

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора технических наук, профессора Кузнецова Сергея Дмитриевича **на диссертационную работу ЦЫМБЛЕРА Михаила Леонидовича «Интеллектуальный анализ данных в СУБД»**, представленную на соискание ученой степени доктора физико-математических наук по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Актуальность диссертационного исследования М.Л. Цымблера определяется тем, что в настоящее время перспективным является направление, связанное с хранением, параллельной обработкой и интеллектуальным анализом сверхбольших баз данных. Последние исследования показывают, что для решения указанных задач могут эффективно использоваться вычислительные кластеры с узлами, оснащенными многоядерными ускорителями. В данной области, однако, остается ряд крупных нерешенных научных проблем, в том числе связанных с задачей разработки эффективных и экономичных параллельных СУБД. Внедрение интеллектуального анализа данных в параллельные СУБД, рассматриваемое в диссертации, является в настоящее время одним из актуальных направлений научных исследований, поскольку такой подход исключает значительные накладные расходы на экспорт-импорт данных и результатов. Разработка аналитических алгоритмов, внедряемых в СУБД, для платформы многоядерных ускорителей, также представляет собой актуальное направление исследований, поскольку на сегодня многоядерные ускорители (GPU, Intel Xeon Phi и др.) значительно опережают традиционные процессоры в производительности по арифметическим операциям и, как показывают недавние исследования, эффективны для решения задач интеллектуального анализа данных.

Целью диссертационного исследования М.Л. Цымблера является разработка комплекса параллельных алгоритмов, составляющих программную платформу интеллектуального анализа данных средствами

СУБД с открытым кодом. Для выполнения данной цели М.Л. Цымблером были решены следующие **задачи**:

1. Разработаны методы и алгоритмы для распараллеливания СУБД с открытым кодом и выполнена проверка данных решений на СУБД PostgreSQL.
2. Разработаны методы и алгоритмы, обеспечивающие проведение интеллектуального анализа данных в рамках параллельной СУБД на современных вычислительных кластерах с многоядерными ускорителями.
3. Разработан комплекс параллельных алгоритмов для решения задач интеллектуального анализа данных в параллельной реляционной СУБД:
 - алгоритм кластеризации графов социальных сетей для параллельной СУБД;
 - алгоритм нечеткой кластеризации данных для параллельной СУБД;
 - параллельный алгоритм кластеризации данных с шумами и выбросами для многоядерного ускорителя;
 - параллельный алгоритм поиска похожих подпоследовательностей временного ряда для кластерной вычислительной системы с многоядерными ускорителями;
 - параллельный алгоритм поиска аномальных подпоследовательностей временного ряда для многоядерного ускорителя;
 - параллельный алгоритм поиска частых наборов для многоядерного ускорителя.
4. Проведены вычислительные эксперименты, в которых предложенные методы и алгоритмы продемонстрировали свою высокую эффективность и преимущество над ведущими аналогами.

Текст диссертации аккуратно оформлен в соответствии с требованиями Минобрнауки РФ, в изложении используются строгий язык и стиль научных публикаций. **Текст автореферата** достаточно полно отражает содержание диссертации.

Достоверность полученных результатов подтверждается вычислительными экспериментами с реальными и синтетическими данными, которые проведены согласно общепринятым стандартам, и выполнено сравнение полученных результатов с существующими подходами.

Научная новизна диссертационного исследования заключается в следующем:

1. Предложен новый метод внедрения параллелизма в последовательную СУБД с открытым исходным кодом, подразумевающий сравнительно небольшие по объему изменения в исходном коде. Разработаны новые параллельные алгоритмы интеллектуального анализа данных для параллельной СУБД.
2. Предложен оригинальный метод выполнения интеллектуального анализа данных внутри реляционной СУБД, инкапсулирующий параллельные алгоритмы анализа данных на многоядерных ускорителях.
3. Разработан ряд новых параллельных алгоритмов для интеллектуального анализа данных на многоядерном ускорителе: поиск частых наборов, кластеризация данных, поиск похожих подпоследовательностей и поиск аномальных подпоследовательностей во временных рядах.

Теоретическая ценность диссертационного исследования заключена в предложенных соискателем методах интеграции параллельной обработки и анализа данных в СУБД, а также в параллельных алгоритмах анализа данных для многоядерных ускорителей, показывающих близкое к линейному ускорение.

Практическая ценность диссертационного исследования заключается в том на основе предложенных методов и подходов разработан прототип параллельной СУБД для вычислительных кластеров, в которую внедрены разработанные параллельные алгоритмы для многоядерных ускорителей. Результаты исследования могут быть применены в разработке системного

программного обеспечения для параллельной обработки и анализа больших объемов реляционных данных.

В качестве **замечаний** к диссертационной работе можно отметить следующие моменты.

1. Поскольку существенной частью работы является разработка СУБД с фрагментным параллелизмом, стоило бы включить в диссертацию более широкий обзор родственных работ, тем более, что уже неоднократно СУБД PostgreSQL использовалась в качестве основы таких проектов.
2. В диссертации достаточно большое внимание уделяется аналитическим аспектам задачи разрезки графов (graph partitioning). Однако та же задача очень актуальна и для СУБД с фрагментным параллелизмом, в которых ее решение может позволить для заданной нагрузки выбрать способ фрагментирования базы данных, сокращающий число обменов между узлами кластера при выполнении запросов. На эту сторону задачи следовало бы обратить внимание.
3. В работе много говорится про интеграцию аналитики в СУБД. Стоило бы продолжить эту тему (или хотя бы обозначить ее) и в связи с СУБД с фрагментным параллелизмом. Интересной задачей является распараллеливание аналитических программ внутри того же кластера, на котором работает СУБД.

