

ИССЛЕДОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ С МНОГОЯДЕРНЫМИ УСКОРИТЕЛЯМИ В РАМКАХ ПРИЛОЖЕНИЙ БАЗ ДАННЫХ

П.С. Костенецкий, А.И. Семенов

Южно-Уральский госуниверситет, Челябинск

На сегодняшний день уже существуют работы, посвященные оптимизации процесса интеллектуального анализа данных [2, 3] и ускорению обработки запросов к базе данных с использованием графических ускорителей [1, 4]. Однако работы в этих научных направлениях не привели к созданию полноценных прототипов и, тем более, промышленных параллельных СУБД. В первую очередь это связано с ограничениями скорости системной шины, по которой данные передаются между CPU и GPU. Необходимость учитывать множество узких мест в архитектуре гибридных узлов значительно усложняет разработку параллельных систем баз данных. В связи с этим возникает вопрос о необходимости исследования и анализа гибридных аппаратных архитектур при помощи математических моделей. Данная работа посвящена моделированию подобных систем баз данных и поиску оптимальных гибридных аппаратных архитектур для обработки запросов не только с использованием центрального процессора, но и с использованием GPU либо многоядерных ускорителей класса Intel Xeon Phi.

В качестве первого этапа работы был реализован эмулятор СУБД, моделирующий параллельную обработку запросов [7] на графическом ускорителе, а также обработку этих же запросов на многоядерном центральном процессоре. При помощи эмулятора исследована эффективность использования графических ускорителей для хранения и обработки баз данных. В результате исследования определено, что с использованием многоядерных графических ускорителей можно добиться многократного ускорения обработки запроса SELECT или вычисления агрегатных функций. По сравнению с обработкой на 4 ядрах центрального процессора удалось добиться ускорения в 83 раза при обработке на 512 нитях GPU. В то же время значительно ускорить алгоритм JOIN с использованием GPU не удается [6]. Исходя из результатов экспериментов, можно сделать вывод, что эффективным будет следующий комбинированный алгоритм работы системы баз данных: обработка запроса JOIN выполняется на ядрах центрального процессора. В это же время на графическом ускорителе, установленном в вычислительном узле, выполняется обработка запросов SELECT или вычисление агрегатных функций, таких как MIN, MAX, SUM, AVG или COUNT. SELECT является наиболее распространенным запросом SQL, поэтому использование такой схемы работы может позволить значительно снизить нагрузку на центральные процессоры. Для реализации данного подхода будет эффективным использование нескольких многоядерных ускорителей. Каждый ускоритель обрабатывает свою часть отношения, над которым требуется выполнить запрос. Далее полученные результаты объединяются в результирующее отношение. Кроме того, ускорители могут быть установлены в вычислительные узлы, объединенные в кластер.

Для реализации подобного комбинированного подхода к обработке запросов в сверхбольших базах данных потребуются новые аппаратные архитектуры систем баз данных, содержащие ускорители GPU либо MISC и жесткие диски. Сборка прототипов подобных высокопроизводительных кластеров с гибридными узлами является чрезвы-

чайно дорогостоящей. В связи с этим для выбора оптимальной конфигурации системы баз данных предварительно необходимо выполнять моделирование возможных аппаратных архитектур при помощи математических моделей. Для моделирования аппаратных архитектур параллельных систем баз данных наиболее подходит *модель мульти-процессоров баз данных DMM* [5]. Однако для анализа гибридных иерархических архитектур необходимо выполнить расширение для данной модели. Это является направлением дальнейших исследований.

Работа выполнена при поддержке РФФИ, грант № 12-07-31082 (2012-2013 гг.)

Литература

1. Bakum P., Skadron K. Accelerating SQL database operations on a GPU with CUDA // Proceedings of the 3rd Workshop on General-Purpose Computation on Graphics Processing Units (GPGPU'10), Pittsburgh, USA, March 14, 2010. ACM. P. 94–103.
2. Ding S., He J., Yan H., Suel T. Using graphics processors for high performance IR query processing. In WWW '09: Proceedings of the 18th international conference on World wide web, New York, NY, USA, 2009. ACM. P. 421–430.
3. Fang W., Lau K., Lu M. et al. Parallel data mining on graphics processors: Technical report, Hong Kong University of Science and Technology, 2008.
4. Govindaraju N., Lloyd B., Wang W. et al. Fast computation of database operations using graphics processors. In SIGGRAPH '05: ACM SIGGRAPH 2005 Courses, New York, NY, USA, 2005. ACM. P. 206.
5. Костенецкий П.С. Моделирование параллельных систем баз данных для вычислительных кластеров // Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность: Труды всероссийской научной конференции (21-26 сентября 2009 г., Новороссийск). М.: Изд-во МГУ. 2009. С. 300–304.
6. Костенецкий П.С., Семенов А.И. Разработка методов обработки запросов к базе данных на многоядерных ускорителях, поддерживающих технологию CUDA // Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции (17-22 сентября 2012 г., Новороссийск). М.: Изд-во МГУ, 2012. С. 435–441.
7. Соколинский Л.Б. Организация параллельного выполнения запросов в многопроцессорной машине баз данных с иерархической структурой // Программирование. 2001. № 6. С. 13–19.