

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Развитие компьютерной инфраструктуры

Для достижения главных целей ведущих проектов ОИЯИ потребуется обрабатывать огромное количество экспериментальных данных. Согласно весьма грубой оценке — это десятки тысяч процессорных ядер. В частности, для проекта NICA необходимы грид-инфраструктуры уровней Tier0, Tier1 и Tier2, для нейтринной программы ОИЯИ – вычислительные ресурсы и ресурсы хранения. Для поддержки стратегических исследований в ОИЯИ необходимо развивать распределенные многоуровневые гетерогенные вычислительные среды, в том числе и на ресурсах участников экспериментов.

Предполагается, что центры Tier0 и Tier1 для проекта NICA будут построены на ресурсах ОИЯИ, включая сотни петабайт долговременного хранилища необработанных данных. Это позволит обеспечить 25–30 % всех вычислительных ресурсов в распределенной системе, предоставление и поддержку основных сервисов для распределенной вычислительной системы (DIRAC, PanDA и др.).

Емкость хранилища данных и вычислительных мощностей для проекта WLCG, направленного на решение задач, связанных с участием ОИЯИ в экспериментах ЦЕРН, должны увеличиваться ежегодно на 10–20 %, что позволит сохранить требуемую скорость обработки данных.

Для разработки новых алгоритмов обработки и анализа данных на основе глубокого и машинного обучения потребуется поддержка и развитие инфраструктуры высокопроизводительных вычислений. Суперкомпьютер «Говорун» — это гибкая, масштабируемая, гиперконвергентная система, сочетающая в себе вычислительные архитектуры разных типов, иерархическую систему обработки и хранения данных. Развитие СК «Говорун» направлено на создание среды для суперкомпьютерного моделирования и решения ресурсоемких теоретических и экспериментальных задач ОИЯИ. Такая исследовательская среда необходима для параллельных вычислений, задач ML/DL/AI, квантовых вычислений, инструментов анализа и визуализации данных, прикладных пакетов, веб-сервисов для прикладных программ, учебных курсов и практик.

Одним из главных приоритетов Семилетнего плана является расширение облачной инфраструктуры ОИЯИ и создание интегрированной облачной среды для экспериментов ОИЯИ и его стран-участниц на основе технологий контейнеризации. Прогресс в этой области будет во многом зависеть от готовности экспериментов к переходу на такой рабочий процесс.

Развитие информационных технологий напрямую связано с дальнейшим развитием сетевой инфраструктуры ОИЯИ. Поддержка современных сетевых технологий включает программно-определяемые сети (SDN), сети доставки содержимого (CDN), именованные сети передачи данных (NDN) и технологии построения распределенных центров обработки данных (DCI) Data Center Interconnect.

Стратегия развития Больших данных в ОИЯИ включает широкий спектр исследований: подготовку инфраструктуры хранения и обработки Больших данных (аппаратное и программное обеспечение, безопасность); разработку современных методов и алгоритмов Больших данных для решения прикладных задач; интеллектуальный мониторинг функционирования и безопасности распределенных вычислительных систем; предоставление инфраструктуры Больших данных для конечных пользователей.

Первоочередной задачей в области разработки и применения квантовых вычислений, квантовой программной инженерии и квантового интеллектуального управления является построение квантовых систем интеллектуального управления физическими экспериментальными установками, в том числе для случаев непредвиденных и непредсказуемых ситуаций.

Исходя из этих потребностей, основное направление развития ИТ-экосистемы ЛИТ связано с модернизацией сетевых каналов связи, инженерной и вычислительной инфраструктуры Многофункционального информационно-вычислительного комплекса (МИВК), а также с развитием технологий обработки и хранения данных для экспериментов NICA и нейтринной программы

ОИЯИ. Для обеспечения стабильной работы МИВК требуется регулярное обновление и обслуживание существующей инфраструктуры.

ИТ-экосистема станет базовой платформой для подготовки ИТ-специалистов, способных разрабатывать алгоритмические и программные решения для задач ОИЯИ.

Все работы будут проводиться в тесном сотрудничестве с исследовательскими группами и ИТ-специалистами из всех лабораторий ОИЯИ и стран-участниц.



Рис. 32. Элементы Многофункционального информационно-вычислительного комплекса

Ожидаемые результаты:

- Модернизация инженерной инфраструктуры МИВК ОИЯИ (реконструкция в соответствии с современными требованиями машинного зала 4-го этажа ЛИТ).
- Модернизация и развитие офлайн распределенной вычислительной платформы для проекта NICA с привлечением вычислительных центров коллаборации NICA.
- Создание грид-кластера Tier0 для экспериментов мегапроекта NICA для хранения экспериментальных и смоделированных данных. Расширение производительности и емкости систем хранения грид-кластеров Tier1 и интегрированного Tier2/CCIC в качестве центров обработки данных для экспериментов мегапроекта NICA, нейтринной программы ОИЯИ и экспериментов на LHC.
- Расширение облачной инфраструктуры ОИЯИ с целью увеличения предоставляемого пользователям спектра сервисов на основе технологий контейнеризации. Автоматизация развертывания облачных технологий в организациях стран-участниц ОИЯИ.
- Расширение гетерогенной платформы HybriLIT, включая суперкомпьютер «Говорун», как гиперконвергентной программно-определяемой среды с иерархической системой хранения и обработки данных.

