

Ю
льного директора
ЛВИС»
Ю.В. Мосолова

23 апреля

2025 г.

ОТЗЫВ

ведущей организации на диссертационную работу

Мунермана Виктора Иосифовича «Алгебраические модели и методы для
разработки программно-аппаратных комплексов массовой обработки

данных», представленную на соискание учёной степени доктора технических
наук по специальности 2.3.5 – «Математическое и программное обеспечение
вычислительных систем комплексов и компьютерных сетей»

Актуальность темы исследования

Исследование, результат которого представлен в диссертации
В.И. Мунермана, посвящено разработке алгебраических моделей и методов
для проектирования программно-аппаратных комплексов, реализующих
массовую обработку данных. Автор рассматривает широко распространенный
класс обработки данных, в котором в обработке участвуют только
структурированные данные. Современные системы управления базами
данных, такие как Oracle, DB2, MS SQL server, PostgreSQL, реализуют
хранение и обработку именно такого класса данных. Рассматриваемый в
работе класс массовой обработки данных находит широкое применение в
задачах управления организациями и предприятиями, в финансовой сфере и
задачах, относящихся с активно развивающихся в настоящее время областях,
связанных с задачами в области искусственного интеллекта.

Особенность массовой обработки данных состоит в том, что
обрабатываются большие и сверхбольшие объемы данных, причем в
практически все они включаются обработку, что в работе называется
"высокой активностью данных". Это обуславливает высокую
вычислительную сложность решаемых задач. Поэтому заявленная и
достигнутая в диссертационной работе цель: "создание алгебраических
моделей и методов синтеза программно-аппаратных комплексов массовой
обработки высокоактивных данных путем выбора на их основе подходящих
архитектур и способов построения вычислительных процедур",
свидетельствует о ее актуальности.

Структура и содержание диссертации

Диссертация состоит из введения, 6 глав, заключения, списка литературы и приложений. Общий объем диссертации – 286 стр., в том числе 50 рисунков, 25 таблиц, 4 приложения. Список литературы состоит из 244 наименований.

Во введении обоснована актуальность темы диссертации, сформулированы основные задачи и научные результаты, дано краткое содержание глав.

Первая глава посвящена общим вопросам проектирования архитектуры параллельных систем баз данных, описана проблемная область и определены объекты исследования. В начале главы дано определение массовой обработки данных, указано, что в работе рассматривается и исследуется обработка структурированных данных. В качестве основного объекта для формализации принят структурированный файл, состоящий из однотипных записей (структур). Файл рассматривается как некоторое абстрактное понятие, которому в реальности могут соответствовать не только «физические» файлы в операционных системах, но и таблицы в базах данных и более формальные объекты типа матриц. Введено понятие коэффициента активности файла как отношения общего числа записей в файле к числу записей, подвергающихся обработке в процессе выполнения операции. Сделан вывод о том, что в рассматриваемых задачах коэффициенты активности всех обрабатываемых файлов близки к единице. Проведен анализ архитектур известных промышленных программно-аппаратных комплексов с позиции использования их для реализации массовой обработки данных. Определены цели работы, состоящие в разработке моделей данных, обеспечивающих наибольшее соответствие существующим моделям данных и архитектурам вычислительных комплексов. Показано, что упорядоченность обеспечивает ускорение выполнения всех операций над файлами. При использовании параллельных вычислительных комплексов упорядоченность обеспечивает лучшее распределение файлов и их фрагментов по ресурсам комплекса и эффективное использование этих ресурсов. Рассмотрены проблемы оптимизации процессов массовой обработки данных. Предложены теоретико-множественная (файловая), и многомерно-матричная модели данных и сформулированы требования, которым эти модели должны удовлетворять: процедурность; параллелизм алгебраических операций; оптимизация запросов; объектно-ориентированная парадигма.

