

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА
Д 212.298.18, СОЗДАННОГО НА БАЗЕ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО АВТОНОМНОГО
ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО
ОБРАЗОВАНИЯ «ЮЖНО-УРАЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»
МИНИСТЕРСТВА НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ ПО ДИССЕРТАЦИИ
НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ ДОКТОРА НАУК

аттестационное дело № _____

решение диссертационного совета от 12 февраля 2020 г. № 2

О присуждении Цымблеру Михаилу Леонидовичу, гражданину РФ, ученой степени доктора физико-математических наук.

Диссертация «Интеллектуальный анализ данных в СУБД» по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей – принята к защите 30 октября 2019 г., протокол заседания № 2/п, диссертационным советом Д 212.298.18, созданным на базе Федерального государственного автономного образовательного учреждения высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)», Министерство науки и высшего образования РФ, 454080, г. Челябинск, пр. им. В.И. Ленина, 76, № 105/нк от 11 апреля 2012 г.

Соискатель Цымблер Михаил Леонидович 1971 года рождения, диссертацию на соискание ученой степени кандидата физико-математических наук на тему «Методы построения программного комплекса для управления данными в вычислительных системах с массовым параллелизмом» защитил в 2000 г. в диссертационном совете Д 064.19.03, созданном на базе Челябинского государственного университета. В 2012-2015 гг. обучался в докторантуре Федерального государственного бюджетного образовательного учреждения высшего образования «Южно-Уральский государственный университет» (национальный исследовательский университет). В 2002 г. присвоено ученое звание доцента по кафедре математического обеспечения ЭВМ. С 2004 г. соискатель работает в Южно-Уральском государственном

университете и в настоящее время является доцентом кафедры системного программирования и начальником отдела интеллектуального анализа данных и виртуализации Лаборатории суперкомпьютерного моделирования (по совместительству) Федерального государственного автономного образовательного учреждения высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)», Министерство науки и высшего образования Российской Федерации.

Диссертация выполнена на кафедре системного программирования Высшей школы электроники и компьютерных наук Федерального государственного автономного образовательного учреждения высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)», Министерство науки и высшего образования Российской Федерации.

Научный консультант – доктор физико-математических наук, профессор Соколинский Леонид Борисович, проректор по информатизации, зав. кафедрой системного программирования Высшей школы электроники и компьютерных наук Федерального государственного автономного образовательного учреждения высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)».

Официальные оппоненты:

- Воеводин Владимир Валентинович, доктор физико-математических наук, профессор, член-корреспондент РАН, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В. Ломоносова», Москва, директор Научно-исследовательского вычислительного центра МГУ им. М.В. Ломоносова;
- Кузнецов Сергей Дмитриевич, доктор технических наук, профессор, Федеральное государственное бюджетное учреждение науки «Институт системного программирования имени В.П. Иванникова РАН», Москва, главный научный сотрудник;
- Елизаров Александр Михайлович, доктор физико-математических наук, профессор, Федеральное государственное автономное образовательное

учреждение высшего образования «Казанский (Приволжский) федеральный университет», Казань, профессор кафедры программной инженерии Высшей школы информационных технологий и интеллектуальных систем

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО», Санкт-Петербург, в своем положительном отзыве, подписанном Бухановским Александром Валерьевичем, доктором технических наук, профессором, директором Научно-исследовательского института наукоемких компьютерных технологий Университета ИТМО, указала, что диссертационная работа М.Л. Цымблера является законченной научно-квалификационной работой, в которой разработан комплекс новых параллельных алгоритмов для интеллектуального анализа данных в рамках параллельной СУБД на платформе кластерных вычислительных систем с многоядерными ускорителями и предложен метод инкапсуляции параллелизма в последовательные свободные СУБД, что в совокупности можно квалифицировать как научное достижение. Работа в полной мере отвечает требованиям Положения о порядке присуждения ученых степеней (в т.ч. п. 9), а ее автор заслуживает присуждения ученой степени доктора физико-математических наук по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Соискатель имеет 82 опубликованных научных работы, в том числе по теме диссертации опубликована 21 работа, из них в рецензируемых научных изданиях опубликовано 12 работ, также имеется 3 свидетельства о государственной регистрации программы для ЭВМ:

Статьи в журналах из Перечня рецензируемых научных изданий ВАК

1. Пан К.С., Цымблер М.Л. Внедрение фрагментного параллелизма в СУБД с открытым кодом // Программирование. 2015. Т. 41, № 5. С. 18–32.
2. Пан К.С., Соколинский Л.Б., Цымблер М.Л. Интеграция параллелизма в СУБД с открытым кодом // Открытые системы. СУБД. 2013. № 9. С. 56–58.

