



САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

федеральное государственное автономное
образовательное учреждение высшего образования
«Самарский национальный исследовательский университет
имени академика С.П. Королева»

ул. Московское шоссе, д. 34, г. Самара, 443086
Тел.: +7 (846) 335-18-26, факс: +7 (846) 335-18-36
Сайт: www.ssau.ru, e-mail: ssau@ssau.ru
ОКПО 02068410, ОГРН 1026301168310,
ИНН 6316000632, КПП 631601001

У Т В Е Р Ж Д А Ю

Первый проректор –
проректор по научно-исследовательской
работе
доктор технических наук, доцент



А.Б. Прокофьев

« 03 » 09 2021 г.

03.09.2021 № 22-4493/1

На № _____ от _____

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертацию Потапова Данилы Романовича

«Разработка и исследование метода и алгоритмов адаптации ассоциативного
контейнера данных» по специальности 05.13.17 – теоретические основы
информатики, представленной к защите на соискание ученой степени
кандидата физико-математических наук

Актуальность работы

Диссертационная работа Потапова Данилы Романовича посвящена созданию алгоритмов адаптации ассоциативного контейнера данных к изменяющимся условиям работы и структуре нагрузки.

В настоящее время формируется понимание необходимости введения нового слоя в стек программного обеспечения предприятия – dataware или информационного обеспечения. Информационное обеспечение – это уровень абстракции, который позволяет управлять данными как важнейшим корпоративным ресурсом, отделенным от любых других зависимостей. Предыдущие поколения систем управления данными были точечными продуктами или инструментами, решающими конкретную проблему. Концептуально это похоже на сетевые устройства или сервисы, рассматриваемые как управляемый ресурс. Информационное обеспечение эффективно оперирует разнообразными типами данных, способами доступа к

данным и системными инструментами, необходимыми для управления данными как корпоративным ресурсом, независимо от базовой инфраструктуры и местоположения.

Информационное обеспечение предоставляет предприятиям согласованный подход к управлению и защите данных. Набор стандартных API-интерфейсов позволяет использовать данные широким спектром приложений и инструментов. Разнообразие и большой объём данных преодолеваются за счёт возможности обработки данных в разных местах с использованием как локальной, так и облачной аппаратной инфраструктуры, а также контейнеров.

В этих условиях актуальной задачей является разработка контейнеров данных, которые изменяют свое поведение при изменении условий работы или нагрузки, оптимизируют свою эффективность за счёт использования различных политик и настройки параметров кэширования. На данный момент не существует универсального алгоритма кэширования, удовлетворяющего тому разнообразию процедур доступа к данным, который необходимо обеспечить при подключении широкого спектра приложений к информационному слою предприятия. Возникает необходимость моделирования и разработки ассоциативного самоадаптирующегося контейнера данных, который изменяет логику работы и свои параметры в зависимости от нагрузки и условий работы.

Таким образом, тема диссертации Потапова Данилы Романовича актуальна и соответствует специальности 05.13.17 – теоретические основы информатики.

Структура и содержание диссертационной работы, автореферата характеризуется логической связностью и единством.

Диссертация состоит из четырёх глав, заключения, списка литературы из 151 наименования. Общий объём диссертации составляет 137 страниц.

Новизна исследований и полученных результатов

Целью диссертационного исследования является разработка и исследование самоадаптирующегося контейнера данных.

Для достижения поставленной цели автор разработал формализованную модель самоадаптирующегося контейнера данных, включающую основной ассоциативный контейнер (тип ассоциативной структуры данных и её параметры), алгоритм кэширования и размер кэша, алгоритм адаптации в зависимости от типа и параметров нагрузки с учётом внешних ограничений. В

общем виде алгоритм адаптации может изменять основной ассоциативный контейнер, алгоритм кэширования и размер кэша. В диссертации исследован только один из аспектов адаптации – оптимизация размера кэша по времени обработки запроса в зависимости от параметров нагрузки при чтении данных и соотношения скоростей кэша и основного хранилища. Вместо разработки алгоритма адаптации, подбирающего оптимальный кэширующий контейнер, предложен новый алгоритм кэширования, который имеет механизм адаптации к текущей нагрузке, периодически изменяющий состояние кэша в зависимости от значений интервального статистического ряда, вычисляемого по текущим запросам на чтение из контейнера.

Предложенные в диссертации механизмы адаптации контейнера данных используют достаточно простую модель нагрузки на основе трасс запросов на чтение в виде целочисленных ключей, имеющих нормальное распределение. Поскольку более адекватной представляется модель данных, в которой учитываются несколько одномерных нормальных распределений с различными весами, то в диссертации разработан алгоритм определения параметров сложной нагрузки, состоящей из смеси гауссовских распределений, отличающийся применением метода Expectation Maximization с использованием буфера.

