

# Моделирование системы распределенной обработки данных эксперимента VM@N в составе комплекса T0-T1 NICA

**В.В. Кореньков<sup>1,2</sup>, А.В. Нечаевский<sup>1,a</sup>, Г.А. Ососков<sup>1</sup>, Д.И. Пряхина<sup>1</sup>,  
В.В. Трофимов<sup>1</sup>, Ю.К. Потребеников<sup>1</sup>, А.В. Ужинский<sup>1</sup>**

<sup>1</sup>Объединенный институт ядерных исследований,  
Россия, 141980, Московская обл., г. Дубна, ул. Жолио-Кюри, д. 6

<sup>2</sup>РЭУ им. Г. В. Плеханова, Россия, 117997, Москва, Стремянный пер., д. 36

E-mail: <sup>a</sup>symsim@jinr.ru

В Лаборатории информационных технологий Объединенного Института Ядерных Исследований проводятся работы по созданию компьютерного off-line комплекса для моделирования, обработки, анализа и хранения данных комплекса NICA. В состав комплекса входят различные экспериментальные установки, в том числе VM@N для проведения экспериментов на Нуклотроне с выведенными пучками тяжелых ионов. В 2017 году планируется запуск эксперимента VM@N, в связи с чем необходимо создать систему распределенной обработки данных, полученных с установки. Для выбора архитектуры распределенной системы сбора, хранения и обработки данных и определения необходимого состава оборудования используется программа моделирования SyMSim. В работе представлены результаты моделирования вычислительного комплекса по приему и обработке данных с эксперимента VM@N и некоторые рекомендации по выбору оборудования.

Ключевые слова: NICA, распределенная обработка данных, моделирование

© 2016 Владимир Васильевич Кореньков, Андрей Васильевич Нечаевский, Геннадий Алексеевич Ососков,  
Дарья Игоревна Пряхина, Юрий Константинович Потребеников, Владимир Валентинович Трофимов,  
Александр Владимирович Ужинский

## Введение

В Лаборатории физики высоких энергий (ЛФВЭ) Объединенного института ядерных исследований (ОИЯИ) ведутся работы по созданию сверхпроводящего ускорителя на встречных пучках тяжелых ионов NICA, который планируется сдать в эксплуатацию в 2020 году [Bashashin, 2016]. В состав комплекса входят различные экспериментальные установки, в том числе детектор VM@N для физических исследований по изучению плотной барионной материи на выведенных пучках Нуклотрона [Conceptual Design Report, 2016].

Весной 2015 года был выполнен первый запуск детектора VM@N с целью технической проверки установки. На 2017 год запланирован следующий запуск установки VM@N, что требует разработки распределенной вычислительной системы для хранения и обработки физических данных с учетом конкретных параметров эксперимента. Создание подобного вычислительного комплекса для обработки больших потоков данных является весьма сложной задачей. Для выбора оптимальной архитектуры будущего комплекса с учетом предполагаемых ресурсов и загрузки необходимо определить его основные параметры, структуру и проверить предлагаемые технические решения путем тщательного моделирования, как состава разрабатываемой структуры вычислительного центра, так и потока поступления заданий с учетом их типов и статистических данных о распределении.

В Лаборатории информационных технологий (ЛИТ) ОИЯИ создана программа SyMSim для имитационного моделирования систем хранения и обработки данных крупных научных экспериментов, в которой существует специальная база данных для хранения различных параметров моделирования и результатов работы. Загрузка параметров и визуализация результатов моделирования выполняется через веб-портал [Korenkov, 2015]

## Система обработки данных

Как показывает опыт мировых центров экспериментальной ядерной физики и физики высоких энергий, основные этапы обработки данных - моделирование, реконструкция и анализ проводятся в распределенных вычислительных структурах, построенных по иерархическому принципу типа Tier-0 - Tier-1 [Korenkov, 2016]. Такая распределенная грид-инфраструктура будет использована для создания компьютерного комплекса проекта NICA.