Вторая глава посвящена рассмотрению методов формализации алгебраических моделей данных и вычислений. В ней приводятся основные определения универсальных многоосновных алгебраических систем и приводятся примеры таких систем, например, обычная алгебра матриц, которая представляет собой очень важный с точки зрения массовой обработки данных пример универсальной двухосновной алгебраической системы. Проведен анализ

современных интуитивных подходов к объектно-ориентированному моделированию, проектированию и программированию и осуществлен переход к формальному – алгебраическому подходу к объектно-ориентированным технологиям. Абстрактный тип данных определен как универсальная многоосновная алгебраическая система. Приведено описание объектно-ориентированной технологии программирования для создания программно-аппаратных комплексов на основе универсальной (абстрактной) алгебраической машины – универсальной двухосновной алгебраической системы вида где одна называется структурой, а вторая типом. Предложена алгебра кортежей, содержащих разнотипные атрибуты. Дано определение файла как фактор-множества множества однотипных записей по отношению эквивалентности, порожденному множеством ключей, и на его основе построены алгебраические определения операций массовой обработки данных. Дано описание аксиоматической теории массовой обработки данных. Показано, что запросы на получение данных есть теоремы в аксиоматических теориях – моделях массовой обработки данных, и на основе теоремы об эквивалентности аксиоматических теорий доказано соответствие этих моделей.

В третьей главе рассмотрены вопросы, связанные с алгебраическими методами оптимизации запросов на решение задач массовой обработки данных посредством преобразования алгебраических выражений запросов и их параллельной реализации. Сделано обобщение алгоритма оптимизации последовательности умножения матриц на случай произведения многомерных матриц. На основе применения принципа инвариантного погружения проведено доказательство того, что для синтеза оптимизированного алгебраического выражения запроса на массовую обработку данных может быть использован метод динамического программирования. Также показано, что особенности этих процессов МОД позволяют эффективно использовать этот метод оптимизации, а полученный процесс может быть улучшен использованием алгоритма ветвей-границ. Сделано обобщение алгоритма Кэннона параллельного умножения матриц на случай произведения многомерных матриц. Предложен способ представления многомерных матриц в виде условно-трехмерных. Доказано, что результат произведения многомерных матриц есть многомерная матрица, составленная из сечений, каждое из которых есть произведение сечений исходных матриц по скоттовым индексам. Проведен анализ алгоритмов, реализующих операцию слияния не строго упорядоченных файлов. Предложен способ параллельной реализации этой операции на основе принципа симметричного горизонтального распределения файлов-операндов. Предложен эвристический алгоритм для этого распределения

данных, который обеспечивает равномерное распределение фрагментов по процессорам, и приведено описание ряда экспериментов, подтвердивших его эффективность. Приведена его реализация языковыми средствами реляционных СУБД. Предложена стратегия повышения эффективности процессов массовой обработки данных.

В четвертой главе рассмотрены этапы построения программно-аппаратных комплексов, реализующих массовую обработку данных. Предложена архитектура программно-аппаратного комплекса для реализации многомерно-матричной модели данных. Подробно рассмотрена архитектура, ориентированная на реализацию операции произведения многомерных матриц на основе сечений по совокупности скоттовых индексов. Приведен пример произведения условно-трехмерных матриц. Предложены архитектуры программно-аппаратных комплексов для реализации простых операций теоретико-множественной модели данных и архитектура программно-аппаратного комплекса для параллельной реализации операции слияния нестрого упорядоченных файлов в теоретико-множественной и реляционной моделях данных на основе принципа симметричного горизонтального распределения. Предложены параллельная реализация последовательности операций слияния нестрого упорядоченных файлов на основе конвейерной архитектуры и параллельная реализация операции слияния нестрого упорядоченных файлов на основе архитектуры вычислительных систем с ассоциативным распределением ресурсов.

