

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 720.001.04
НА БАЗЕ ОБЪЕДИНЕННОГО ИНСТИТУТА ЯДЕРНЫХ ИССЛЕДОВАНИЙ
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ ДОКТОРА НАУК

аттестационное дело № _____

Решение диссертационного совета от 16 марта 2018 г. № 70

О присуждении Климентову Алексею Анатольевичу, Россия,
учёной степени доктора физико-математических наук.

Диссертация «Методы обработки сверхбольших объёмов данных в распределенной гетерогенной компьютерной среде для приложений в ядерной физике и физике высоких энергий» по специальности «05.13.11 - математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» принята к защите 27 октября 2017 г., протокол № 69, диссертационным советом Д 720.001.04, созданным 11 апреля 2012 г., приказ 105/нк, на базе Объединённого института ядерных исследований, 141980, Московская область, г. Дубна, ул. Жолио-Кюри, 6. Соискатель Климентов Алексей Анатольевич 1959 года рождения.

Диссертацию на соискание учёной степени кандидата физико-математических наук «Создание комплекса автоматизированных стендов для проведения тестовых испытаний при производстве, сборке и запуске адронного калориметра установки ЛЗ на ускорителе ЛЕП», специальность 01.04.01 – техника физического эксперимента, физика приборов, автоматизация физических исследований, соискатель защитил в 1991 г. в диссертационном совете ИТЭФ, г. Москва.

Соискатель работает начальником лаборатории технологий больших данных для экспериментов класса мегасайенс Курчатовского комплекса НБИКС-технологий НИЦ КИ.

Диссертация выполнена в отделе информационных технологий математического моделирования Курчатовского комплекса НБИКС-технологий НИЦ КИ.

Официальные оппоненты:

1. Аветисян Арутюн Ишханович, доктор физико-математических наук, профессор, член-корреспондент РАН, директор Института системного программирования РАН (г. Москва);
2. Воеводин Владимир Валентинович, доктор физико-математических наук, профессор, член-корреспондент РАН, заместитель директора НИВЦ МГУ, заведующий кафедрой суперкомпьютеров и квантовой информатики Московского государственного университета им. М.В. Ломоносова (г. Москва);
3. Оныкий Борис Николаевич, доктор технических наук, профессор, заведующий кафедрой анализа конкурентных систем НИЯУ МИФИ (г. Москва), дали положительные отзывы на диссертацию.

Все отзывы официальных оппонентов положительные, в них отмечается актуальность диссертационного исследования, научная новизна, адекватность использованных современных методов решаемым задачам, достоверность полученных результатов, их теоретическая и практическая значимость.

В отзыве Оныкия Б.Н. критических замечаний нет.

В отзыве Аветисяна А.И. присутствуют два замечания:

1. Не хватает сравнения отдельных компонентов разработанной системы по хранению данных и управлению заданиями с известными открытыми решениями работы с большими данными (Apache Spark, Apache Ignite, Slurm).
2. Не приведены данные, позволяющие судить об эффективности масштабируемости полученной распределенной системы для обработки данных, что несколько нивелируются приведенными абсолютными цифрами количества обработанных заданий.

В отзыве Воеводина В.В. присутствует ряд замечаний и вопросов:

1. Автор в процессе изложения приводит много примеров реального использования созданных средств в рамках больших физических экспериментов. Один из ключевых моментов, определяющих не только эффективность работы, но и саму возможность функционирования

программных систем в столь масштабных условиях - это правильный выбор значений основных внутренних технологических параметров: размеров таблиц, буферов, значений таймаутов и других. В связи с этим хотелось бы в работе увидеть отдельный раздел, посвященный как описанию таких параметров (чтобы понимать, какие свойства действительно важны на практике), так и способу выбора правильных значений (на основе предварительного моделирования, эмпирически, динамический подбор в процессе работы или что-то иное). Неявно в работе это представлено в различных частях текста, но такой материал, собранный воедино в одном месте, безусловно, имеет самостоятельное значение и ценность.

2. Аналогично, очень хотелось бы видеть критический анализ предложенных подходов и методов, а также анализ функционирования созданных на их основе программных средств.