Комплекс Tier-0, который будет расположен в ЛФВЭ, предназначен для сбора данных эксперимента VM@N. «Сырые» (Raw) данные эксперимента записываются на дисковые накопители для последующей передачи по коммуникационным линиям связи в центр уровня Tier-1, расположенный в ЛИТ. Планируется, что количество данных, полученных с установки, будет на уровне 1 ПБ за запуск. Данные будут записываться на диски и одновременно обрабатываться на компьютерной инфраструктуре ЛИТ. На сегодняшний день между VM@N и ЛИТ существует линия связи пропускной способностью в 10 Гбит/с, которой будет достаточно для передачи всех экспериментальных данных на диски ЛИТ. Планируется поступление четырех типов задач: кроме задач сохранения Raw, существуют также задачи экспресс-анализа, реконструкции и моделирования событий.

В таблице 1 приведены прогнозные величины по потоку данных эксперимента VM@N. Количество событий обусловлено удобством манипуляции с файлами. Каждую секунду с VM@N поступает 2 файла. 5% в каждом файле занимают события, предназначенные для экспресс анализа.

Таблица 1. Прогнозные параметры данных эксперимента VM@N

Параметр	Значение
Частота появления событий на выходе электроники	10000 Гц
Размер события на выходе электроники	1 МБ
Количество событий в файле	5000 событий
Количество событий для одного экспресс анализа	250 событий
Время обработки одного события	1 с
Время обработки одного события на стандартизованном процессоре	1 с

## Результаты моделирования

Моделирование работы эксперимента в течении 270 минут программой SyMSim показало, что максимальная загрузка канала связи достигает 600 МБ/с (см. рис. 1а)).

Вычислительная ферма в ЛИТ содержит 1000 ядер для обработки заданий. На рис. 1б) изображено количество свободных ядер в процессе обработки заданий экспресс анализа. По графику можно сделать вывод, что при постоянной скорости набора данных для полной обработки заданий достаточно 880 ядер, а для экспресс обработки — 120 ядер.

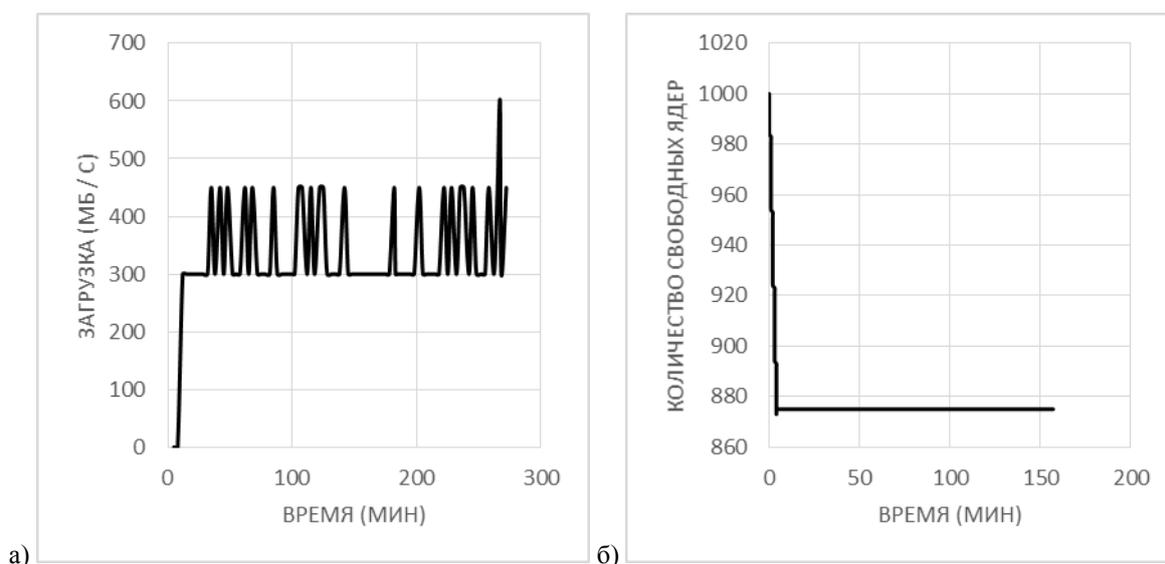


Рис. 1: а) Загрузка канала связи VM@N - ЛИТ. б) Экспресс обработка на ферме ЛИТ

Моделировалось также количество свободных ядер на ферме ЛИТ во время выполнения заданий реконструкции в зависимости от частоты их запуска, при этом полная обработка осуществлялась одновременно с поступлением событий.

