



Разработка сервиса для проведения радиобиологических исследований на платформе HybriLIT

Юрий Бутенко

Группа по гетерогенным вычислениям «HybriLIT»
Лаборатория информационных технологий, ОИЯИ



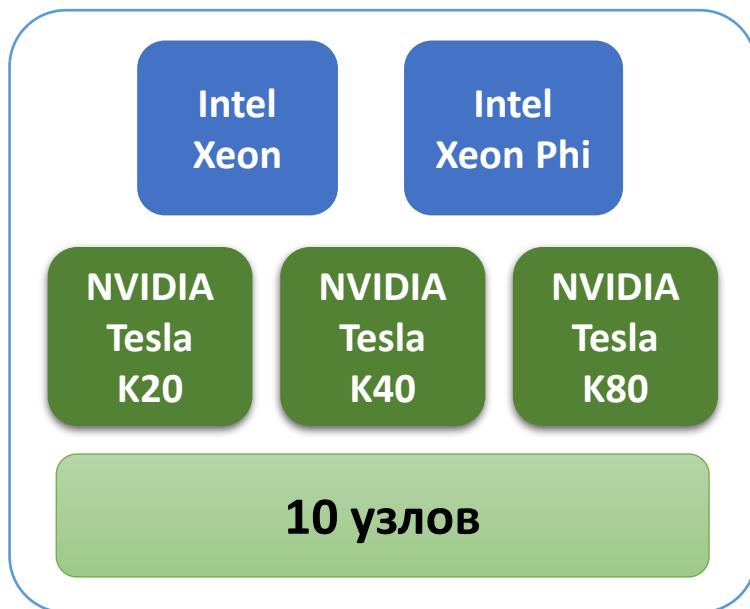
Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



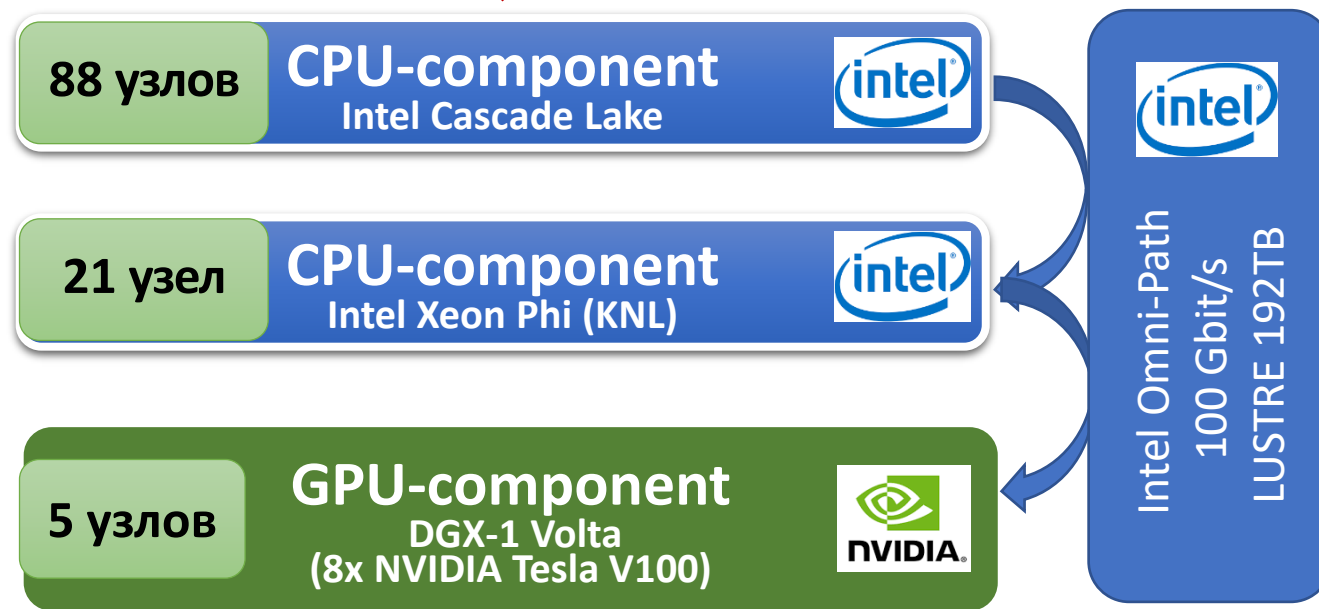
Гетерогенная платформа HybriLiT

Общее информационно-программное окружение

Учебно-тестовый полигон
«HybriLiT»



Суперкомпьютер «Говорун»