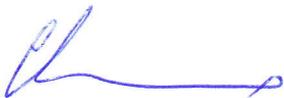
Указанные замечания не снижают общей значимости результатов диссертационного исследования, выполненного М.Л. Цымблером.

Публикации и апробации. Соискателем опубликованы 12 статей в журналах Перечня ВАК и 9 статей в изданиях библиографических баз Scopus и Web of Science, которые с достаточной полнотой отражают основные результаты диссертации. Результаты исследования прошли апробацию на 17 международных и всероссийских научных конференциях.

ЗАКЛЮЧЕНИЕ. Диссертационная работа М.Л. Цымблера представляет собой самостоятельно выполненную и законченную научно-

квалификационную работу, в которой автором предложен новый метод внедрения фрагментного параллелизма в СУБД с открытым кодом и разработан ряд новых параллельных алгоритмов интеллектуального анализа данных в СУБД на вычислительных кластерах с многоядерными ускорителями. Совокупность полученных М.Л. Цымблером результатов можно квалифицировать как научное достижение, и его диссертационная работа соответствует требованиям Положения о порядке присуждения ученых степеней, включая п. 9. М.Л. Цымблер заслуживает присуждения ученой степени доктора физико-математических наук по специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент:



Сергей Дмитриевич Кузнецов

«13» сентября 2020 г.

Доктор техн. наук, профессор,
главный научный сотрудник Федерального государственного бюджетного учреждения науки «Институт системного программирования имени В.П. Иванникова РАН», г. Москва
Адрес организации: 109004, Москва, ул. Александра Солженицына, дом 25
Телефон: +7 (495) 912-15-24
Email: kuzloc@ispras.ru

Подпись С.Д. Кузнецова удостоверяю:

Ученый секретарь ИСП РАН,
Кандидат физико-математических наук



О.И. Самоваров

СВЕДЕНИЯ ОБ ОФИЦИАЛЬНОМ ОППОНЕНТЕ

диссертации Цымблера М.Л. «Интеллектуальный анализ данных в СУБД», представленной на соискание ученой степени доктора физико-математических наук по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

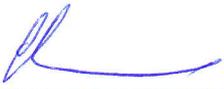
Фамилия, имя, отчество	Кузнецов Сергей Дмитриевич
Ученая степень (с указанием номера и цифра специальности)	Доктор технических наук, 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей
Ученое звание	Профессор
Организация основного места работы	Федеральное государственное бюджетное учреждение науки «Институт системного программирования имени В.П. Иванникова РАН»
Занимаемая должность	Главный научный сотрудник
Почтовый адрес	109004, Москва, ул. Александра Солженицына, дом 25
Телефон	+7 (495) 912-15-24
Адрес электронной почты	kuzloc@ispras.ru

Список основных публикаций по теме диссертации в рецензируемых научных изданиях

1. Y. R. Nedumov and S. D. Kuznetsov. Exploratory Search for Scientific Articles // Programming and Computer Software, 2019, vol. 45, issue 7, pp. pp 405–416.
2. Sergey D. Kuznetsov. Towards a Native Architecture of in-NVM DBMS // In Proc. of the VI International Conference on Actual Problems of System and Software Engineering (invited papers), IEEE Xplore, 2019, pp. 77 – 89.
3. Кузнецов С.Д. Новые устройства хранения данных и их влияние на технологию баз данных // Программная инженерия. 2018. Т. 9. № 4. С. 147-155.
4. Кузнецов С.Д. Управление данными: 25 лет прогнозов // Труды Института системного программирования РАН. 2017. Т. 29. № 2. С. 117-160.
5. Mikhail Drobyshevskiy, Denis Turdakov, Sergey Kuznetsov. Reproducing Network Structure: A Comparative Study of Random Graph Generators // In Proc. of the Ivannikov ISPRAS Open Conference, 2017, pp. 83-89.
6. С.Д. Кузнецов, А.Г. Гомзин. Метод автоматического определения

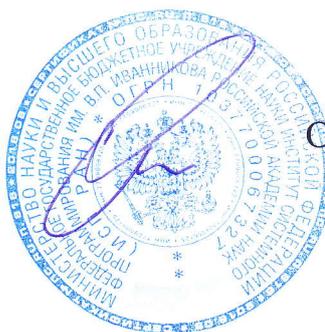
возраста пользователей с помощью социальных связей // Труды Института системного программирования РАН, 2016 г., том 28, выпуск 6, стр. 111-120.

7. Кузнецов С.Д. Объектные модели ODMG и SQL десять лет спустя: нет противоречий // Труды Института системного программирования РАН. 2015. Т. 27. № 1. С. 173-192.
8. Кузнецов С.Д., Посконин А.В. Системы управления данными категории NoSQL // Программирование. 2014. № 6. С. 34-47.

 / Кузнецов С.Д. /

Подпись С.Д. Кузнецова удостоверяю:

Ученый секретарь ИСП РАН,
Кандидат физико-математических наук,



О.И. Самоваров