3. Пан К.С., Цымблер М.Л. Разработка параллельной СУБД на основе последовательной СУБД PostgreSQL с открытым исходным кодом // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. 2012. № 18(277). Вып. 12. С. 112–120.
4. Мовчан А.В., Цымблер М.Л. Обнаружение подпоследовательностей во временных рядах // Открытые системы. СУБД. 2015. № 2. С. 42–43.
5. Миниахметов Р.М., Цымблер М.Л. Интеграция алгоритма кластеризации Fuzzy c-Means в PostgreSQL // Вычислительные методы и программирование: Новые вычислительные технологии. 2012. Т. 13. С. 46–52.
6. Пан К.С., Цымблер М.Л. Параллельный алгоритм решения задачи анализа рыночной корзины на процессорах Cell // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. 2010. № 16(192). Вып. 5. С. 48–57.
7. Речкалов Т.В., Цымблер М.Л. Параллельный алгоритм вычисления матрицы Евклидовых расстояний для многоядерного процессора Intel Xeon Phi Knights Landing // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2018. Т. 7, № 3. С. 65–82.
8. Краева Я.А., Цымблер М.Л. Совместное использование технологий MPI и OpenMP для параллельного поиска похожих подпоследовательностей в сверхбольших временных рядах на вычислительном кластере с узлами на базе многоядерных процессоров Intel Xeon Phi Knights Landing // Вычислительные методы и программирование: Новые вычислительные технологии. 2019. Т. 20, № 1. С. 29–43.
9. Цымблер М.Л. Параллельный алгоритм поиска диссонансов временного ряда для многоядерных ускорителей // Вычислительные методы и программирование: Новые вычислительные технологии. 2019. Т. 20, № 3. С. 211–223.
10. Речкалов Т.В., Цымблер М.Л. Параллельный алгоритм кластеризации данных для многоядерных ускорителей Intel MIC // Вычислительные методы и программирование: Новые вычислительные технологии. 2019. Т. 20, № 2. С. 104–115.
11. Цымблер М.Л. Параллельный поиск частых наборов на многоядерных ускорителях Intel MIC // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2019. Т. 8, № 1. С. 54–70.

12. Цымблер М.Л. Обзор методов интеграции интеллектуального анализа данных в СУБД // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2019. Т. 8, № 2. С. 32–62.

Статьи в изданиях, индексируемых в Scopus и Web of Science

13. Zymbler M. Parallel Algorithm for Frequent Itemset Mining on Intel Many-core Systems // Journal of Computing and Information Technology. 2018. Vol. 26, No. 4. P. 209–221.
14. Zymbler M. Best-match Time Series Subsequence Search on the Intel Many Integrated Core Architecture // Proceedings of the 19th East-European Conference on Advances in Databases and Information Systems, ADBIS 2015 (September 8–11, 2015, Poitiers, France). Lecture Notes in Computer Science. Springer, 2015. Vol. 9282. P. 275–286.
15. Kraeva Ya., Zymbler M. Scalable Algorithm for Subsequence Similarity Search in Very Large Time Series Data on Cluster of Phi KNL // 20th International Conference on Data Analytics and Management in Data Intensive Domains, DAMDID/RCDL 2018, Moscow, Russia, October 9–12, 2018, Revised Selected Papers. Communications in Computer and Information Science. 2019. Vol. 1003. P. 149–164.
16. Movchan A., Zymbler M. Parallel Algorithm for Local-best-match Time Series Subsequence Similarity Search on the Intel MIC Architecture // Procedia Computer Science. 2015. Vol. 66. P. 63–72.
17. Rechkalov T., Zymbler M. Accelerating Medoids-based Clustering with the Intel Many Integrated Core Architecture // Proceedings of the 9th International Conference on Application of Information and Communication Technologies, AICT'2015 (October 14–16, 2015, Rostov-on-Don, Russia). IEEE, 2015. P. 413–417.
18. Rechkalov T., Zymbler M. Integrating DBMS and Parallel Data Mining Algorithms for Modern Many-Core Processors // Revised Selected Papers of the 19th International Conference on Data Analytics and Management in Data Intensive Domains, DAMDID/RCDL 2017 (October 10–13, 2017, Moscow, Russia). Communications in Computer and Information Science. Springer, 2018. Vol. 822. P. 230–245.
19. Pan C., Zymbler M. Taming Elephants, or How to Embed Parallelism into PostgreSQL // Proceedings of the 24th International Conference on Database and Expert Systems Applications, DEXA 2013 (August 26–29, 2013, Prague,

