

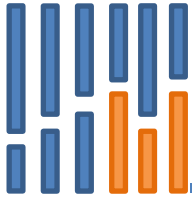
66-я научная конференция профессорско-преподавательского состава ЮУрГУ 2014

---

# Методы обработки запросов с использованием распределенных колоночных индексов

**Е.В. Иванова**

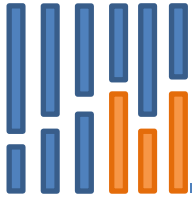
Национальный исследовательский  
Южно-Уральский государственный университет



# 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

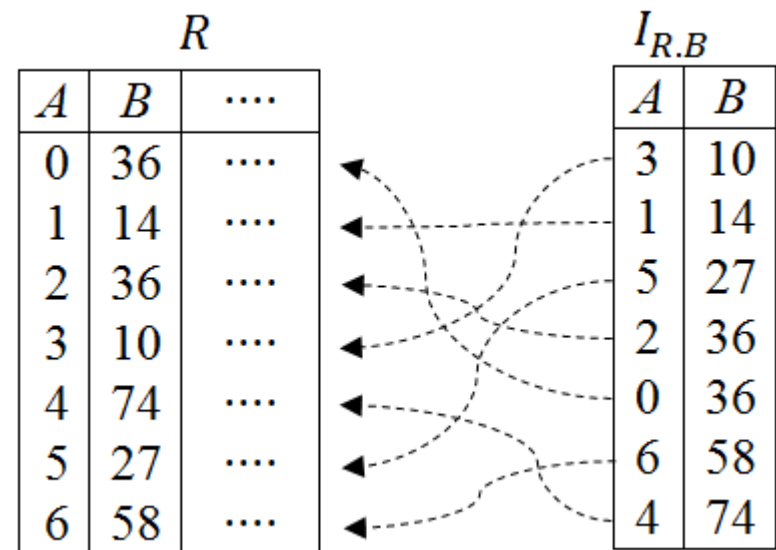
- Пусть  $R(A^*, B, \dots)$  – отношение  $R$  с первичным ключом  $A$  и атрибутом  $B$ .  $A$  представляет собой *служебный первичный ключ* и состоит из целочисленных неотрицательных элементов.
- $\mathcal{D}_B$  – домен атрибута  $B$ . На множестве  $\mathcal{D}_B$  задано отношение линейного порядка.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

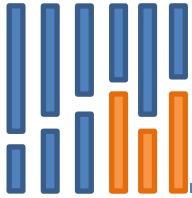
- **Колоночным индексом**  $I_{R.B}$  атрибута  $B$  отношения  $R$  называется упорядоченное отношение, удовлетворяющее следующим требованиям:

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

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

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





# Доменно-интервальная фрагментация

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

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

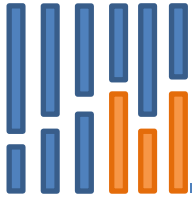
$$v_0 < v_1 < \dots < v_k;$$

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

- **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

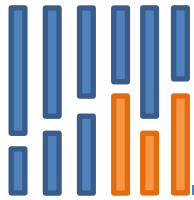
- **Функция фрагментации индекса:**

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

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

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

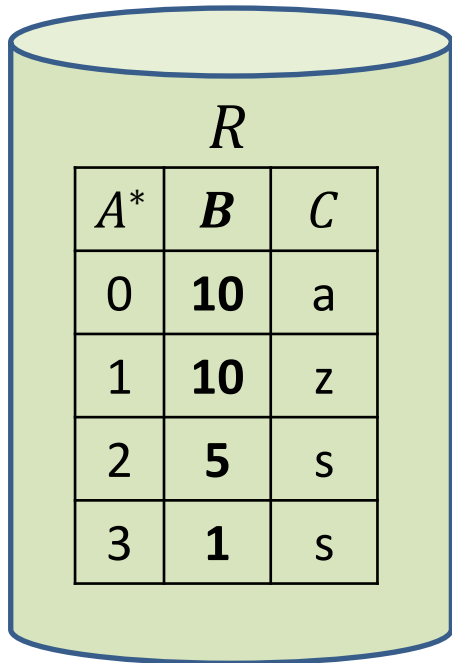
- Будем называть фрагментацию, построенную таким образом, *доменно-интервальной*.
- Количество фрагментов  $k$  будем называть *степенью фрагментации*.



