

Обновление распределенных сжатых колоночных индексов для сверхбольших баз данных*

А.Э. Залятов, Е.В. Иванова
ФГБОУ ВПО ЮУрГУ (НИУ)

Один из подходов, применяемых для обработки сверхбольших баз данных, основывается на использовании распределенных колоночных индексов, которые обеспечивают параллельную обработку запросов на многопроцессорных вычислительных системах [1, 2]. Колоночный индекс представляет собой таблицу из двух колонок A и B , колонка A содержит уникальные ключи, колонка B – повторяющиеся упорядоченные значения. В данной работе рассматриваются целочисленные ключи и значения. Колоночный индекс сегментируется по значению B на основе доменно-интервальной фрагментации и сохраняется в оперативной памяти в виде массива структур `segment` с полями `size` (количество строк в сегменте), `ptr_key` (указатель на область памяти, где хранятся ключи сегмента) и `ptr_value` (указатель на область памяти, где хранятся значения сегмента).

В работе рассматривается подход к обновлению распределенного колоночного индекса, сжатого с помощью алгоритма RLE. Данный алгоритм эффективно сжимает в сегменте колонку B . Алгоритм RLE позволяет обрабатывать данные без их распаковки, что обеспечивает более эффективное использование оперативной памяти, но требует разработки специализированных алгоритмов для обновления сжатых данных. На рис. 1 представлен алгоритм вставки строки $(key, value)$ в сжатый сегмент колоночного индекса.

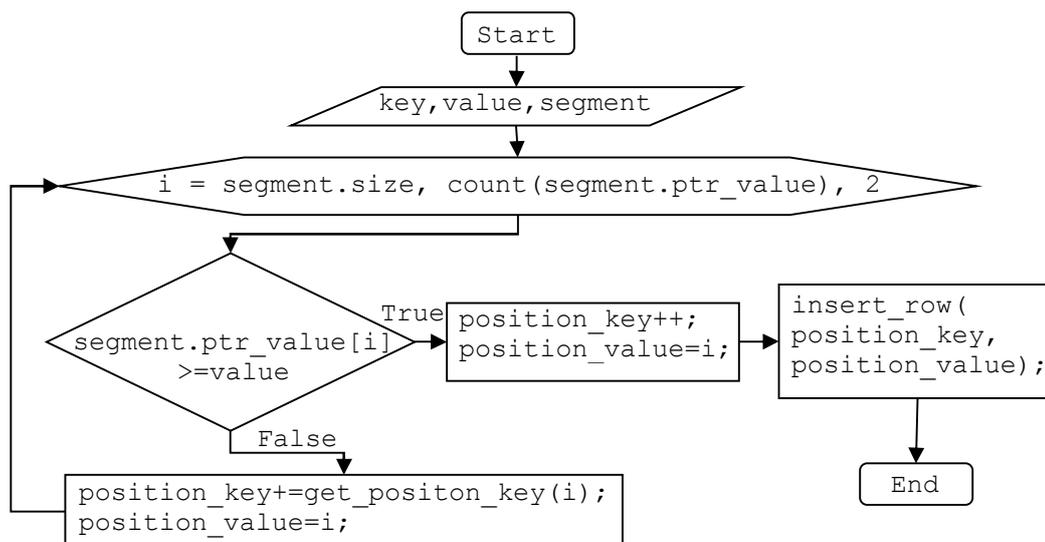


Рис. 1. Алгоритм вставки строки $(key, value)$ в сжатый сегмент колоночного индекса

Литература

1. Иванова Е.В., Соколинский Л.Б. Использование распределенных колоночных индексов для выполнения запросов к сверхбольшим базам данных // Параллельные вычислительные технологии (ПаВТ'2014): труды международной научной конференции (1–3 апреля 2014 г., г. Ростов-на-Дону). Челябинск: Издательский центр ЮУрГУ, 2014. С. 270–275.
2. Соколинский Л.Б. Параллельные машины баз данных // Природа. 2001. № 8. С. 10-17.

* 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).