Важной составляющей диссертационной работы являются вычислительные эксперименты, в ходе которых получены зависимости размера кэша от среднеквадратичного отклонения ключей в трассе и соотношения скоростей кэша и основного хранилища, исследован модуль определения параметров сложной нагрузки и проведено сравнение адаптивного кэширующего контейнера данных с использованием интервального статистического ряда с рядом известных и эффективных алгоритмов кэширования причём не только на модельных трассах ключей с нормальным распределением, но и на различных, в том числе реальных трассах, на которых оценивается эффективность алгоритмов кэширования.

В диссертации получены следующие результаты, которые могут быть квалифицированы как обладающие научной новизной:

1. Предложена формализованная модель самоадаптирующегося контейнера данных, включающая основной ассоциативный контейнер (тип ассоциативной структуры данных и её параметры), алгоритм кэширования и размер кэша, алгоритм адаптации в зависимости от типа и параметров нагрузки с учётом внешних ограничений.

2. Получена зависимость размера кэширующего контейнера от среднеквадратичного отклонения ключей в трассе и соотношения скоростей кэша и основного хранилища, позволяющая определить минимальный по времени работы размер кэша.
3. Разработан алгоритм определения параметров сложной нагрузки из смеси гауссовских распределений, отличающийся применением метода Expectation Maximization с использованием буфера.
4. Разработан новый алгоритм кэширования, который имеет механизм адаптации к текущей нагрузке, периодически изменяющий состояние кэша в зависимости от значений интервального статистического ряда, вычисляемого по текущим запросам на чтение из контейнера.

Достоверность научных положений диссертации

Достоверность научных положений, выводов и результатов диссертации обосновывается в целом корректным использованием математического аппарата, в частности, дискретной математики и математической статистики, подтверждается выполненными вычислительными экспериментами. Полученные диссертантом результаты согласуются с результатами, полученными другими исследователями.

Основные результаты диссертации прошли апробацию на научных конференциях, опубликованы в рецензируемых научных изданиях.

Значимость полученных результатов для науки и практики

Значимость результатов *для науки* определяется тем, что в работе:

- предложена и на теоретическом уровне исследована формализованная модель самоадаптирующегося контейнера данных;
- выявлено три направления адаптации контейнера данных в зависимости от типа и параметров нагрузки с учётом внешних ограничений: основной ассоциативный контейнер, алгоритм кэширования и размер кэша;
- исследован один из аспектов адаптации – оптимизация размера кэша по времени обработки запроса в зависимости от параметров нагрузки при чтении данных и соотношения скоростей кэша и основного хранилища;
- в рамках направления адаптации алгоритма кэширования предложен новый алгоритм кэширования, который имеет механизм адаптации к текущей

нагрузке, периодически изменяющий состояние кэша в зависимости от значений интервального статистического ряда, вычисляемого по текущим запросам на чтение из контейнера.

Значимость результатов, полученных в диссертации, для практики обусловлена необходимостью повышения эффективности существующих контейнеров данных и политик кэширования. Созданные в диссертации алгоритмы, реализующие некоторые модули самоадаптирующегося контейнера данных, позволяют увеличить эффективность работы нового слоя в стеке программного обеспечения предприятия – dataware или информационного обеспечения.

Рекомендации по использованию результатов диссертации

Результаты диссертации рекомендуются к использованию в организациях, которые создают распределённые программные системы обработки данных, являются интеграторами сложных программных систем, имеющими в своём составе широкий спектр приложений с различными паттернами доступа к данным, например ООО «Открытый код», ООО «Холмонт Девелопмент».

Также результаты диссертации могут быть полезны для ряда образовательных программ Воронежского государственного университета, Самарского национального исследовательского университета имени академика С.П. Королева и других российских вузов, осуществляющих подготовку бакалавров и специалистов в области разработки программных систем обработки данных, в том числе распределённых.

Замечания по диссертационной работе

1. Декларируется научный результат, в котором предложен метод адаптации ассоциативного контейнера данных, позволяющий изменять структуру данных основного хранилища данных, алгоритм кэширования и размер кэша в зависимости от условий работы и нагрузки на контейнер. Фактически в диссертации указано, что исследование ограничивается формализацией способа задания самоадаптирующегося контейнера данных, поиском зависимости размера кэша от параметров нагрузки при чтении данных, а также разработкой нового адаптивного алгоритма кэширования с использованием интервального статистического ряда, вычисляемого по

текущим запросам на чтение из контейнера. Собственно метод адаптации ассоциативного контейнера данных в диссертации не сформулирован.

