

ИСПОЛЬЗОВАНИЕ РАСПРЕДЕЛЕННЫХ КОЛОНОЧНЫХ ХЕШ-ИНДЕКСОВ ДЛЯ ОБРАБОТКИ ЗАПРОСОВ К СВЕРХБОЛЬШИМ БАЗАМ ДАННЫХ*

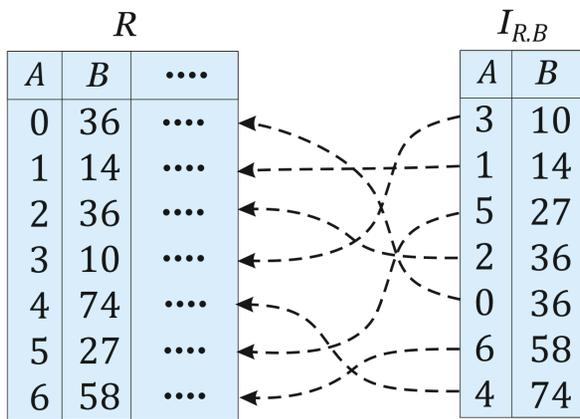
Е.В. Иванова

ivanovaev@susu.ac.ru

ФГБОУ ВПО "ЮУрГУ" (НИУ)

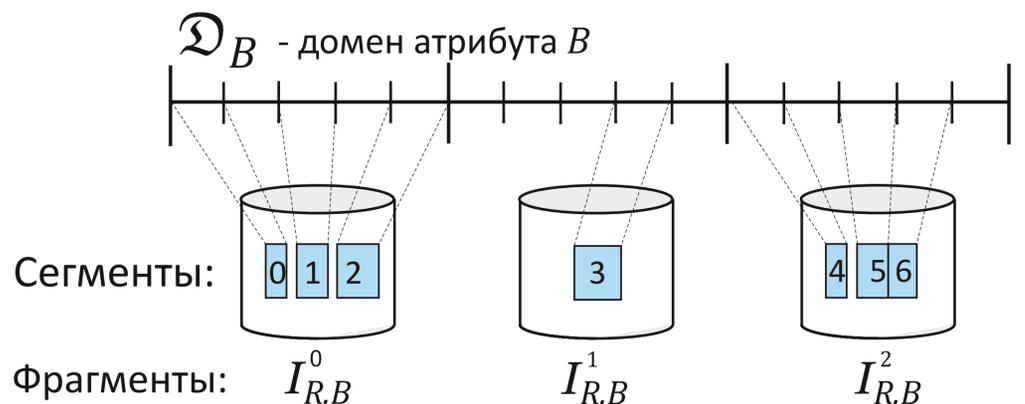
РАСПРЕДЕЛЕННЫЕ КОЛОНОЧНЫЕ ИНДЕКСЫ

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.

Доменно-интервальная фрагментация



i -тый фрагмент индекса:
 $I_{R,B}^i = \{x \mid x \in I_{R,B}; \varphi_{I_{R,B}}(x) = i\}$

