

Исследование возможности использования хеширования при построении управляющих таблиц дискового кэша в высокопроизводительных системах хранения данных

М.А. Сибиряков

Поволжский государственный технологический университет

Системы хранения данных (СХД) являются неотъемлемой частью практически любой организации начиная с небольшого офиса, и, заканчивая крупными компаниями и предприятиями. Если раньше для хранения данных вполне хватало нескольких жестких дисков, то сегодня этого уже не достаточно. Объемы хранимых данных за последние 25 лет колоссально возросли, к 2020 году объем хранимой информации может вырасти примерно в 50 раз. Разработчики СХД постоянно создают новые архитектуры и программное обеспечение, пытаясь удовлетворить всевозрастающий спрос на хранилища данных.

В рамках предшествующих исследований были проанализированы основные архитектуры высокопроизводительных систем хранения данных таких компаний как IBM, EMC и Hitachi, являющихся мировыми лидерами в этой области [1]. В процессе анализа было выяснено, что в последние годы рост производительности за счет совершенствования аппаратных компонентов практически достиг своего предела. Наметила тенденция увеличения производительности путем масштабирования аппаратных компонентов. В высокопроизводительных системах хранения используются новейшие шины ввода/вывода, процессоры и т.д., однако, как и в любой вычислительной системе, узким местом остаётся подсистема памяти. Производители СХД пытаются решить эту проблему путем увеличения объемов буферной памяти (кэш-памяти). Например, архитектура Hitachi Lighting 9960 в 2000 году имела 32 Гб буферной памяти, а новейшие архитектуры (например, Hitachi VSP, EMC Symmetrix V-Max, IBM DS8870) имеют уже 1 Тб памяти и это далеко не предел. Однако, путем простого наращивания невозможно значительно повысить производительность таких систем. Необходимо также создавать новые или оптимизировать существующие методы управления кэш-памятью, которые включают в себя множество сложных алгоритмов обработки операций ввода/вывода.

Также был проведен детальный анализ существующих методов управления кэш-памятью в СХД, в результате были выявлены их достоинства и недостатки [2]. Самый новый из методов (взят в качестве базового метода для исследований) лишен основных недостатков предшественников, однако опираясь на их основы, он также использует списковые структуры (двухсвязные кольцевые списки) для построения управляющих таблиц, применяемые для удобства реализации алгоритма вытеснения данных LRU.

Недостатком двухсвязного кольцевого списка является увеличение времени поиска элемента при увеличении объемов кэш-памяти системы хранения данных, и, как следствие, снижение времени выполнения алгоритмов обработки кэшируемых данных. В качестве решения данной проблемы предлагается использовать другую структуру данных – хеш-таблицу. В связи с этим основной целью данной работы является исследование возможности применения структуры хеш-таблиц для построения управляющих таблиц кэша для систем хранения данных с большим объемом кэш-памяти. Для выполнения дальнейших исследований планируется использование компьютерного имитационного моделирования.

Литература

1. Сибиряков М.А., Васяева Е.С. Сравнительный анализ основных моделей интеллектуальных систем хранения данных в процессе их эволюции // Информационные технологии в профессиональной деятельности и научной работе. 2012. Ч1. С. 99-105
2. Vasyaeva E.S., Sibiryaev M.A., Koshpaev A.A. Analysis and comparison of cache memory control methods in storage systems // In the World of Scientific Discoverie. Krasnoyarsk: Publishing House Science and Innovation Center, 2014. №10 (58) (Natural & Technical Sciences) p. 263-280