3. У автора огромный опыт организации и проведения крупномасштабных вычислительных экспериментов, в ходе которых наверняка становились понятными и "узкие места", и источники ненадёжности, и проблемы с масштабированием и многое другое, что всегда возникает в реальной работе распределенных программных комплексов. В чем основные проблемы? Как они устранялись? Какие внутренние параметры программных комплексов наиболее критичны для нормальной работы систем? Как обрабатывались пиковые нагрузки? Какие режимы работы систем особенно важны? На что следует разработчикам подобных систем обращать внимание в будущем? Это не описание "отрицательных" или же "негативных" сторон данной работы - ни в коем случае - это описание степени эластичности предложенных подходов и их поведения в условиях реальной нагрузки.

При этом во всех отзывах указывается, что данные замечания и вопросы не снижают достоинств работы, диссертация представляет собой законченную научно-квалификационную работу. Во всех отзывах подчёркивается, что диссертация удовлетворяет всем требованиям, предъявляемым к докторским диссертациям по специальности 05.13.11 - математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Ведущая организация Федеральное государственное бюджетное учреждение науки Институт ядерной физики им. Г.И. Будкера СО РАН (г. Новосибирск) в своём положительном отзыве, подписанном Эйдельманом Семеном Исааковичем, доктором физико-математических наук, главным научным сотрудником ИЯФ СО РАН, и утверждённом директором ИЯФ СО РАН, академиком РАН, доктором физико-математических наук, Павлом Владимировичем Логачевым, указала, что диссертационная работа Климентова А.А., посвящённая актуальной проблеме – обработке сверхбольших объёмов данных, развитию методов, архитектур, компьютерных моделей и программных систем, а также реализации соответствующих программных и инструментальных средств в ядерной физике и физике высоких энергий, является законченным научным исследованием, в котором получены новые фундаментальные результаты, и выполненной на высоком научном и методическом уровне, имеет научную новизну и практическую значимость. Достоверность данных и полученных результатов не вызывает сомнения, все выводы в достаточной мере обоснованы. В отзыве отмечается, что *«реализация разработанной диссертантом модели и её приложения в области физики элементарных частиц впервые позволила использовать различные архитектуры: грид, суперкомпьютеры, ресурсы облачных вычислений для обработки данных физического эксперимента через единую систему управления потоками заданий, сделав это "прозрачно" для пользователя. Разработка и реализация новых методов, архитектуры и программного инфраструктура системы для глобальной распределенной обработки данных позволили создать на этой основе уникальную систему управления потоками заданий, высокой производительности и масштабируемости (более двух миллионов задач, выполняемых ежедневно в 250 вычислительных центрах по всему миру). Была успешно решена проблема разделения вычислительных ресурсов между различными потоками научных заданий (обработка данных, Монте-Карло моделирование, физический анализ данных, приложения для триггера)»*. В качестве замечаний указано наличие некоторого количества опечаток и

пропущенных или неверно поставленных знаков препинания, а также нестандартных сокращений, например, «д.б.» вместо «должно быть». В отзыве ведущей организации указано, что диссертация является законченным научным исследованием, в котором получены новые фундаментальные результаты, и, безусловно, отвечает требованиям «Положения о присуждении учёных степеней» (п.9-14), утверждённого постановлением Правительства Российской Федерации N842 от 24 сентября 2013 г., предъявляемым к докторским диссертациям, а сам автор, Климентов Алексей Анатольевич, несомненно, заслуживает присуждения ему учёной степени доктора физико-математических наук по специальности 05.13.11 -математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Выбор официальных оппонентов и ведущей организации обусловлен авторитетностью и компетентностью назначенных оппонентов и широкой известностью достижений ведущей организации в тех областях математического моделирования, которые затрагивает диссертация.

По теме диссертации автором опубликовано свыше 150 печатных работ, в том числе по основным результатам — 68 (из них 47 из перечня ведущих рецензируемых научных изданий). Результаты работы также опубликованы в отчётах по руководимым автором инфраструктурным и научным проектам в рамках мегагранта Правительства РФ и проектам, поддержанным РНФ и РФФИ.

Основные работы, вошедшие в диссертацию, опубликованы в российских и международных журналах, входящих в Перечень ведущих рецензируемых научных изданий, неоднократно докладывались на семинарах и рабочих совещаниях таких экспериментов, как L3, AMS, COMPASS, ATLAS, ALICE, а также на российских и международных конференциях.