- Создание многоцелевой программно-аппаратной платформы аналитики Больших данных на основе гибридных аппаратных ускорителей (GPU, FPGA, квантовые системы); алгоритмов машинного обучения; инструментов аналитики, отчетов и визуализации; поддержки пользовательских интерфейсов и задач.
- Проектирование и разработка распределенной программно-конфигурируемой высокопроизводительной вычислительной платформы, объединяющей суперкомпьютерные (гетерогенные), грид- и облачные технологии для эффективного использования новых вычислительных архитектур.
- Разработка системы защиты компьютерной инфраструктуры на основе принципиально новых парадигм, включая квантовую криптографию, нейрокогнитивные принципы организации данных и взаимодействия объектов данных, глобальную интеграцию информационных систем, универсальный доступ к приложениям, новые интернет-протоколы, виртуализацию, социальные сети, данные мобильных устройств и геолокации.

Таблица 7. Примерная оценка требуемых вычислительных ресурсов

		2024	2025	2026	2027	2028	2029	2030
LHC Tier1 (CMS)	CPU (Pflops)	1,53	1,69	1,84	2,03	2,22	2,45	2,68
	Disk (PB)	18	20	25	28	31	34	40
	Tape (PB)	46	50	60	70	80	90	100
	Network (Gbps)	200	400	400	600	600	800	800
LHC Tier2 (ATLAS, CMS, ALICE, LHCb и др.)	CPU (Pflops)	0,73	0,81	0,88	0,96	1,04	1,15	1,27
	Disk (PB)	7,7	8,5	9,2	10	11,	12,80	14
	Network (Gbps)	200	400	400	600	600	800	800
СК «Говорун»	CPU (Pflops)	1,2	2,2	3,2	4,2	5,2	6,2	7,2
	Disk (PB)	8	9	10	11	12	13	14
Озеро данных	Disk (PB)	60	60	60	80	80	80	100
*NICA Tier 0,1,2	CPU (Pflops)	2,2	2,6	8,6	8,6	15,6	15,6	15,6
	Disk (PB)	17	24	47	75	96	119	142
	Tape (PB)	45	88	170	226	352	444	536
	Network (Gbps)	400	400	400	400	400	400	400
*Baikal-GVD, NOvA, JUNO, DUNE Tier 0,1,2	CPU (Pflops)	0,94	1,02	1,2	1,28	1,36	1,54	1,62
	Disk (PB)	1,9	3,2	3,5	3,8	4,6	4,9	5,2
	Tape (PB)	9	12	15	18	21	24	27
	Network (Gbps)	200	200	200	200	200	200	200

* Финансирование вычислительных ресурсов на компьютеринг по NICA и Нейтринной программе ОИЯИ будет осуществляться в рамках бюджетов соответствующих направлений.

Математическая поддержка исследований, проводимых в ОИЯИ

Важным направлением деятельности является обеспечение математической, алгоритмической и программной поддержки экспериментальных и теоретических исследований, проводимых в ОИЯИ. Целью являются моделирование физических процессов, создание алгоритмов и программных комплексов для обработки и анализа экспериментальных данных, разработка алгоритмов в области машинного и глубокого обучения, искусственного интеллекта и когнитивной интеллектуальной робототехники, систем квантового интеллектуального управления, развитие методов компьютерной алгебры и квантовых вычислений, а также аналитики Больших данных.

Ожидаемые результаты

1. Развитие информационно-вычислительных систем для анализа и обработки экспериментальных данных в области радиобиологии.

2. Развитие алгоритмов на основе рекуррентных и сверточных нейронных сетей для задач машинного и глубокого обучения и аналитики Больших данных, предназначенных в первую очередь для решения различных задач в экспериментах по физике частиц, в том числе для мегапроекта NICA и нейтринных экспериментов.

3. Создание современных инструментов исследований для международных коллабораций (NICA, нейтринная программа ОИЯИ, эксперименты на LHC).

4. Развитие новых численных и вычислительных моделей, включая квантовые вычисления, для теоретических исследований, проводимых в ОИЯИ.

5. Разработка алгоритмов интеллектуального управления физическими экспериментальными установками ОИЯИ на основе квантового подхода.