В пятой главе приведены результаты экспериментов, проведенных с целью проверки полученных в предыдущих главах результатов. Эксперименты проводились на рабочих станциях, локальных сетях рабочих станций и с использованием облачных технологий. Приведены результаты эксперимента для анализа параллельной реализации операции умножения многомерных матриц. Показано, что с увеличением размерности индексов возрастает эффективность параллельного алгоритма. Анализ операции слияния нестрого упорядоченных файлов показал, что параллельная реализация симметрично горизонтально распределенных файлов (таблиц в реляционной базе данных) дает существенное ускорение. Кроме того, показано, что в этом случае загрузка процессоров максимальна.

В шестой главе рассмотрено применение предложенного подхода к решению практических прикладных и системных задач массовой обработки данных. Рассмотрено использование предложенного метода для решения задач о путях в графе. Показано, что при возведении в степень исходной матрицы графа операцией умножения многомерных матриц, набор значений индексов отличного от

нейтрального элемента очередной степени этой матрицы есть последовательность номеров вершин графа, через которые проходит путь. Показано, что этот метод может быть использован для решения задачи вывода ассоциативных правил. Показано, что возможно и эффективно использование принципа симметричного горизонтально распределения использование для решения задачи поиска изображений в базах данных на основе перцептивного хеширования. Предложен и реализован метод параллельного сравнения ключей изображений с использованием SIMD-регистров процессора. Предложена архитектура программно-аппаратного комплекса для поиска изображений в базах данных. Показана реализация алгоритма шифрования Хилла на основе алгебры многомерных матриц. Показано, что криптостойкость обобщенного алгоритма Хилла повышается за счет свойств многомерных матриц, таких как количество и размерности индексов, выбора скоттовых и кэлиевых индексов, возможности построения различных единичных и, соответственно, обратных матриц.

В заключении сформулированы основные научные и практические результаты, представленные в диссертации.

Приложения содержат свидетельства о регистрации разработанных алгоритмов и программ, патенты и акт внедрения результатов исследований, изложенных в диссертации.

С учетом изложенного можно считать, что научно-техническая новизна диссертационной работы состоит в следующем.

- Предложен новый метод формализации моделей данных и моделей вычислений, основанный на универсальных многоосновных алгебраических системах и объектно-ориентированном подходе к проектированию и разработке программно-аппаратных комплексов для решения задач массовой обработки данных.
- Предложена и разработана теоретико-множественная (файловая) алгебраическая система, которая используется для верификации соответствия известных и используемых на практике моделей данных и моделей вычислений.
- Описана алгебра многомерных матриц и предложен метод – абстрактная алгебраическая машина, который позволяет использовать в качестве элементов многомерных матриц различные, как простые, так и структурные (кортежи), типы данных.
- Разработаны алгебраический и аксиоматический методы подтверждения соответствия (изоморфизма или гомоморфизма) моделей данных и моделей вычислений и осуществлено подтверждение гомоморфизма, а для конкретных задач – изоморфизма, теоретико-множественной, многомерно-матричной и реляционной моделей на основе использования обоих методов

– Разработано обобщение алгоритма выбора последовательности операций умножения матриц методом динамического программирования для произведения многомерных матриц, показано, что синтез оптимизированного процесса массовой обработки данных может быть реализован методом динамического программирования. Приведен пример синтеза такого процесса.

– На основе параллельной реализации операций: выбран и обобщен для параллельной реализации произведения многомерных матриц алгоритм Кэннона; для файловой модели разработан алгоритм операции слияния нестрого упорядоченных файлов на основе симметричного горизонтального распределения файлов-операндов; для реализации симметричного горизонтального распределения разработан оригинальный эвристический алгоритм; показано, что на основе симметричного горизонтального распределения может быть эффективно распараллелена реляционная операция Join.

– Разработаны этапы построения программно-аппаратного комплекса для реализации МОД, архитектура для реализации многомерно-матричной модели данных, архитектуры для реализации простых (однопроходных) операций теоретико-множественной модели данных, архитектура для параллельной реализации операции слияния нестрого упорядоченных файлов в теоретико-множественной и реляционной моделях данных для различных алгоритмов, в том числе с использованием ассоциативных вычислительных систем.