Наиболее важные публикации:

1. A. Klimentov, M. Grigorieva, A. Kiryanov, A. Zarochentsev. BigData and Computing Challenges in High Energy and Nuclear Physics, JINST Volume 12, June 2017, DOI:10.1088/1748-0221/12/06/C06044.
2. K. De, S. Jha, A. Klimentov, T. Maeno, R. Mashinistov, P. Nilsson, A. Novikov, D. Oleynik, S. Panitkin, A. Poyda, K.F. Read, E. Ryabinkin, A. Teslyuk, V. Velikhov,

J.C. Wells, and T. Wenaus. Integration of Panda Workload Management System with Supercomputers, ISSN 1547-4771, Physics of Particles and Nuclei Letters, 2016, Vol. 13, No. 5, c. 647–653.

3. F.B. Megino, A. Klimentov et al. on behalf of ATLAS collaboration. PanDA: Exascale Federation of Resources for the ATLAS Experiment at the LHC, EPJ Web Conf. 108 (2016) 01001.

4. A. Klimentov, K. De, S. Jha, T. Maeno, P. Nilsson, D. Oleynik, S. Panitkin, J. Wells, T. Wenaus. Integration of PanDA Workload Management System with Supercomputers for ATLAS and Data Intensive Science. Journal of Physics: Conference Series, 2016, Vol. 762, No. 1.

5. M. Borodin, K. De, J. Garcia Navarro, D. Golubkov, A. Klimentov, T. Maeno and A. Vaniachine, Scaling up ATLAS production system for the LHC Run 2 and beyond : project ProdSys2. J.Phys.Conf.Ser. 664 (2015) no.6, 062005.

6. A. Klimentov, P. Buncic, K. De, T. Maeno, T. Wenaus. Next Generation Workload Management System For Big Data on Heterogeneous Distributed Computing, J Phys.Conf. Ser. 608 (2015) no.1, 012040.

7. G. Aad, A. Klimentov et al. (ATLAS collaboration). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Physics Letters B, 716, 2012, p. 1–29.

8. A.Klimentov et al. (ATLAS collaboration). The ATLAS Simulation Infrastructure.2010. 53 pp. Eur.Phys.J. C70 (2010) 823-874 DOI: 10.1140/epjc/s10052-010-1429-9.

9. A.Klimentov et al (ATLAS collaboration). The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 S08003 DOI: 10.1088/1748-0221/3/08/S08003 (2008).

10. A. Klimentov et al. (ATLAS collaboration). Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. arXiv:0901.0512 [hep-ex]. 2008.

11. A. Klimentov et al (AMS Collaboration). The Alpha Magnetic Spectrometer (AMS) on the International Space Station, Part I, Results from the test flight on the Space Shuttle, Physics reports, vol.366/6, 331-404.

В опубликованных работах соискателя полностью отражены все

основные результаты диссертации и положения, выносимые на защиту. Вклад автора диссертационной работы в рассматриваемые публикации является определяющим: большинство работ написано им без соавторов.

На автореферат поступили отзывы от:

- 1) Егорычева Виктора Юрьевича, д.ф.-м.н., директора Федерального государственного бюджетного учреждения «Институт теоретической и экспериментальной физики имени А.И. Алиханова Национального исследовательского центра «Курчатовский институт»;
- 2) Якобовского Михаила Владимировича, д.ф.-м.н., член-корреспондента РАН, профессора, зам. директора ИПМ им. М.В. Келдыша РАН;
- 3) Беднякова Вадима Александровича, д.ф.-м.н., директора Лаборатории ядерных проблем ОИЯИ;
- 4) Петрова Игорь Борисович, д.ф.-м.н., профессора, член-корреспондента РАН, заведующего кафедрой информатики МФТИ;
- 5) Дегтярева Александра Борисовича, д.т.н., доцента, профессора кафедры компьютерного моделирования и многопроцессорных систем СПбГУ;
- 6) Федина Олега Львовича, к.ф.-м.н., руководителя Отделения физики высоких энергий НИЦ «Курчатовский институт» - ПИЯФ;
- 7) НИЦ «Курчатовский институт» - ИТЭФ, подписанного Соколовым Михаилом Михайловичем, начальником отдела 240 НИЦ «Курчатовский институт» - ИТЭФ;
- 8) Коноплича Ростислава Всеволодовича, д.ф.-м.н., профессора, зав. кафедрой физики Манхаттэновского Колледжа Нью-Йоркского университета.

