

ОТЗЫВ

официального оппонента САЕНКО Игоря Борисовича
на диссертацию ФИЛИМОНОВА Алексея Валерьевича,
выполненную на тему «Математическое и программное обеспечение
процесса приближенной обработки запросов в реляционных системах
управления базами данных»
и представленную на соискание ученой степени кандидата
технических наук по специальности 2.3.5. Математическое и
программное обеспечение вычислительных систем, комплексов и
компьютерных сетей

1. Актуальность темы

В настоящее время аналитическая обработка данных в реальном времени является одной из наиболее упоминаемых технологий в области баз данных. В широком смысле эта технология должна обеспечить оперативный анализ потока данных в реальном масштабе времени. Одним из проблемных вопросов, связанных с аналитической обработкой данных, является ограничение времени ответа на сложный в вычислительном плане запрос. Приближенная обработка запросов представляется как один из возможных подходов, который способен разрешить данное противоречие. Применение механизмов приближенной обработки запросов может привести к возможности получения достаточно быстрого ответа на запрос без необходимости обработки полного набора данных. При этом точность ответа может в полной мере удовлетворить пользователя. Тем самым кратно снижается стоимость выполнения запроса.

Однако несмотря на то, что методы приближенной обработки запросов могут значительно повысить производительность систем управления базами данных (СУБД), они также имеют ограничения и недостатки. Одним из наиболее значимых недостатков является существенная потеря точности при выполнении аналитических запросов на некоторых наборах данных. В этой связи представляется вполне справедливым выбор цели диссертации, которая заключается в разработке математического и программного обеспечения, которые обеспечивают эффективную приближенную обработку конвейера запросов вне зависимости от обрабатываемых данных.

Для этого в диссертационной работе предложено усовершенствовать метод выборки путем учета характера обрабатываемых данных. Данная возможность достигается при помощи использования выявленной в исследовании зависимости частоты выборки данных от коэффициента вариации оцениваемых значений.

На основе усовершенствованного метода выборки разработаны алгоритмы приближенной обработки запросов и конвейера запросов. Представленные алгоритмы применены для синтеза архитектуры и элементов программного обеспечения процессора запросов СУБД с функцией приближенной обработки конвейера запросов.

Таким образом, диссертация Филимонова А.В., выполненная на тему «Математическое и программное обеспечение процесса приближенной обработки запросов в реляционных системах управления базами данных», сущностью которой является разработка средств математического и специального программного обеспечения процесса приближенной обработки конвейера запросов, находится в рамках известного научного направления развития современных систем и механизмов управления базами данных, что подтверждает ее высокую актуальность.

2. Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Представленные автором положения, выносимые на защиту, выводы по результатам анализа математических методов и экспериментального оценивания, а также рекомендации по теоретическому и практическому применению полученных в диссертации результатов являются в достаточной степени обоснованными и достоверными. Их обоснованность и достоверность обеспечивается:

- тщательностью проведенного анализа литературы, включающего в себя работы в отечественных и международных рецензируемых научных изданиях;
- адекватностью используемых методов теоретического и экспериментального исследования и сформулированных постановок задач;
- результатами компьютерных экспериментов;
- корректным применением известных методов исследования;
- непротиворечивостью исходных данных и воспроизводимостью результатов, полученных в процессе экспериментов.

3. Научная новизна результатов

Научная новизна результатов, полученных автором в диссертации, обусловлена следующим:

- математическое обеспечение приближенной обработки конвейера запросов в реляционных системах управления базами данных отличается тем, что в нем применяются коэффициент вариации как метрика качества выборки данных, что обеспечивает получение равноточных оценок случайных величин из различных кластеров;

- алгоритм приближенной обработки конвейера запросов отличается от известных аналогов тем, что он имеет двухпроходную процедуру расчета требуемых статистических характеристик в цепочках запросов, что обеспечивает получение результатов выполнения конвейера запросов с оптимальным распределением пространства выборки;

- архитектура процессора запросов реляционной СУБД отличается наличием механизмов кэширования планов и результатов выполнения промежуточных запросов из конвейера, тем самым обеспечивая снижение ресурсоемкости повторной обработки частных запросов в других конвейерах;

- структура программного обеспечения приближенной обработки конвейера запросов, в отличие от известных, реализует взаимодействие с модулем сбора статистики и механизмом кэширования СУБД, что обеспечивает повышение точности результатов выполненных запросов в заданных временных ограничениях.

