

О Т З Ы В

официального оппонента о диссертационной работе
Мунермана Виктора Иосифовича

“Алгебраические модели и методы для разработки программно-аппаратных комплексов массовой обработки данных”,

представленной на соискание учёной степени

доктора технических наук по специальности 2.3.5

“Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей”

В диссертационной работе В.И. Мунермана рассматриваются задача создания математической основы (в первую очередь – на основе алгебраических систем) и последующей разработке алгоритмической поддержки т.н. массовой обработки данных (МОД).

Актуальность темы. В настоящее время массовую обработку данных связывают с направлением, получившим название Big Data (большие данные), который обозначает вновь создающиеся структурированные, неструктурированные и полуструктурированные данные больших и постоянно возрастающих объёмов. Загрузка неструктурированных и полуструктурированных данных в обычную (например, реляционную) базу данных и последующая обработка требуют слишком больших затрат ресурсов вычислительных комплексов.

Работа В.И. Мунермана посвящена рассмотрению распространенного класса массовой обработки – обработке высокоактивных структурированных данных. Высокая активность данных означает, что при выполнении операций над данными в обработку включается их большая часть, близкая к 100%.

Постановка проблемы. Современные системы высокой доступности должны обладать высокой скоростью реакции на запросы пользователей. К числу таких систем относятся системы, основанные на взаимодействии с базами данных. Способы организации баз данных для таких систем зависят от специфики решаемых ими задач. Для систем, используемых для анализа больших объёмов данных и формирования отчётов в управлении производствами, маркетинге и финансовых технологиях, целесообразно использование массовой обработки данных.

Особенность представления данных в виде многомерного куба порождает их основной недостаток – длительная предварительная подготовка агрегированной информации для последующего к ней обращения, которое реализуется только унарными и тривиальными операциями, такими как транспонирование, сечение и свёртка. Следовательно, любое изменение данных, на основе которых строится многомерный куб, приводит к полной его перестройке.

Цель работы. Разработать теоретические и практические методы создания таких СУБД и машин баз данных, которые могли бы обеспечить включение в обработку результирующего гиперкуба с целью его оперативной корректировки.

Теоретическая часть работы базируется на двух алгебраических моделях МОД. Это: теоретико-множественная (или файловая) и многомерно-матричная модели данных. Теоретическая часть является основой всей диссертационной работы, её главной составной частью.

Среди множества произвольных абстрактных типов данных (АТД) предлагается рассмотреть специфический АТД, называемый в дальнейшем *абстрактной алгебраиче-*

ской машиной. Абстрактная алгебраическая машина (определение 2.11) – это двухосновная алгебраическая система вида $E = \langle S, T, \Omega, \Pi \rangle$. Основа S называется *структурой*, а основа T – *типом*. Структура представляет собой некоторую конструкцию, составленную из экземпляров данного типа. Примеры такого рода структур – векторы, матрицы, графы. Выбор структуры и типа определяется особенностями решаемой задачи. Причём для некоторых классов задач одной структуре могут соответствовать несколько типов.

Особый интерес представляет абстрактная алгебраическая машина, которая имеет следующий вид: $EM = \langle M, X, \Omega, \Pi \rangle$, где X – тип, а M – множество квадратных матриц, составленных из элементов типа. Минимальное требование к типу X состоит в том, чтобы на X были определены две алгебраические операции, одна из которых трактуется как аддитивная, а вторая – как мультипликативная. То есть тип X должен быть по каждой из этих операций, по крайней мере, алгебраической структурой, называемой группоидом. В реальных задачах типами могут быть достаточно сложные алгебраические структуры, такие как кольца и поля.

Практическая ценность универсальных алгебраических машин состоит в том, что суть операций над элементами структуры S не изменяется при изменении сути операций над элементами типа T . Это свойство универсальных алгебраических машин может быть полезным в практическом программировании. Если типы $T_1, \dots, T_n$ – гомоморфные или изоморфные универсальные алгебраические системы, то становится возможной отладка операций над структурой на наиболее простом типе данных. Отлаженная таким образом структура становится базовым АТД, от которого можно порождать конкретные АТД (реализации), предназначенные для решения задач на сложных типах данных.

При таком подходе реальные алгебраические машины, работающие с конкретными типами данных, разрабатываются как АТД – наследники абстрактных алгебраических машин, у которых они наследуют операции над структурами и включают в себя операции над реальными элементами этих структур.

Автором реализованы два эквивалентных метода описания объектов предметной области и операций преобразования одного объекта в другой – аксиоматический и алгебраический.

Предложена теоретико-множественная (файловая) модель массовой обработки данных. Дано определение файла, как фактор-множества множества однотипных записей по отношению эквивалентности, порождённому множеством ключей. На основе этого определения сделаны формальные определения основных операций массовой обработки данных с использованием алгебраического и объектно-ориентированного подхода, которые позволили формализовать как операции над структурами данных высокого уровня – файлами, так и над составляющими их элементами – записями (кортежами).

Предложена алгебраическая модель данных и вычислений на основе созданной Н.П. Соколовым в 1960-1970-х гг. алгебры многомерных матриц. Доказано соответствие этой модели теоретико-множественной и реляционной моделям.