6. Разработка на базе платформы аналитики Больших данных системы для анализа и защиты данных компьютерной сети ОИЯИ в режиме реального времени на основе сетевого трафика.

7. Развитие алгоритмов машинного обучения и искусственного интеллекта для оптимизации функционирования и интеллектуального мониторинга распределенных вычислительных систем.

8. Создание аналитической системы нового поколения на основе эффективных методов и алгоритмов формализации, извлечения знаний и обработки Больших данных.

9. Разработка интеллектуальных информационных систем для научных исследований и приложений.

10. Развитие квантовых ИТ-технологий обработки данных с доступом к NISQ (Noisy Intermediate-Scale Quantum) компьютерам/квантовым компьютерам с надежной защитой от ошибок.

11. Разработка масштабируемых алгоритмов и программного обеспечения для обработки многопараметрических, многомерных, иерархических наборов данных эксабайтного объема.

Цифровая экосистема ОИЯИ

Одной из наиболее важных задач Семилетнего плана является создание общеинститутской цифровой платформы «Цифровая экосистема ОИЯИ». Основной целью является организация цифрового пространства с единым доступом и обменом данными между электронными системами, а также автоматизация действий, требовавших ранее личного или письменного обращения. Платформа должна обеспечить интеграцию существующих и перспективных сервисов поддержки научной, административной и социальной деятельности, а также сопровождение инженерной и ИТ-инфраструктур Института.

Пользователь получит возможность единой точки входа в цифровую среду ОИЯИ, через которую будет осуществляться доступ к масштабной сети разнообразных сервисов. Интерфейс «Цифровой экосистемы» будет представлять собой «витрину» цифровых сервисов и ресурсов с возможностью осуществления определенного набора действий (например, управления учетными записями)



Рис. 33. Направления развития в рамках математической поддержки исследований, проводимых в ОИЯИ

либо перехода на полнофункциональную версию сервиса. Примерами сервисов являются ресурсы для пользователей базовых установок, библиотечные сервисы, серверы документов, вычислительные ресурсы МИВК, административные сервисы 1С (финансы, кадры, электронный документооборот) и т.д.

В рамках создаваемой платформы зарегистрированные пользователи (имеющие учетную запись ОИЯИ – Single Sign-On, SSO) смогут оформлять и согласовывать различные документы в электронном виде, а также регистрироваться и использовать научные и административные сервисы без заполнения бумажных форм и личного посещения ответственных за них сотрудников. В личном кабинете сотрудника будет доступна система оповещений от различных сервисов (например, о документах, ожидающих подписания). Уровень доступа к сервисам будет зависеть от должности сотрудника и выполняемых им функциональных обязанностей. Для администраторов сервисов будет организован удобный интерфейс, позволяющий оперативно обновлять информацию. Часть ресурсов станет доступна и для незарегистрированных пользователей: телефонный справочник, информация по диссертационным советам, научное программное обеспечение, карта ОИЯИ.

В рамках цифровой платформы будет развиваться геоинформационная система ОИЯИ, включающая интерактивную карту, информацию по зданиям и прочим объектам ОИЯИ (планы зданий, инженерные и прочие сети, размещение персонала, учет и анализ использования помещений с учетом класса, типа и предназначения) и т.д. Геоинформационная система позволит осуществлять быстрый и удобный поиск информации как по зданиям, так и по сотрудникам ОИЯИ. Предполагается использование технологии мобильных роботов и элементов квантового управления для решения задач автоматической экспликации помещений (построения планов зданий) и локализации объектов на карте.

Платформа должна предоставлять надежный и безопасный доступ к данным различного типа, возникающим в процессе работы Института, — от открытых до конфиденциальных. Выборка данных из ключевых сервисов будет помещаться в хранилище для дальнейшего совместного анализа с использованием технологий Больших данных и искусственного интеллекта. На основе таких данных, как сведения о публикациях сотрудников, финансовая информация, использование вычислительных ресурсов, будет возможен автоматизированный мониторинг показателей функционирования как отдельных проектов, так и Института в целом.

Ожидаемые результаты

Создание платформы «Цифровая экосистема ОИЯИ».



Рис. 34. Цифровая экосистема ОИЯИ

Таблица 8. МИВК

(тыс. долларов США)

Виды расходов	2024	2025	2026	2027	2028	2029	2030	ВСЕГО
Расходы на персонал (ст. 1-3)								
Материальные расходы на создание, развитие, модернизацию (ст. 5, 6, 9, 10, 18, 19)								40 000,0
Расходы на эксплуатацию и обслуживание (ст. 5, 6, 7, 8, 10, 14)								14 757,4
Расходы на международное сотрудничество (ст. 4)								
Сервисные расходы (ст. 11,12,13,14,15,16,17))								
ВСЕГО								