Результаты исследования реализованы в виде элементов специального программного обеспечения для прототипа хранилища данных реляционного типа и применены для решения задачи совершенствования программного обеспечения интерактивной аналитической обработки распределенных данных.

Результаты исследования отражены в 9 публикациях, из которых 3 статьи опубликованы в журналах, рекомендованных ВАК РФ, 1 статья – в издании, индексируемом в международной базе цитирования Scopus, а также в публикациях в других научных изданиях и материалах международных и всероссийских конференций различного уровня. На элементы разработанного специального программного обеспечения получено 2 свидетельства о государственной регистрации программы для ЭВМ, выданный Федеральным институтом промышленной собственности.

4. Теоретическая и практическая значимость исследования

Теоретическая и практическая значимость исследования заключается в разработке средств математического и специального программного обеспечения процесса приближенной обработки запросов в реляционных СУБД, обеспечивающего оптимизацию пространства выборки конвейера запросов. Высокая теоретическая и практическая значимость исследования также обусловлена доведением разработанных алгоритмов до уровня программной реализации и их исполнением в виде элементов специального программного обеспечения с возможностью применения как при проектировании новых СУБД, так и при разработке программных инструментов приближенной обработки запросов в существующих системах.

Кроме того, подтверждением высокой теоретической и практической значимости исследования является реализация полученных результатов в научных исследованиях Института системного программирования РАН, проводимых по тематике совершенствования программного обеспечения распределенных СУБД, а также в образовательном процессе Академии ФСО России в дисциплине «Базы данных», о чем свидетельствуют соответствующие акты.

5. Замечания по диссертационной работе

В качестве замечаний по работе можно отметить следующее:

1. Разработанное в диссертации математическое обеспечение применимо к запросам, в которых оцениваемые величины имеют только метрические типы данных. Запросы к данным других типов входят в ограничения работы, что несколько сужает область применения полученных результатов.

2. В представленном математическом обеспечении недостаточно полно обоснована выбор евклидовой нормы и норма Чебышева в качестве метрик качества выборки данных.

3. В диссертации не в полной мере представлены результаты оценки сходимости алгоритма приближенной обработки конвейера запросов при выполнении многотабличных и вложенных запросов.

4. Из приведенного в работе описания архитектуры программного обеспечения процессора СУБД с функцией приближенной обработки конвейера запросов не до конца ясна роль и место механизма кеширования промежуточных запросов.

При этом следует отметить, что отмеченные недостатки носят рекомендательный характер и в целом не влияют на общую высокую оценку качества полученных теоретических и практических результатов диссертационного исследования.

6. Заключение

Диссертация Филимонова Алексея Валерьевича выполнена на актуальную тему, обладает внутренним единством, содержит новые научные результаты и является законченным научно-исследовательским трудом, в котором изложены подходы к разработке математического и специального программного обеспечения процесса приближенной обработки запросов, что соответствует п. 3 «Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем»; п. 4 «Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов» паспорта

специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Полученные автором результаты имеют достаточную степень обоснованности и достоверности. Материал диссертации логически взаимоувязан, в его рамках по каждому из положений сделаны корректные и логически верные выводы. Автореферат соответствует основному содержанию диссертации. В целом диссертационная работа отвечает критериям Положения о присуждении ученых степеней, а ее автор, Филимонов Алексей Валерьевич, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Официальный оппонент:

ведущий научный сотрудник лаборатории проблем компьютерной безопасности СПИИРАН Санкт-Петербургского Федерального исследовательского центра Российской академии наук
доктор технических наук, профессор

Игорь Борисович Саенко

22 марта 2023 г.

Почтовый адрес: Россия, 199178, Санкт-Петербург, 14-я линия Васильевского острова, д. 39
Телефон: 8-812-328-34-11
E-mail: ibsaen@comsec.spb.ru

Подг
Наче

зверяю

окарев