Все отзывы положительные, в них отмечается актуальность диссертационного исследования, научная новизна, адекватность использованных современных методов решаемым задачам, достоверность полученных результатов, их теоретическая и практическая значимость.

Можно отметить отзыв Беднякова В.А., в котором он, как руководитель группы сотрудников ОИЯИ, с самого начала участвующих в эксперименте АТЛАС (ЦЕРН), отмечает, что созданная соискателем концепция и реализованный на ее основе компьютерный комплекс для обработки данных этого эксперимента оказались очень успешными, надёжно работающими и

имеющими ясные перспективы дальнейшего использования. По существу, именно правильно выбранная «компьютерная модель» эксперимента АТЛАС и обеспечила успешное решение главной задачи первого этапа экспериментов на Большом адронном коллайдере – открытие бозона Хиггса. Бедняков В.А. также отмечает, «что огромная и очень непростая работа 20-летнего периода по организации эффективного, быстрого, надёжного и удобного для пользователей компьютерно-математического обеспечения уникальных физических исследований нашего времени, наконец-то была оформлена Климентовым А.А. в виде диссертации на соискание учёной степени доктора физико-математических наук. Для меня совершенно очевидно, что соискатель уже давно заслуживает этой степени».

В двух отзывах присутствует несколько замечаний.

Так в отзыве Дегтярева А.Б. указывается:

1. «Положения, выносимые на защиту сформулированы не в виде утверждений, а вольными предложениями, из которых не до конца ясно, что выносятся на защиту».
2. В тексте и названии встречаются термины «большие данные» и «сверхбольшие объёмы данных», вместе с тем непонятно в каком смысле автор использует термины и какая метрика введена для их измерения.

В отзыве Беднякова В.А. отмечено, что автор злоупотребляет жаргонными словами и определёнными штампами, не все из которых расшифрованы автором и известны простому читателю (эксабайт, «приведённые данные», стрес-тест, T0-2, WAN и т.д.).

При этом в отзывах указывается, что данные замечания и вопросы не снижают достоинств работы, диссертация представляет собой законченную научно-квалификационную работу.

Во всех отзывах подчёркивается, что диссертация удовлетворяет всем требованиям, предъявляемым к докторским диссертациям по специальности 05.13.11 - математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Кроме отзывов на реферат в поддержку диссертации А.А. Климентова поступило семь писем от:

1. Киселя Ивана Васильевича, д.ф.-м.н., профессора Университета Франкфурта-на-Майне, Германия.
2. Andrea Contin, Professor, Physics Department, University of Bologna. Italia.
3. Martin Pohl, Professeur honoraire, La Faculté des Sciences, Université de Genève, Suisse.
4. Ian Bird, Dr., LHC Computing Grid Project Leader, IT Department CERN, Suisse.
5. Borut Paul Kersevan, Dr., Professor, Full Professor of Physics, University of Ljubljana, Slovenia.
6. Kaushik De, Professor of Physics, Director, Center of Excellence in HEP, University of Texas at Arlington, USA.
7. Peter Fisher, Professor of Physics, Physics Department Head, MIT, USA.

Диссертационный совет отмечает, что в результате выполненных соискателем исследований:

1. Разработана компьютерная модель для экспериментов в области ФВЭ и ЯФ, позволяющая объединять высокопропускные вычислительные мощности (грид), высокоскоростные вычислительные мощности суперкомпьютеры, ресурсы облачных вычислений и университетские кластеры в единую вычислительную среду.

2. Разработаны принципы построения и архитектура системы для глобальной обработки данных эксабайтного масштаба для тысяч пользователей в гетерогенной вычислительной среде.

3. Разработаны методы управления потоками заданий в гетерогенной вычислительной среде, позволяющие учитывать неоднородность потоков заданий и запросов пользователей, с целью оптимального использования вычислительных ресурсов, доступных в современном физическом эксперименте.