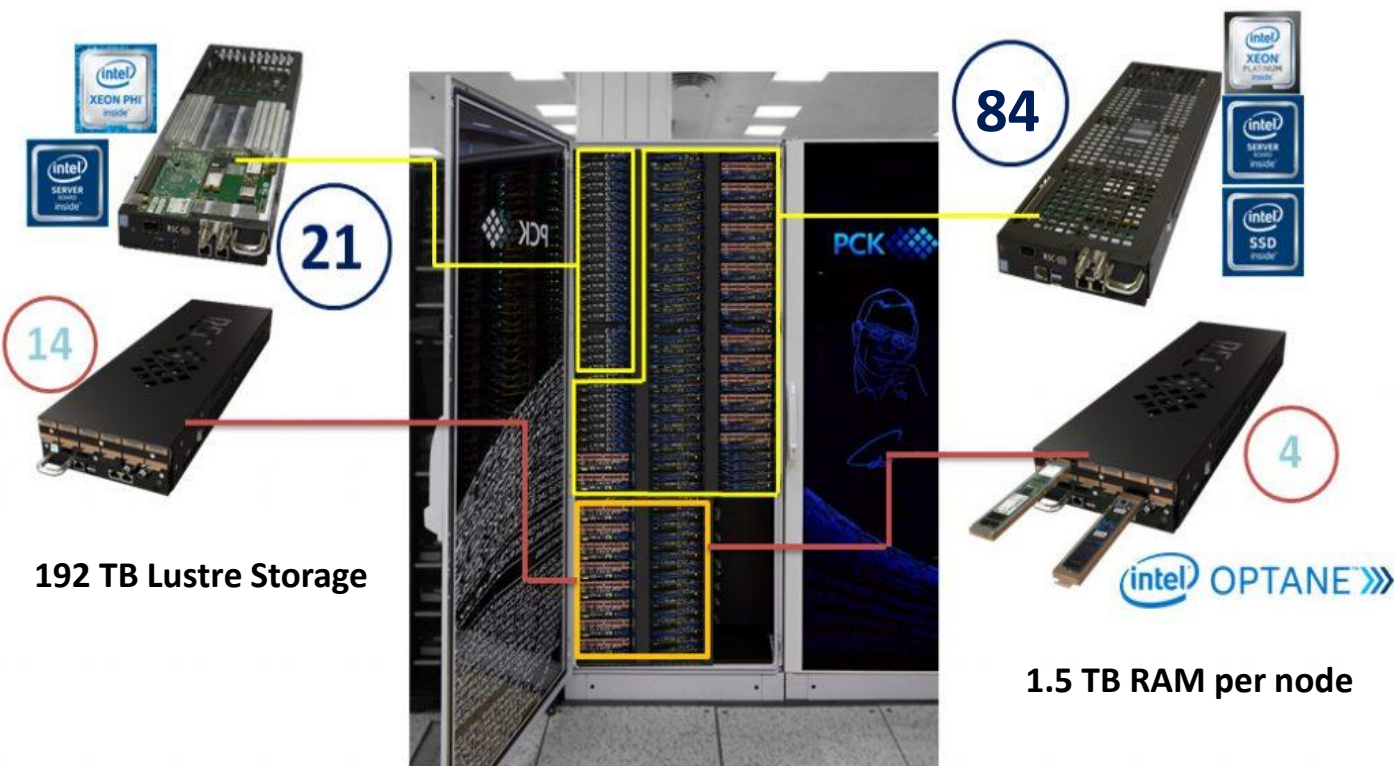




Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Суперкомпьютер "Говорун"



RSC Tornado nodes based on Intel® Xeon Phi™:

- Intel® Xeon Phi™ 7190 processors (72 cores)
- Intel® Server Board S7200AP
- Intel® SSD DC S3520 (SATA, M.2)
- 96GB DDR4 2400 GHz RAM
- Intel® Omni-Path 100 Gb/s adapter

RSC Tornado nodes based on Intel® Xeon® Scalable gen 2:

- Intel® Xeon® Platinum 8268 processors (24 cores)
- Intel® Server Board S2600BP
- Intel® SSD DC S4510 (SATA, M.2),
- 2x Intel® SSD DC P4511 (NVMe, M.2) 2TB
- 192GB DDR4 2933 GHz RAM
- Intel® Omni-Path 100 Gb/s adapter



Cascade lake: 4224 ядер (не учитывая HT) + 22 TB RAM

KNL: 1512 ядер (не учитывая HT) + 2 TB RAM



Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Суперкомпьютер «Говорун»

NVIDIA DGX-1 – это первая система, разработанная специально для задач глубокого обучения. В основе системы лежат графические ускорители поколения Volta, которые обеспечивают скорость обработки данных, сравнимую с 250 серверами x86 архитектуры.

