

ОТЗЫВ

на автореферат диссертации Ивановой Елены Владимировны
«Методы параллельной обработки сверхбольших баз данных
с использованием распределенных колоночных индексов», представленной на
соискание ученой степени кандидата физико-математических наук по специальности
05.13.11 — «Математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей».

Диссертационное исследование Е.В. Ивановой предлагает способы решения проблемы обработки сверхбольших баз данных. Актуальность данной проблемы связана с тем, что рост объемов информации, требующей хранения и обработки, опережает рост производительности серверных систем. Решение, описанное в автореферате, заключается в разработке на основе колоночной модели хранения информации вспомогательных структур данных (колоночных индексов), позволяющих ускорить выполнение ресурсоемких реляционных операций на кластерных вычислительных системах с многоядерными ускорителями. К колоночным индексам применяется доменно-интервальная фрагментация, позволяющая минимизировать обмены данными между вычислительными узлами при выполнении реляционных операций. Разработанные модели были реализованы в виде программной системы «Колоночный сопроцессор КСОП», проведены эксперименты, демонстрирующие их эффективность.

Для анализа результатов работы в дополнение к автореферату необходимо использовать саму диссертацию, так как в автореферате не полностью описаны эксперименты и не приведены теоретические результаты, полученные в ходе исследования.

К недостаткам работы можно отнести следующее. При анализе эффективности разработанного решения по внешним характеристикам используется ограниченный набор показателей – ускорение по времени выполнения запроса, что не является достаточным. Необходимо также использовать показатели эффективности использования вычислительных ресурсов и рассматривать изменение характеристик решения при согласованном изменении размера задачи и размера вычислительного комплекса.

В работе также не исследовано влияние характеристик дисковой подсистемы и подсистемы кэширования на эффективность предлагаемого решения. А именно, по-видимому, наблюдаемое при сравнительном анализе скорости выполнения SQL запроса ускорение почти в 300 раз при $sel=0.0005$, может быть связано с «эффектом кэша», и может исчезнуть, если на сервере будут выполняться и другие приложения, использующие дисковую подсистему, или при одновременном выполнении нескольких запросов.

Проведенные эксперименты демонстрируют работу разработанного решения для одного вида SQL запросов, и не дают возможности оценить эффективность решения в более сложных сценариях работы, что ограничивает практическую применимость результатов работы на данном этапе.

Также было бы интересно провести сравнение с современными алгоритмами, которые используют bitmap индексы со сжатием (в т.ч. WAN), реализованные в FastBit, ParStream.

Тем не менее, я считаю, что исследования выполнены на высоком уровне. По возможности, учитывая актуальность и важность темы исследований, их следует продолжить. Работа соответствует требованиям, предъявляемым к кандидатским диссертациям по специальности 05.13.11 — «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Е.В. Иванова достойна присуждения ученой степени кандидата физико-математических наук.

Доцент кафедры «Алгоритмы
и технологии программирования»
Московского физико-технического
института (государственного университета),
кандидат физ.-мат. наук

А.С. Хританков

«11» декабря 2015 г.

E-mail: anton.khritanov@gmail.com

Тел.: +7 (495) 408-78-81

Адрес организации: Московская область,
г. Долгопрудный, Институтский пер., 9

Подпись А.С. Хританкова заверяю:

СПЕЦИАЛИСТ ПО КАДРАМ
ОТДЕЛА КАДРОВ
СТАРЦЕВА Л. Е.