Если задачи поступают каждые 9 с., то более 100 ядер остаются свободными, то есть ферма используется не полностью. При поступлении заданий каждые 7 с. ферма будет загружена полностью, но как показало моделирование, возникнут растущие очереди заданий на обработку, что может привести к переполнению ресурсов фермы при продолжительной работе эксперимента. Сравнение количества поступающих задач и задач, находящихся в очереди на исполнение показало, что проблем переполнения ресурсов можно избежать, если уменьшить темп поступления заданий, например, до 8 с., или ускорить их обработку.

Предполагается, что эксперимент VM@N будет продолжаться 1500 часов. Первые 120 часов работы эксперимента будет осуществляться настройка оборудования, поэтому информация записываться не будет. Затем, в течение 1000 часов будут накапливаться данные с эксперимента при интенсивности потока 70% от максимальной и записываться на диски ЛИТ,

при этом полная обработка заданий будет осуществляться на 880 ядрах, экспресс обработка – на 120 ядрах. Оставшиеся 380 часов данные будут накапливаться с эффективностью 100%. После завершения эксперимента, т.е. окончания набора данных, полная обработка заданий будет осуществляться на 1000 ядрах.

Ранее предполагалось, что для хранения данных с эксперимента будет достаточно 1 ПБ на дисках. Моделирование будущей системы хранения и обработки данных показало, что такого объема памяти будет недостаточно и для хранения всех файлов потребуется 1200 ТБ дискового пространства.

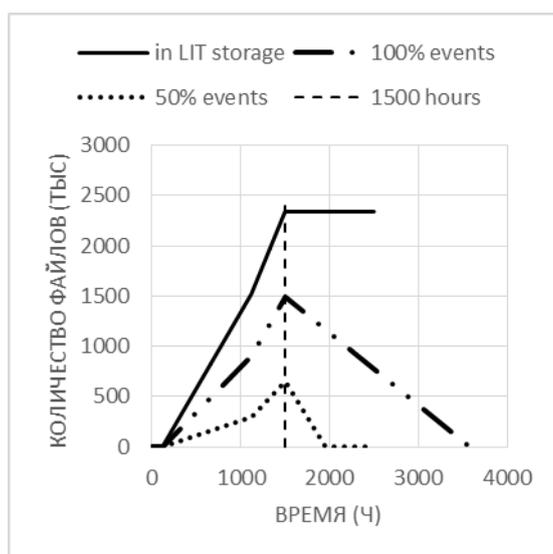


Рис. 2: Два способа обработки событий

Также моделировались две стратегии пофайловой обработки событий: обработка 100% событий каждого файла, что приведет к накоплению большой очереди необработанных событий, которые придется обрабатывать потом после окончания сеанса, или обработка только 50%, т.е. каждого второго события, что существенно уменьшит очередь. Одно событие, в среднем, обрабатывается 1 с, за один запуск эксперимента будет получено 2500000 файлов, а on-line обработка событий осуществляется одновременно с их поступлением. Это значит, что потребуется некоторое количество времени после окончания эксперимента, чтобы полностью обработать все поступившие события. Как показывает рис.2 с результатами моделирования этих двух стратегий, если будет обрабатываться каждое событие в файле (100% событий), то после окончания эксперимента потребуется 2000 часов, чтобы закончить обработку всех событий на 1000 ядрах фермы. Если же обрабатывать каждое второе событие (50%), то для завершения обработки потребуется только 500 часов.