В состав СК «Говорун» входит 5 серверов DGX-1 с вычислительными процессорами двух типов:

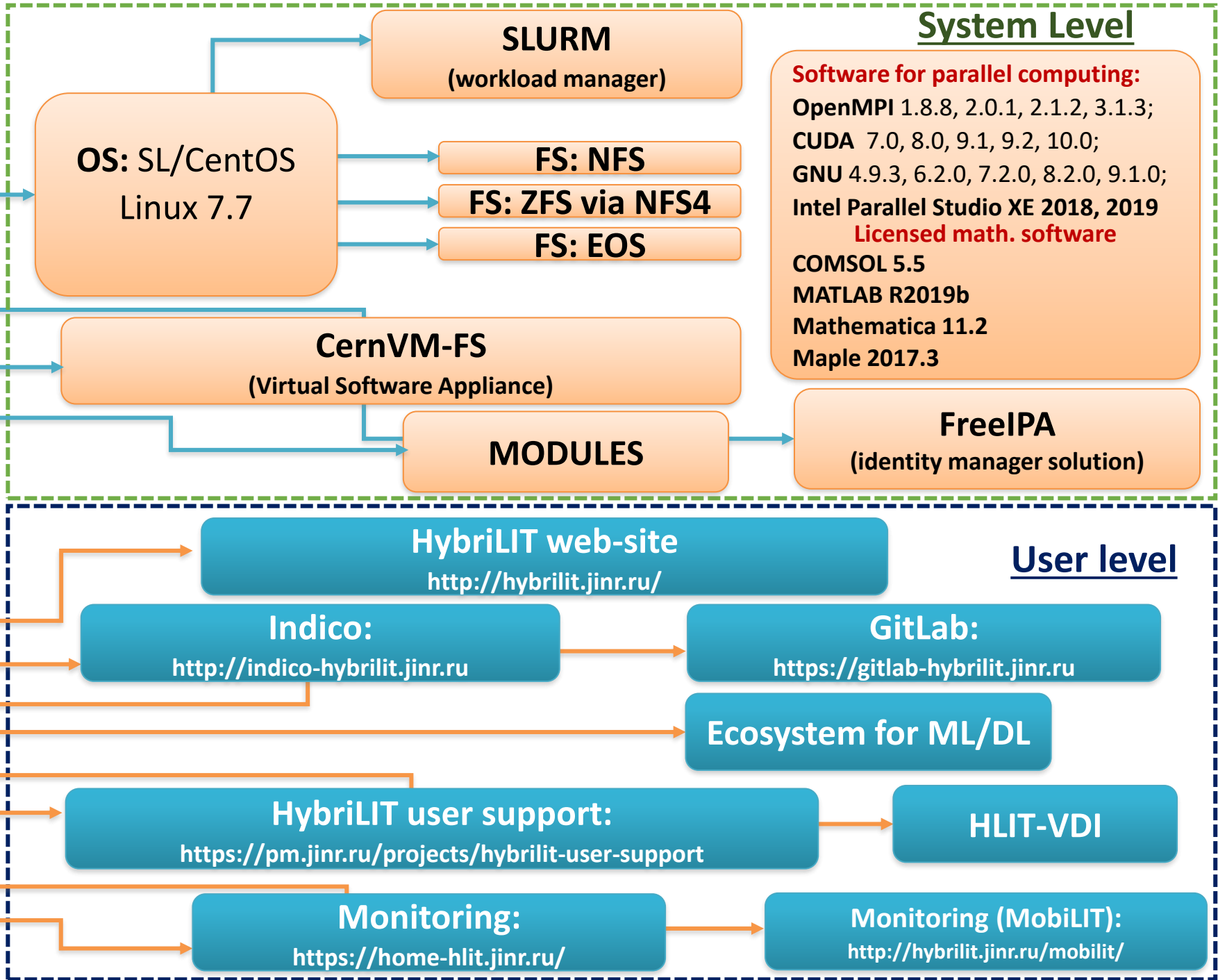
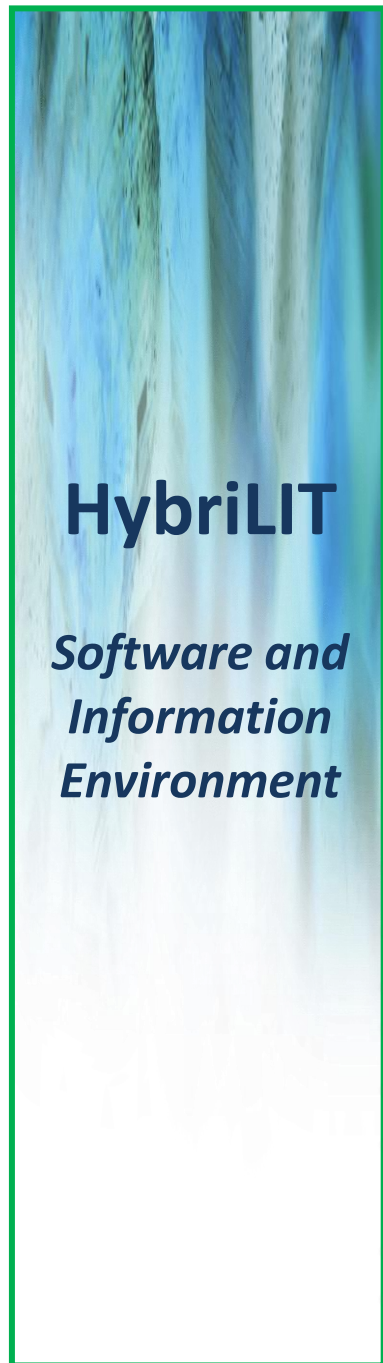
- 2x Intel Xeon E5-2698 v4 20 ядер;
- 8x NVIDIA Tesla V100