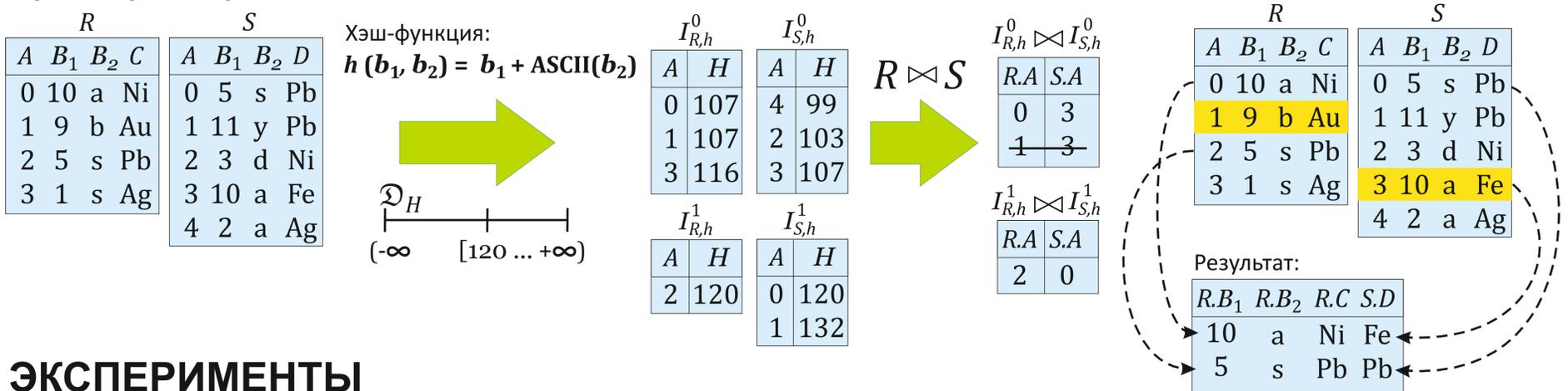
Колоночный хэш-индекс

Хэш-индекс позволяет использовать один колоночный индекс для нескольких атрибутов одного отношения.

Пусть задано отношение $R(A^*, B_1, \dots, B_U, C, \dots)$. Пусть задана хэш-функция $h: \mathcal{D}_{B_1} \times \dots \times \mathcal{D}_{B_U} \rightarrow \mathbb{Z}_{\geq 0}$. **Хэш-индексом**

$I_h(A^*, H)$ атрибутов B_1, \dots, B_U отношения R называется упорядоченное отношение: $I_h = \tau_H(\pi_{A,h(B_1, \dots, B_U)}(R))$

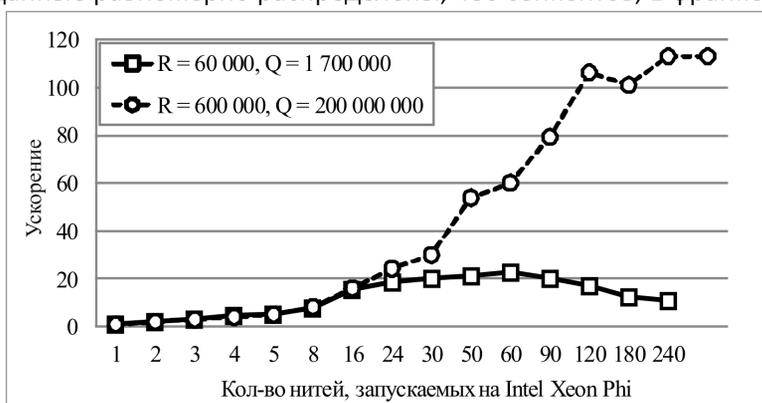
Пример операции естественного соединения с колоночным хэш-индексом



ЭКСПЕРИМЕНТЫ

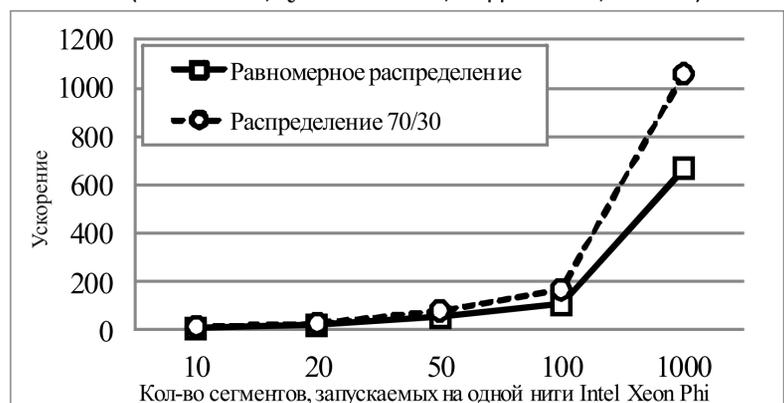
Исследование масштабируемости операции естественного соединения с использованием колоночного индекса

(данные равномерно распределены, 480 сегментов, 1 фрагмент)



Исследование масштабируемости операции естественного соединения при изменении количества сегментов

($R = 50\,000$, $Q = 1\,700\,000$, 1 фрагмент, 4 нити)



* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 12-07-00443 а.