# Пример доменно-интервальной фрагментации колоночного индекса

Степень фрагментации: 2

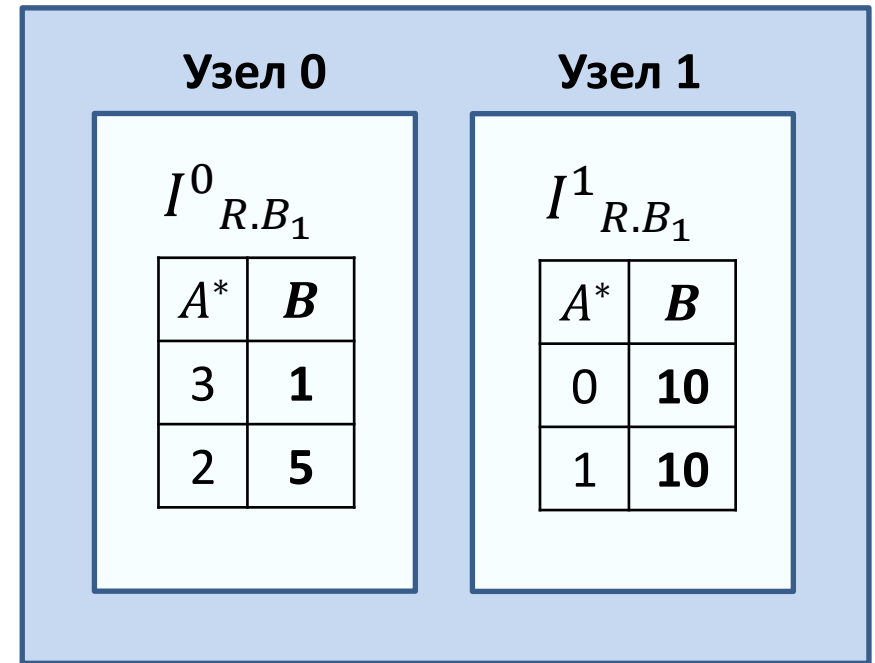
Устройство хранения

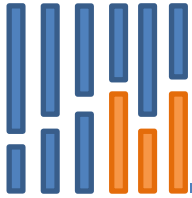


$I^0_{R.B_1}$

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

Кластерная вычислительная система





# 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  $k$ :

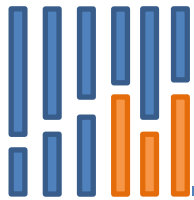
$$I_{R.B_j} = \bigcup_{i=0}^{k-1} I_{R.B_j}^i ; \quad I_{S.B_j} = \bigcup_{i=0}^{k-1} I_{S.B_j}^i .$$

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

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

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

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



# Пример операции естественного соединения

Определим  $R \bowtie S$

Степень фрагментации: 2

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

$R$

$A$	$B_1$	$B_2$	$C$
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

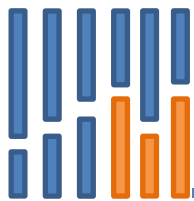
**Узел 0**

$I^0_{R.B_1}$	$I^0_{R.B_2}$	$I^0_{S.B_1}$	$I^0_{S.B_2}$																										
<table border="1"> <tr><th><math>A</math></th><th><math>B_1</math></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 border="1"> <tr><th><math>A</math></th><th><math>B_2</math></th></tr> <tr><td>0</td><td>a</td></tr> </table>	$A$	$B_2$	0	a	<table border="1"> <tr><th><math>A</math></th><th><math>B_1</math></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 border="1"> <tr><th><math>A</math></th><th><math>B_2</math></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																												

**Узел 1**

$I^1_{R.B_1}$	$I^1_{R.B_2}$	$I^1_{S.B_1}$	$I^1_{S.B_2}$																										
<table border="1"> <tr><th><math>A</math></th><th><math>B_1</math></th></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 border="1"> <tr><th><math>A</math></th><th><math>B_2</math></th></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 border="1"> <tr><th><math>A</math></th><th><math>B_1</math></th></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 border="1"> <tr><th><math>A</math></th><th><math>B_2</math></th></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																												





# Пример операции естественного соединения

Узел 0

$I^0_{R.B_1}$		$I^0_{R.B_2}$		$I^0_{S.B_1}$		$I^0_{S.B_2}$	
A	B <sub>1</sub>	A	B <sub>2</sub>	A	B <sub>1</sub>	A	B <sub>2</sub>
3	1	0	a	4	2	3	a
2	5			2	3	4	a
				0	5	2	d



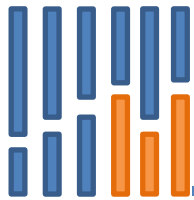
$P^0_1$		$P^0_2$	
R.A	S.A	R.A	S.A
2	0	0	3
		0	4