2. В диссертационной работе не проведены исследования по идентификации параметров сложной нагрузки как смеси нормальных распределений для тестовых трасс, на которых оценивается эффективность алгоритмов кэширования, которые были использованы при исследовании эффективности адаптивного кэширующего контейнера данных с использованием интервального статистического ряда (п. 4.3 диссертации). Такие исследования могли бы стать необходимым этапом для проведения исследований, указанных в следующем пункте.
3. В диссертационной работе не проведены комплексные исследования эффективности адаптивного кэширующего контейнера, в котором и размер кэша также автоматически определяется в процессе вычислительного эксперимента на основе текущей оценки параметров сложной нагрузки как смеси нормальных распределений. Таким образом, разработанный модуль определения параметров сложной нагрузки так и не был использован в самоадаптирующихся контейнерах данных.
4. Некоторые из приведенных в первой главе сведений, в частности, подробное описание устройства существующих алгоритмов кэширования, в дальнейшем никак не используются, поэтому такое подробное рассмотрение не имеет смысла.
5. В разделе 2.8 отсутствует обоснование выбора соотношения скоростей между основным хранилищем и кэшем в качестве главного параметра условий адаптации для анализа зависимости размера кэша от параметров нагрузки и условий работы.
6. В формулировках Лемм 1 и 3 в разделе 2.8 используются качественные сравнения соотношений скоростей хранилищ – достаточно небольшие и достаточно большие, необходимость которых в дальнейших формализованных рассуждениях отсутствует.
7. Блок-схема на рис. 3.7 не использует общепринятые соглашения для отображения блок-схем, в результате возникает неоднозначность её интерпретации – на схеме, скорее, используется шаблон Read-Through, а не шаблон Cache-aside, как это указано в тексте при описании данной блок-схемы.

8. В диссертации и автореферате присутствуют опечатки и стилистические неточности.

Заключение

Диссертация Д.Р. Потапова является завершенной научно-квалификационной работой, выполненной на актуальную тему. Научные результаты, полученные диссертантом, имеют определенное значение для науки и практики. Выводы и положения работы в основном достоверны и обоснованы. Диссертация по ее целям, задачам, содержанию, положениям, выносимым на защиту, соответствует заявляемой специальности.

Результаты, полученные в диссертации, опубликованы в 12 работах, пять из которых - в рекомендуемых ВАК РФ рецензируемых научных изданиях. Получено свидетельство о регистрации программы для ЭВМ.

Автореферат полностью отражает содержание диссертации.

Подводя итог, можно констатировать, что диссертационная работа Потапова Данилы Романовича «Разработка и исследование метода и алгоритмов адаптации ассоциативного контейнера данных» по своему научному уровню, практической значимости, степени новизны полученных результатов соответствует всем требованиям Положения о присуждении ученых степеней (утверждено постановлением Правительства РФ от 24 сентября 2013 г. №842, в ред. от 01.10.2018 с изм. 26.05.2020), предъявляемым к кандидатским диссертациям. В ней изложены новые научно обоснованные решения и разработки, имеющие существенное значение для развития отрасли информационных технологий. Автор диссертации Потапов Данила Романович заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.17 – Теоретические основы информатики.

Диссертационная работа Потапова Данилы Романовича «Разработка и исследование метода и алгоритмов адаптации ассоциативного контейнера данных», автореферат и настоящий отзыв на диссертацию были заслушаны и обсуждены на заседании научно-технического семинара кафедры геоинформатики и информационной безопасности федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева», протокол № 3 от 30 августа 2021 года.

Отзыв составили:

Руководитель НТС, заведующий кафедрой
геоинформатики и информационной безопасности
д.т.н., профессор



Сергеев Владислав Викторович

Профессор кафедры геоинформатики
и информационной безопасности
д.ф.-м.н., доцент



Мясников Владислав Валерьевич

Сведения о составителях отзыва:

Сергеев Владислав Викторович, доктор технических наук, профессор.

Защитил докторскую диссертацию в 1993 году по специальности 05.13.16 – Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях. Ученое звание профессора получил в 1996 году по кафедре технической кибернетики.

Должность: директор института информатики, математики и электроники, заведующий кафедрой геоинформатики и информационной безопасности.

Контактный телефон: +7(846) 267-49-06, e-mail: vserg@geosamara.ru

Мясников Владислав Валерьевич, доктор физико-математических наук, доцент.

Защитил докторскую диссертацию в 2008 году по специальности 05.13.17 – Теоретические основы информатики. Учёное звание доцента получил в 2006 году по кафедре геоинформатики.

Должность: профессор кафедры геоинформатики и информационной безопасности

Контактный телефон: +7(846) 267-49-04, e-mail: vmyas@geosamara.ru

Сведения об организации:

федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» (Самарский университет).

Адрес: ул. Московское шоссе, д. 34, г. Самара, 443086

<http://ssau.ru/> , телефон: +7 (846) 335-18-26, e-mail: ssau@ssau.ru