- Czech Republic). Lecture Notes in Computer Science. Springer, 2013. Vol. 8055. Part I. P. 153–164.
20. Pan C., Zymbler M. Very Large Graph Partitioning by Means of Parallel DBMS // Proceedings of the 17th East-European Conference on Advances in Databases and Information Systems, ADBIS 2013 (September 1–4, 2013, Genoa, Italy). Lecture Notes in Computer Science. Springer, 2013. Vol. 8133. P. 388–399.
 21. Rechkalov T., Zymbler M. A Study of Euclidean Distance Matrix Computation on Intel Many-core Processors // Revised Selected Papers of the 12th International Conference, PCT 2018 (April 2–6, 2018, Rostov-on-Don, Russia). Communications in Computer and Information Science. Springer, 2018. Vol. 910. P. 200–215.

Свидетельства о регистрации программы

22. Мовчан А.В., Цымблер М.Л. Свидетельство Роспатента о государственной регистрации программы для ЭВМ «Программный комплекс для поиска похожих подпоследовательностей временного ряда на многоядерном сопроцессоре Intel Xeon Phi» № 2015618537 от 11.08.2015.
23. Соколинский Л.Б., Цымблер М.Л., Пан К.С., Медведев А.А. Свидетельство Роспатента о государственной регистрации программы для ЭВМ «Параллельная СУБД PargreSQL» № 2012614599 от 23.05.2012.
24. Цымблер М.Л., Пан К.С. Свидетельство Роспатента о государственной регистрации программы для ЭВМ «Программный комплекс для решения задачи анализа рыночной корзины на многоядерных процессорах с поддержкой векторных вычислений» № 2011610732 от 11.01.2011.

Все положения, вынесенные на защиту, получены М.Л. Цымблером единолично. Все совместные публикации выполнены М.Л. Цымблером в соавторстве с аспирантами и студентами, у которых он являлся научным руководителем (за исключением статьи [2]). В статье [2], написанной совместно с Л.Б. Соколинским и К.С. Паном, являвшимся аспирантом М.Л. Цымблера, соискателю принадлежит описание архитектуры и методов реализации СУБД PargreSQL (стр. 57), Л.Б. Соколинскому принадлежит введение (стр. 56), К.С. Пану принадлежит описание тестов СУБД PargreSQL (стр. 58).

На диссертацию и автореферат поступили положительные отзывы от нижеперечисленных ученых.

1. Губайдуллин И.М., доктор физ.-мат. наук, доцент, Институт нефтехимии и катализа – обособленное структурное подразделение Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра РАН, Уфа, старший научный сотрудник лаборатории математической химии. Замечаний в отзыве нет.
2. Легалов А.И., доктор техн. наук, профессор, Федеральное государственное автономное образовательное учреждение высшего образования «Сибирский федеральный университет», Красноярск, профессор кафедры вычислительной техники, руководитель научно-учебной лаборатории технологий программирования Института космических и информационных технологий. Замечание: отсутствие в научных результатах констатации качественных положительных отличий от уже существующих решений.
3. Райхлин В.А., доктор физ.-мат. наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего образования «Казанский национальный исследовательский технический университет имени А.Н. Туполева», Казань, профессор. Замечания: 1) логично было бы выполнять разработки всех алгоритмов для суперкомпьютера «Торнадо ЮУрГУ»; 2) в автореферате не указаны алгоритмы-аналоги, опережаемые разработками автора; 3) в автореферате было бы полезно акцентировать рассматриваемые в работе приложения Data Mining.
4. Глинский Б.М., доктор техн. наук, Федеральное государственное бюджетное учреждение науки «Институт вычислительной математики и математической геофизики Сибирского отделения РАН», Новосибирск, главный научный сотрудник. Замечания: 1) в автореферате для иллюстрации масштабируемости алгоритмов достаточно было бы привести только один из двух графиков: ускорение или эффективность; 2) в автореферате не указано, что в исследовании использованы методы теории вероятностей и математической статистики.
5. Штейнберг Б.Я., доктор техн. наук, старший научный сотрудник, Федеральное государственное автономное учреждение высшего образования «Южный федеральный университет», заведующий кафедрой

