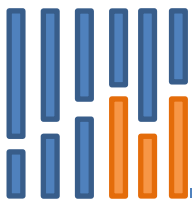
A	B_1	B_2
0	5	s
1	11	y
2	3	d
3	10	a
4	2	a

$$Q = R \bowtie S$$

$$(R.B_1 + R.B_3) < S.B_1$$

Q

$R.A$	$R.B_1$	$R.B_2$	$R.B_3$	$S.A$	$S.B_1$	$S.B_2$
3	1	s	3	0	5	s
3	1	s	3	3	10	a
3	1	s	3	1	11	y
...	...	...	...	...	...	...



Спасибо за внимание!

Вопросы?