



Отзыв рецензента на бакалаврскую работу

студента кафедры Системного Программирования СПбГУ

Смирнова Михаила Александровича

«АВТОМАТИЧЕСКАЯ НАСТРОЙКА ПАРАМЕТРОВ ТИРИНГА, ЗАВИСЯЩАЯ ОТ ВХОДЯЩЕЙ НАГРУЗКИ, В СИСТЕМЕ ХРАНЕНИЯ ДАННЫХ»

В дипломной работе Михаила Смирнова рассматривается проблема гибридного хранения данных, а именно один из подходов – Data Tiering. Многоуровневая организация хранения данных является компромиссом между использованием дорогих, но быстрых устройств (SSD, NVME SSD) и более дешевых и медленных, но обладающих большим объемом (HDD). В работе Михаила рассматривались несколько проблем связанных с таким подходом, а именно определение горячих и холодных областей для последующего переноса, реализация более чем 2 уровней тира, уменьшение количества метаданных, необходимых для работы. Данные задачи очень важны для корректной и быстрой работы тира, в современном мире проводится много исследований по данным вопросам.

В дипломной работе были проведены обзоры технических решений от EMC, IBM и академических исследований по теме определения горячих данных. Обзоры технических решений достаточно поверхностны, но найти детальное описание используемых технологий бывает очень сложно. Многие компании не публикуют подробные описания, либо технологии являются защищены патентами.

Большое внимание в работе уделяется алгоритму основанному на Bloom Filter. Данный алгоритм хорош в том плане, что позволяет достаточно хорошо определять горячие области и в то же время сокращает объем метаданных. С практической точки зрения, большая работа была проделана при реализации алгоритмов. Они были реализованы, модифицированы и оптимизированы в виде модулей ядра Линукс. Программирование на С в ядре имеет определенные сложности, связанные с отладкой и использованием структур ядра операционной системы, поэтому часть работы посвящена проверке корректности работы алгоритмов и измерениям производительности.

Полученные результаты имеют практическую значимость, поскольку полученный алгоритм показал высокую производительность при эффективной организации метаданных.

К недостаткам можно отнести то, что при проведении исследований и практической реализации мало внимания было уделено скорости миграции данных. Например, сейчас миграция реализована как последовательный процесс, но можно его сделать асинхронным.

В целом работа написана понятным языком, содержит поясняющие рисунки, имеют четкую структуру. Содержится несколько опечаток, которые не мешают пониманию материала. Считаю, что работа заслуживает оценки **«отлично»**

Руководитель исследовательской лаборатории RAIDIX

к. т. н. Лазарева С. В.