Узел 1

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



$P^1_1$		$P^1_2$	
R.A	S.A	R.A	S.A
0	3	2	0
1	3	3	0



# Пример операции естественного соединения

Узел 0

$P^0_1$

<i>R.A</i>	<i>S.A</i>
2	0

$P^0_2$

<i>R.A</i>	<i>S.A</i>
0	3
0	4

$P^1_1$

<i>R.A</i>	<i>S.A</i>
0	3
1	3

$P^1_2$

<i>R.A</i>	<i>S.A</i>
2	0
3	0

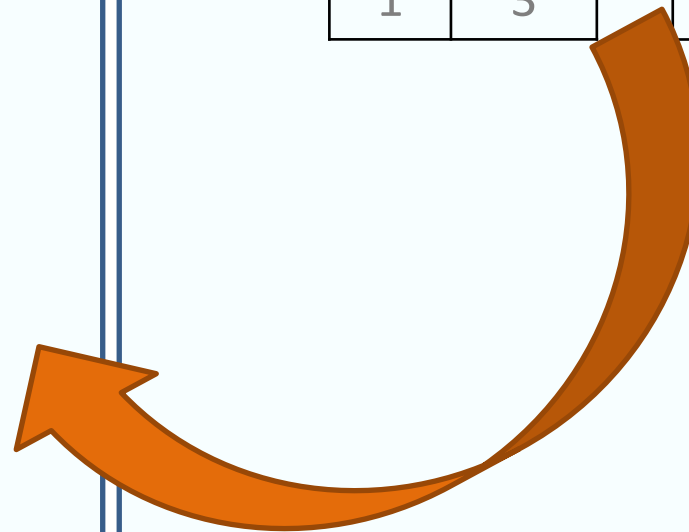
Узел 1

$P^1_1$

<i>R.A</i>	<i>S.A</i>
0	3
1	3

$P^1_2$

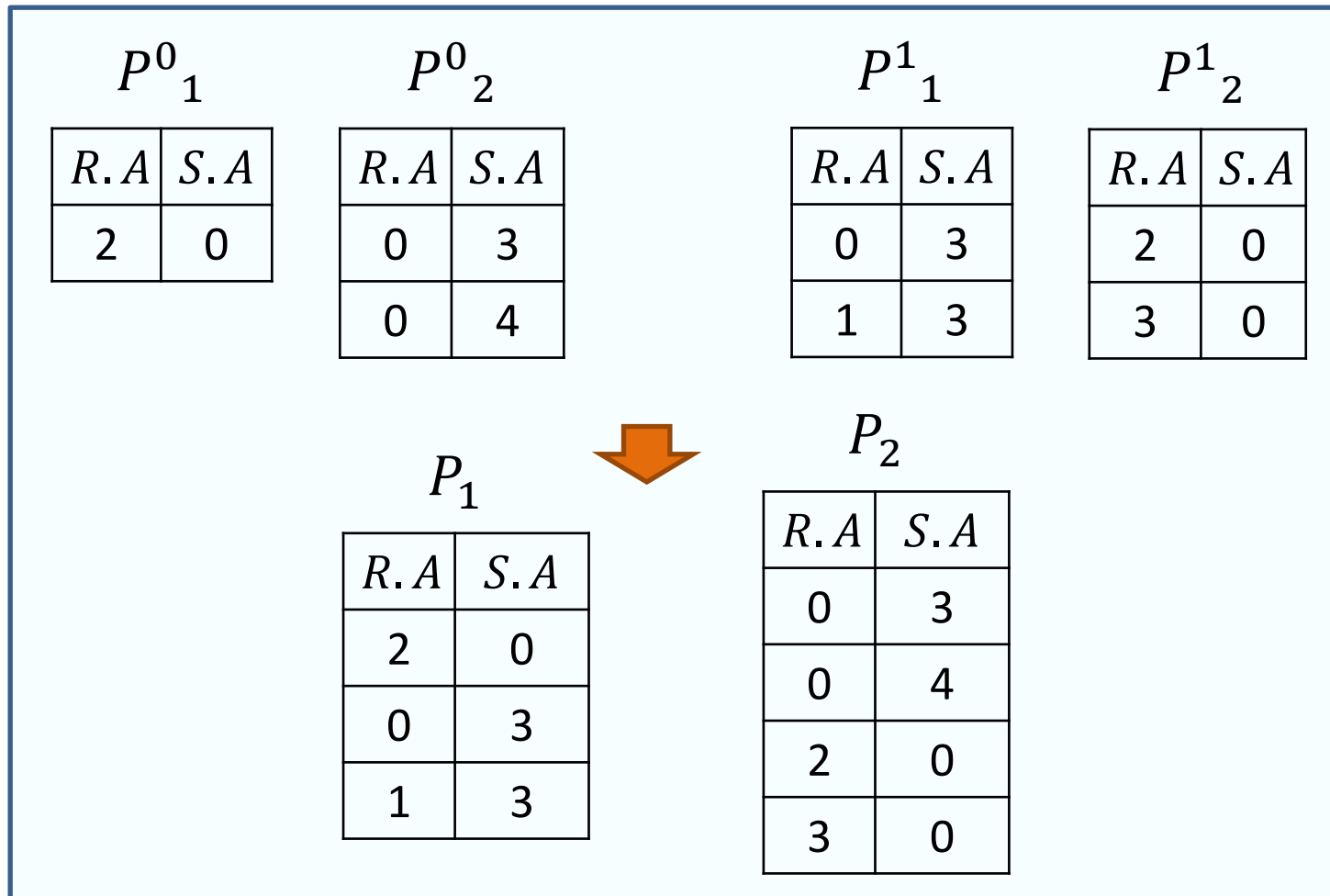
<i>R.A</i>	<i>S.A</i>
2	0
3	0



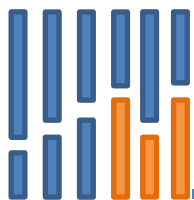


# Пример операции естественного соединения

Узел 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

$P_2$

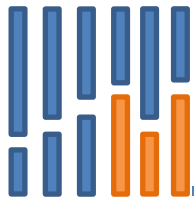
$R.A$	$S.A$
0	3
0	4
2	0
3	0



$P$

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

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



# Пример операции естественного соединения

*R*

<i>A</i>	<i>B</i> <sub>1</sub>	<i>B</i> <sub>2</sub>	<i>C</i>
0	10	a	Ni
1	10	z	Au
2	5	s	Pb
3	1	s	Ag

*S*

<i>A</i>	<i>B</i> <sub>1</sub>	<i>B</i> <sub>2</sub>	<i>D</i>
0	5	s	Pb
1	11	y	Pb
2	3	d	Ni
3	10	a	Fr
4	2	a	Ag

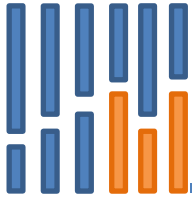
Узел 0

*P*

<i>R.A</i>	<i>S.A</i>
2	0
0	3



<i>R.A</i>	<i>R.B</i> <sub>1</sub>	<i>R.B</i> <sub>2</sub>	<i>R.C</i>	<i>S.D</i>
2	5	s	Pb	Pb
0	10	a	Ni	Fr

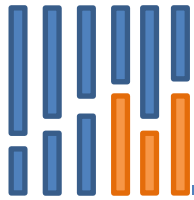