4. На основе разработанных принципов и архитектуры создана масштабируемая (обработка данных эксабайтного диапазона в $O(100)$ центрах $O(1000)$ пользователями $O(10^6)$ научных заданий/день) система для обработки данных современного физического эксперимента.

5. Разработана система мониторинга и оценки эффективности работы глобальной системы для обработки данных в распределенной гетерогенной компьютерной среде.

Практическая значимость. Основные результаты данной работы являются пионерскими и используются в действующих экспериментах в области ФВЭ и ЯФ, в других областях науки. В том числе результаты работ, положенных в основу диссертации, используются в двух крупнейших экспериментах в области ФВЭ и ЯФ – ATLAS и ALICE на LHC, эксперименте COMPASS на SPS, а также для приложений биоинформатики на суперкомпьютерах НИЦ КИ:

- вычислительные модели экспериментов ATLAS и AMS опираются на результаты работ, положенных в основу диссертации;
- разработанная и созданная система управления потоками заданий в гетерогенной компьютерной среде используется в экспериментах на ускорителях LHC и SPS и принята в качестве базовой для будущего коллайдера NICA;
- разработанная система для обработки данных была также применена для исследований ДНК мамонта на суперкомпьютере НИЦ КИ и в европейском проекте BlueBrain.

Разработанная система управления загрузкой не имеет мировых аналогов по масштабируемости и отказоустойчивости. До $2 \cdot 10^6$ задач выполняются ежедневно, в 2016 г. физиками ATLAS было обработано 1.4 Эбайта данных. Таким образом, система уже сейчас работает в эксабайтном диапазоне.

Реализация результатов работы. Результаты диссертации были получены под руководством и при личном участии соискателя в следующих международных проектах: WLCG — проект грид для LHC, megaPanDA — проект по созданию нового поколения системы управления заданиями в гетерогенной компьютерной среде, проект ATLAS на LHC, проекты AMS-01 и AMS-02 на Международной космической станции (МКС), проект metaMiner — по созданию системы поиска аномалий и предсказания поведения комплексных распределенных вычислительных систем, проект Federated Storage — по созданию прототипа распределенной компьютерной среды.

Автор диссертации внёс определяющий вклад при выполнении ряда национальных российских и международных проектов, в том числе L3, AMS, ATLAS, megaPanDA, в которых он являлся одним из руководителей (или руководителем) компьютерной и программной частями проекта и одновременно основным архитектором создаваемых систем и программного обеспечения.

Работы в 2013–2016 гг. были поддержаны грантом Министерства образования и науки РФ по привлечению ведущих учёных, тремя грантами РФФИ и грантом РНФ. В настоящее время автор является руководителем мегагранта и руководителем двух международных проектов совместно с ЦЕРН и DESY — «Создание федеративного распределенного дискового пространства» и «Использование алгоритмов машинного обучения для приложений ФВЭ».

Базовая вычислительная модель реализуется в проекте ATLAS на LHC и рассматривается как основная для ускорительного комплекса NICA (ОИЯИ). Созданы системы управления загрузкой для распределенной обработки данных в НИЦ КИ (для приложений биоинформатики), ОИЯИ (для эксперимента COMPASS в ЦЕРН), ЦЕРН (эксперименты ATLAS), ORNL (для высокоинтенсивных научных приложений), EPFL (проект BlueBrain, Лозанна, Швейцария), ASGC (эксперимент AMS-02, Тайпей, Тайвань).

Все представленные к защите положения и данные являются достоверными. Исследовательские работы и разработки по теме диссертации, от постановки задачи и выбора методики до получения результатов, выполнены соискателем и/или под его непосредственным руководством. Вклад соискателя в эти работы является определяющим. Все выносимые на защиту результаты получены соискателем лично.

На заседании 16 марта 2018 г. диссертационный совет принял решение присудить Климентову А.А. учёную степень доктора физико-математических наук.

На заседании № 70 из 21 человека, входящих в состав совета, присутствовали 17, из них 5 докторов по специальности рассматриваемой диссертации.

Проголосовали **за** присуждение Климентову Алексею Анатольевичу

учёной степени доктора физико-математических наук по специальности «05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» 17 человек, против 0 недействительных бюллетеней – 0.

Председатель

диссертационного совета

/Иванов В.В./

Ученый секретарь

диссертационного совета

/Иванченко И.М./

16 марта 2018 г.