Разработана аксиоматическая теория массовой обработки данных. Показано, что запросы на обработку данных являются теоремами аксиоматической теории. На основе теоремы об эквивалентности аксиоматических теорий доказано, что многомерноматричная и реляционная модели данных соответствуют друг другу.

Научная новизна работы заключается в следующем:

1. Предложен 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.

2. Предложена и разработана теоретико-множественная (файловая) алгебраическая система, которая используется для верификации соответствия известных и используемых на практике моделей данных и моделей вычислений.

3. 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.

4. 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.

5. Разработано обобщение алгоритма выбора последовательности операций умножения матриц методом динамического программирования для (λ, μ) -свернутого произведения многомерных матриц, показано, что синтез оптимизированного процесса МОД может быть реализован методом динамического программирования. Приведён пример синтеза такого процесса.

6. На основе параллельной реализации операций выбран и обобщён для параллельной реализации (λ, μ) -свёрнутого произведения многомерных матриц алгоритм Кэннона; для файловой модели разработан алгоритм операции слияния нестрого упорядоченных файлов.

7. 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.

Теоретическая значимость результатов диссертационной работы состоит в том, что на основе разработанных (файловой и многомерно-матричной) моделей данных, методов доказательства соответствия модели данных и модели вычислений, методов синтеза и оптимизации процессов массовой обработки данных, открывается возможность построения программно-аппаратных комплексов для решения важных задач параллельной реализации массовой обработки данных.

Практическая значимость результатов, полученных в данной работе, состоит в следующем:

1. Предложенные подходы, методы и алгоритмы могут быть использованы для проектирования и разработки программно-аппаратных комплексов МОД на базе широкого спектра многоядерных и многопроцессорных систем.

2. Предложенные алгебраические модели и методы могут быть использованы для разработки технологии коллективного проектирования больших многопроцессорных программно-аппаратных комплексов и для создания сложных программных систем различного назначения.

Замечания.

1. Сложно спорить с тем, что многомерная архитектура может оказаться эффективнее колоночной при правильном дизайне структуры и агрегатов. Но цена этого дизайна часто оказывается слишком высока, что и привело к фактическому вытеснению многомерных баз данных колоночными, дизайн которых на порядок проще.

2. Использование индексов Скотта позволяет значительно повысить производительность выполнения сложных аналитических запросов, снижая нагрузку на систему и сокращая время отклика пользователей. Однако стоит помнить, что предварительное соз-

дание таких индексов требует значительных ресурсов на этапе подготовки данных и обновления структуры при изменении исходных данных. Поэтому решение о применении данной технологии должно приниматься взвешенно, учитывая характер и частоту выполняемых запросов.

3. К замечаниям по работе можно отнести не вполне достаточное измерение эффективности предложенных методов и алгоритмов при решении конкретных практических задач, что могло бы в большей степени обосновать вклад диссертационного исследования в совершенствование технологий МОД.

4. Имеются неточности в описании аксиоматической теории МОД (п. 2.5.2). Например, терм назван формулой. Конечно, то или иное название не играет принципиальной роли, однако, здесь автор противоречит со своим собственным заявлением о том, что термы обозначают объекты (предметы), а формулы – утверждения.

5. Неудобно для читателя организован список литературы: ввиду большого количества публикаций в этом списке его следовало бы упорядочить в алфавитном порядке. Кроме того, некоторые используемые источники упоминаются только в тексте диссертации и отсутствуют в списке литературы (например, Н. Бурбаки, “Теория множеств”).

Однако указанные замечания не снижают оценку значимости как адаптированных, так и разработанных автором алгебраических структур данных, на основе которых открываются перспективы научно обоснованной постановки задач эффективной обработки массовых данных и их дальнейшего решения. Диссертация Мунермана В.И. представляет собой полноценное научное исследование, имеющее, как отмечалось выше, большое теоретическое и практическое значение.

Автореферат диссертации правильно и полно отражает её содержание.

Считаю, что диссертационная работа “Алгебраические модели и методы для разработки программно-аппаратных комплексов массовой обработки данных” удовлетворяет требованиям Положения о порядке присуждения учёных степеней ВАК РФ и отвечает паспорту 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 распределённой обработки данных, языки и инструментальные средства параллельного программирования, 9 – модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределённой обработки данных. Автор работы Мунерман В.И. заслуживает присуждения ему учёной степени доктора технических наук по этой специальности.

Официальный оппонент: доктор физико-математических наук, профессор, специальность 01.01.06 – “Математическая логика, алгебра и теория чисел”, профессор кафедры ВМ-1 Национального исследовательского университета “МИЭТ”

30 апреля 2025 года

Кожухов Игорь Борисович

Подпись Кожухова И.Б. удостоверяю
Учёный секретарь Учёного совета МИЭ

 Козлов А.В.

Сведения об организации: Федеральное
режделение “Национальный исследовательский
124498, г. Москва, Зеленоград, пл. Шокина, д. 1, телефон +7 (499) 731-44-41,
электронная почта: netadm@micee.ru, адрес в сети Интернет: miet.ru

образовательное уч-