## Заключение

Результаты моделирования будущей системы распределенной обработки данных первого рана эксперимента VM@N показали, что:

- максимальная загрузка канала связи между VM@N и ЛИТ составляет 600 Мбайт/с, поэтому для приема всех данных с VM@N на диски ЛИТ достаточно существующего канала связи с пропускной способностью 10 Гбит/с;
- при постоянной скорости набора данных для полной обработки заданий потребуется 880 ядер, а для экспресс анализа — 120 ядер;
- для хранения всех файлов в ЛИТ потребуется 1200 ТБ на дисках;

— если при обработке при одновременном поступлении событий запускать задания каждые 9 с, то ферма ЛИТ будет использована не полностью, а при запуске каждые 7 с возникнут растущие очереди;

— если будет обрабатываться 100% событий файла, то после окончания эксперимента потребуется еще 2000 часов, чтобы полностью закончить обработку поступивших заданий; при обработке 50% событий файла — 500 часов.

Полученные результаты демонстрируют, что моделирование необходимо при разработке вычислительных комплексов обработки больших потоков данных, т.к. позволяет принять оптимальное решение в выборе архитектуры будущего комплекса.

## Список литературы

*Bashashin M. V., Kekelidze D. V., Kostromin S. A., Korenkov V. V., Kuniaev S. V., Morozov V. V., Potrebenikov Yu. K., Trubnikov G. V., Philippov A. V.* NICA project management information system // *Particles and Nuclei Letters*. — 2016. — Vol. 13, No 5. — P. 969–973.

*Conceptual Design Report. BM@N - Baryonic Matter at Nuclotron.* [Electronic resource]. URL: [http://nica.jinr.ru/files/BM@N/BMN\\_CDR.pdf](http://nica.jinr.ru/files/BM@N/BMN_CDR.pdf) (accessed 30.06.2016).

*Korenkov V., Nechaevskiy A., Ososkov G., Pryahina D., Trofimov V., Uzhinskiy A., Balashov N.* Web-Service Development of the Grid-Cloud Simulation Tools // *Procedia Computer Science: 4th International Young Scientist Conference on Computational Science*. — 2015. — Vol. 66 — P. 533–539.

*Korenkov V. V., Nechaevskiy A. V., Ososkov G. A., Pryahina D. I., Trofimov V. V., Uzhinskiy A. V.* Simulation concept of NICA-MPD-SPD Tier0-Tier1 computing facilities // *Particles and Nuclei Letters*. — 2016. — Vol. 13, No 5. — P. 1074–1083.

## Simulation of distributed data processing system for BM@N experiment of T0-T1 NICA project

**V.V. Korenkov<sup>1,2,a</sup>, A.V. Nechaevskiy<sup>1</sup>, G.A. Ososkov<sup>1</sup>, D.I. Pryahina<sup>1</sup>,  
Yu.K. Potrebenikov<sup>1</sup>, V.V. Trofimov<sup>1</sup>, A.V. Uzhinskiy<sup>1</sup>**

<sup>1</sup> Joint Institute for Nuclear Research, 6 Joliot-Curie street, Dubna, Moscow region, 141980, Russia

<sup>2</sup> Plekhanov Russian University of Economics, 36 Stremyanny per., Moscow, 117997, Russia

E-mail: <sup>a</sup>symsim@jinr.ru

Computing for the off-line NICA complex is developing in the Laboratory of Information Technologies of the Joint Institute for Nuclear Research with the aim to simulate, process, analyze and store experimental data. The NICA complex consists of different experimental facilities, including BM@N experiment at the Nuclotron with pass heavy-ion beams. BM@N experiment run is planned in 2017 what requires developing a distributed system for processing data from the facility. The simulation program SyMSim is applied to select the infrastructure architecture of the system for data acquisition, storing and processing. SyMSim also facilitates decisions taking to choose required equipment. The paper presents simulation results of the computing complex for receiving and processing data from the BM@N experiment and some recommendation for choosing equipment.

Keywords: NICA, distributed data processing, simulation