

Моделирование распределенных колоночных индексов в контексте параллельных систем баз данных^{*}

С.О. Приказчиков, П.С. Костенецкий

Южно-Уральский государственный университет

В настоящее время научная и практическая деятельность человечества создает новые задачи, которые требуют обработки больших данных. Объем цифровых данных удваивается каждые два года [1], в связи с этим очевидна необходимость постоянно повышать производительность параллельных систем баз данных. Повышение производительности параллельных систем баз данных с классической аппаратной архитектурой, имеет ряд физических ограничений. Использование при обработке данных многоядерных и графических ускорителей может дать существенный прирост производительности [3], однако это требует разработки и внедрения в параллельные СУБД новых технологий обработки данных. Использование в таких системах распределенных колоночных индексов [1] представляется одним из перспективных подходов к совершенствованию архитектуры параллельных систем баз данных.

Данное незаконченное исследование посвящено расширению математической модели мультипроцессоров баз данных *DMM* [2] для моделирования параллельной обработки транзакций с применением многоядерных и графических ускорителей и разработке на ее основе методов и алгоритмов моделирования процессов параллельной обработки транзакций. Расширенная модель позволит моделировать и исследовать эффективность применения распределенных колоночных индексов в контексте параллельных систем баз данных. На базе расширенной модели будет разработана программная система – эмулятор, которая позволит исследовать различные гибридные аппаратные конфигурации параллельных систем баз данных без их аппаратной реализации.

В настоящее время получены следующие результаты: 1) исследована эффективность применения графических ускорителей для обработки запросов над сжатыми данными в параллельных системах баз данных; 2) исследована эффективность применения сжатия при передаче данных из основной памяти в память графического ускорителя [3]. Ведутся работы по формированию математического фундамента для расширения модели *DMM*. Разрабатываемая модель будет абстрагирована от аппаратного обеспечения и деталей выполнения, обеспечивая высокоуровневый подход к сравнению времени выполнения и определению характеристик различных типов запросов в параллельных колоночных СУБД. Модель будет состоять из четырех подмоделей: 1) модель аппаратной платформы, в рамках которой многопроцессорные системы будут представляться в виде ориентированного графа, будет моделироваться передача сообщений по соединительной сети, дисковые операции ввода/вывода и передача на многоядерные сопроцессоры, 2) модель операционной среды для моделирования фрагментного параллелизма и колоночных индексов, 3) стоимостная, оценивающая трудоемкость выполнения запросов к базе данных и 4) модель транзакций, для моделирования параллельной обработки смеси транзакций.

Литература

1. Ivanova E., Sokolinsky L. Decomposition of Natural Join Based on Domain-Interval Fragmented Column Indices // 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) 2015, Opatija, Croatia, May 25–29, 2015. IEEE, 2015. P. 223–226.
2. Костенецкий П.С., Соколинский Л.Б. Моделирование параллельных систем баз данных: учебное пособие. Челябинск: Издательский центр ЮУрГУ, 2012. 78 с.
3. Приказчиков С.О., Костенецкий П.С. Применение графических ускорителей для обработки запросов над сжатыми данными в параллельных системах баз данных // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2015. Т. 4. № 1. С. 64–70.

^{*} Работа выполнена при поддержке стипендии Президента РФ СП-1229.2015.5 (2015-2016 гг.): "Моделирование параллельной обработки запросов на высокопроизводительных многопроцессорных системах с графическими ускорителями"

Modeling distributed column indexes in the context of parallel database systems

Stepan Prikazchikov and Pavel Kostenetskiy

Keywords: DMM model, database multiprocessor, parallel data processing, multicore accelerators, distributed column indexes

The paper is devoted to extending of the mathematical model of database multiprocessor DMM [Kostenetskii P.S., Sokolinsky L.B. Simulation of Hierarchical Multiprocessor Database Systems // Programming and Computer Software. Vol. 39 No. 1. 2013. P. 10–24.] for parallel data processing using multicore accelerators and development on this basis of methods and algorithms for the simulation of processes in parallel processing of transactions.

Extended model allows simulating and analyzing distributed column indexes in the context of parallel database systems.