Достоверность результатов и выводов, представленных в диссертации, подтверждается проведенными испытаниями программного обеспечения, реализующего разработанные методы и алгоритмы. Эти результаты неоднократно докладывались на международных научных конференциях. Они опубликованы в российских рецензируемых изданиях, в числе которых 33 в журналах, входящих в Перечень ВАК, 9 работах в рецензируемых научных изданиях, входящих в системы цитирования Web of Science и Scopus, и 1 монографии.

Рекомендации по использованию результатов работы

Результаты исследования, представленные в диссертации, могут заинтересовать специалистов, занимающихся исследованиями и разработкой методов параллельной обработки больших данных. Кроме того, они представляют интерес для научно-исследовательских организаций и предприятий, занимающихся разработкой и промышленной реализацией программно-аппаратных комплексов.

Соответствие диссертации и автореферата требованиям

«Положения о присуждении ученых степеней»

Диссертация В.И. Мунермана обладает внутренним единством. Автореферат

диссертации адекватно отражает ее содержание. Диссертация и автореферат удовлетворяют требованиям «Положения о присуждении ученых степеней».

По тексту диссертации есть следующие замечания:

1. В приложении имеется акт внедрению предложенных в работе методов, однако в тексте отсутствуют описания того, как результаты исследования были использованы в этой разработке.
2. В описаниях экспериментов говорится о том, что они проводились с использованием процессоров Intel, но не приводятся результаты экспериментов на отечественных процессорах.
3. В облачных экспериментах использовано программное обеспечение, в частности СУБД Microsoft SQL Server с облачная система Microsoft Azure. Было целесообразно привести описание экспериментов на СУБД PostgreSQL и отечественных облачных системах.

Заключение

Отмеченные замечания не меняют общего положительного впечатления о диссертационной работе. В ней излагаются результаты решения востребованной практикой, стратегически важной задачи создания алгебраических моделей и методов синтеза программно-аппаратных комплексов массовой обработки высокоактивных данных путем выбора на их основе подходящих архитектур и способов построения вычислительных процедур.

Диссертация Мунермана В.И. обладает внутренним единством, представляет завершенные результаты на актуальную тему и является научно-квалификационной работой, в которой на основании выполненных автором исследований разработаны теоретические положения, вносящие значительный вклад в разработку математического и программного обеспечения вычислительных систем, комплексов и компьютерных сетей. Автореферат диссертации адекватно отражает ее содержание.

Диссертационная работа Мунермана В.И. удовлетворяет требованиям, предъявляемым к диссертациям по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», в пунктах: 3 – модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем; 8 – модели и методы создания программ и программных систем для параллельной и распределенной обработки данных, языки и инструментальные средства параллельного программирования; 9 – модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределенной обработки данных, а ее автор заслуживает присуждения ему

ученой степени доктора технических наук по этой специальности.

Мунерман В.И не был и не является сотрудником АО НПЦ «ЭЛВИС»

Отзыв составлен доктором технических наук, доцентом Е.С. Янаковой

Отзыв о диссертации обсужден и одобрен на заседании НТС от 15.04.2025
(протокол № 7/2025 от 15.04.2025).

Начальник отдела
интегрированных программных решений

Е.С. Янакова
15.04.2025

Контактные данные:

ФИО: Янакова Елена Сергеевна

Ученая степень: доктор технических наук

Специальность: 05.13.01 – Системный анализ, управление и обработка информации.

Ученое звание: доцент

Полное название организации: Акционерное общество Научно-производственный центр «Электронные вычислительно-информационные системы»

Должность: начальник отдела интегрированных программных решений

Почтовый адрес: 124460, г. Москва, г. Зеленоград, ул. Конструктора Лукина, д. 14, стр. 14, этаж 6, ком. 6.23

Контактный телефон: 8 (495) 926-79-57 доб. 2100

Адрес электронной почты: helen@elvees.com