алгебры и дискретной математики Института математики, механики и компьютерных наук, Ростов-на-Дону. Замечания: 1) не приводятся оценки сложности разработанных алгоритмов как функции обращений к внешней памяти от объема используемых данных; 2) не приводятся влияющие на быстродействие характеристики алгоритмов и вычислительных систем, на которых производились вычислительные эксперименты.

Выбор официальных оппонентов и ведущей организации обосновывается широкой известностью их научных достижений в областях высокопроизводительных вычислений, параллельного программирования и интеллектуальной обработки больших объемов данных. В ведущей организации действует научная школа в области суперкомпьютерного моделирования и анализа сложных систем.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- **разработан** комплекс новых параллельных алгоритмов интеллектуального анализа данных в параллельной СУБД для вычислительных кластеров с многоядерными ускорителями, позволяющих существенно сократить затраты на экспорт-импорт данных;
- **предложен** новый подход к распараллеливанию последовательных СУБД с открытым исходным кодом, обеспечивающий существенно более экономичное получение параллельной СУБД, чем разработка «с нуля» при сравнимой производительности;
- **введен** новый метод внедрения алгоритмов интеллектуального анализа данных в реляционные СУБД, ориентированный на многопроцессорные системы с многоядерными ускорителями.

Теоретическая значимость исследования обоснована тем, что в диссертационной работе

- **изложена** концепция интеграции параллельных алгоритмов интеллектуального анализа данных в реляционную СУБД;
- **раскрыт** метод инкапсуляции фрагментного параллелизма в последовательные реляционные СУБД с открытым исходным кодом;

- **изучены** генезис и эволюция методов обработки запросов в системах баз данных и их интеграции с методами интеллектуального анализа данных;
- **разработаны новые эффективные параллельные** алгоритмы для решения задач интеллектуального анализа данных (кластеризация, поиск частых наборов, поиск похожих и аномальных подпоследовательностей временного ряда), ориентированные на современные многопроцессорные системы с многоядерными ускорителями.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

- **определена** эффективность разработанных алгоритмов в сравнении с существующими аналогами путем проведения вычислительных экспериментов на реальных и синтетических наборах данных;
- **создан** прототип отечественной параллельной СУБД PargreSQL для обработки запросов к сверхбольшим базам данных;
- **разработана и внедрена** в параллельную СУБД PargreSQL библиотека функций для интеллектуального анализа данных;
- **представлены** методические рекомендации по созданию программной платформы интеллектуального анализа данных средствами СУБД с открытым исходным кодом.

Оценка достоверности результатов исследования выявила:

- **результаты получены** на современных кластерных вычислительных системах с многоядерными ускорителями;
- **экспериментально подтверждена** более высокая эффективность разработанных соискателем алгоритмов по сравнению с лучшими алгоритмами-аналогами, опубликованными в авторитетных международных научных изданиях;
- **использованы** современные методики проведения вычислительных экспериментов и анализа их результатов, обеспечивающие воспроизводимость результатов.

Личный вклад соискателя состоит в том, что:

- **соискателем единолично разработаны** комплекс новых параллельных алгоритмов для интеллектуального анализа данных в рамках параллельной СУБД, метод инкапсуляции параллелизма в последовательные СУБД с открытым исходным кодом и подход к

внедрению параллельных алгоритмов интеллектуального анализа данных в реляционные СУБД.

На заседании 12 февраля 2020 г. диссертационный совет принял решение присудить Цымблеру М.Л. ученую степень доктора физико-математических наук.

При проведении тайного голосования диссертационный совет в количестве 14 человек, из них 7 докторов наук по профилю диссертации, участвовавших в заседании, из 19 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за – 14, против – нет, недействительных бюллетеней – нет.

Зам. председателя
диссертационного совета

И.о. ученого секретаря
диссертационного совета

12 февраля 2020 г.



Танана В.П.

Мокеев В.В.