



**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Направление подготовки/профиль: 09.06.01 — «Информатика и вычислительная техника» / 05.13.11 — «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Исследовательская школа физики высокоэнергетических процессов

Научный доклад об основных результатах подготовленной научно-квалификационной работы

Тема научного доклада
Развитие методов мониторинга систем распределенной обработки данных в гетерогенной компьютерной среде для приложений физики высоких энергий

УДК 004.75 – 047.36 : 539.121.667

Аспирант

Группа	ФИО	Подпись	Дата
A5-39	Корчуганова Татьяна Александровна		

Руководитель профиля подготовки

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор ОИТ, ИШИТР	Спицын Владимир Григорьевич	доцент, д-р техн. наук		

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор ОИТ, ИШИТР	Ким Валерий Львович	профессор, д-р техн. наук		

Научный руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
профессор НИЯУ МИФИ	Климентов Алексей Анатольевич	д-р физ.-мат. наук		

Аннотация

В течение последних 60 лет объем данных производимый установками в области физики высоких энергий увеличивался в геометрической прогрессии с единиц КБ до сотен ПБ и приблизился к эксабайтной отметке. Например, экспериментом ATLAS на Большом адронном коллайдере (БАК) за первые два сеанса сбора данных (2008-2012 и 2014-2018 гг.) было собрано 400 ПБ данных. Для обработки и анализа такого объема данных требуется комплексная вычислительная инфраструктура. Так экспериментами на БАК используется глобальная вычислительная инфраструктура WLCG (World LHC Computing Grid), а также суперкомпьютеры и ресурсы облачных вычислений. Программное обеспечение, используемое физическими экспериментами включает сложные системы по управлению потоками заданий обработки и анализа данных в распределенной компьютерной среде, выполняющих миллионы научных задач в день. Использование распределенных компьютерных ресурсов, а также объемы обрабатываемой информации требуют создания новых систем мониторинга. Системы мониторинга необходимы на каждом этапе обработки, моделирования и анализа данных. Для научного сообщества эти системы являются основным источником структурированной информации для принятия решений по настройке и отладке модулей системы управления загрузкой, а также поиску и выявлению проблем, связанных с самой инфраструктурой. Пользователями таких систем являются сотрудники физического эксперимента, отдельные физические группы, компьютерные профессионалы, обеспечивающие работу вычислительных центров, разработчики ПО для обработки данных. Кроме того система мониторинга должна предоставлять информацию национальным финансирующим организациям об использовании вычислительного ресурса.

В данной работе описываются методы, архитектурные принципы и подходы использующиеся для разработки и реализации системы мониторинга процесса распределенной обработки и анализа данных в гетерогенной компьютерной среде для современного физического эксперимента.