# Хэш-индекс

- 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  $h : \mathcal{D}_{B_1} \times \dots \times \mathcal{D}_{B_u} \rightarrow \mathbb{Z}_{\geq 0}$ . Хеш-индексом  $I_h(A^*, H)$  атрибутов  $B_1, \dots, B_u$  отношения  $R$  называется упорядоченное отношение, удовлетворяющее тождеству:

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

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



## Операция естественного соединения с использованием хэш-индекса

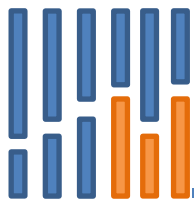
- Пусть  $R(A^*, B_1, \dots, B_u, C_1, \dots, C_v)$  и  $S(A^*, B_1, \dots, B_u, D_1, \dots, D_w)$ .
- Определим  $Q$  как результат операции  $R \bowtie S$ .
- Пусть имеется два хэш-индекса для атрибутов  $B_1, \dots, B_u$ , построенные с помощью одной и той же хэш-функции  $h$  для которых задана доменно-интервальная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

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

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

$$P^i = \pi_{I_{R,h}^i \cdot A \rightarrow A_R, I_{S,h}^i \cdot A \rightarrow A_S} \left( I_{R,h}^i \bowtie_{(I_{R,h}^i \cdot H = I_{S,h}^i \cdot H)} I_{S,h}^i \right) \quad P = \bigcup_{i=0}^{k-1} P^i$$

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



# Пример операции естественного соединения с хэш-индексом

Определим  $R \bowtie S$

Степень фрагментации: 2

$$h(b_1, b_2) = b_1 + ACSII\_КОД(b_2)$$

$R$

$A$	$B_1$	$B_2$	$C$
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

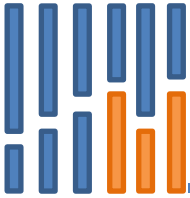
$I^1_{R.h}$

$A$	$H$
2	120
1	132

$I^1_{S.h}$

$A$	$H$
0	120
1	132





Спасибо за внимание!  
Вопросы?