NVIDIA Tesla V100

Двойная точность	7.8 Tflops
Одинарная точность	15.7 Tflops
Глубокое обучение	125 Tflops







Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Экосистема ML/DL и анализа данных + JLabHPC

Удобное и быстрое прототипирование ML/DL алгоритмов в среде Jupyter Notebook

Для обучения сверточных нейронных сетей доступно 4x GPU Volta V100

Собраны и готовы к использованию популярные фреймворки для DL/ML/CV (TensorFlow, Keras, PyTorch, OpenCV итд.)

Большой выбор доступных Python-библиотек для анализа и визуализации данных



Аппаратные характеристики:

- GPU: 4x Nvidia Volta V100-SXM2 *NVLink* 32Gb HBM2
- CPU : 2x Intel Xeon Gold 6148 CPU @ 2.40GHz
20 Cores/40 Threads
- RAM: 512 GB DDR4 2666MHz
- SSD: 2*240 GB





Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Платформа HybriLIT это не только вычислительные мощности

Большое количество предустановленного ПО и инструментов

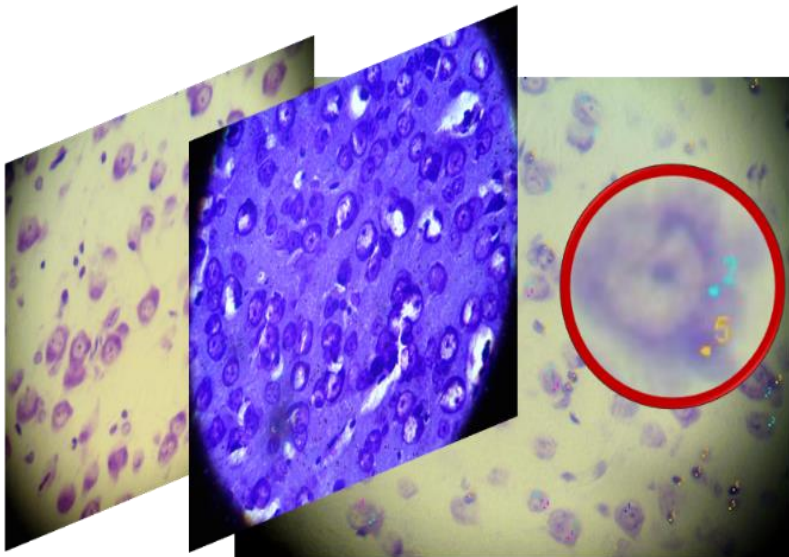
Суммарное дисковое хранилище платформы (Hlit-EOS, NFS, ZFS, Lustre) ~ **792 TB**
+ доступен JINR-EOS **4 PB**

Storage-компоненты и ПО доступно на всех сервисах и узлах платформы

Большое количество доступных сервисов: GitLab, Мониторинг, HLIT-VDI,
Экосистема ML/DL и JLabHPC, Ovirt.



Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



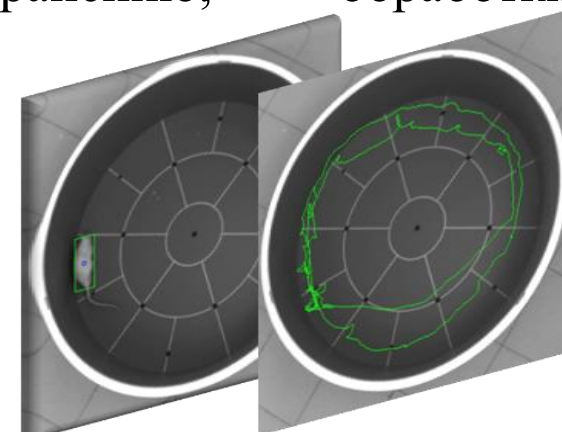
Совместный проект ЛИТ и ЛРБ направлен на создание информационной системы для анализа поведенческих и патоморфологических изменений в центральной нервной системе при исследовании воздействия ионизирующего излучения и других факторов на биологические объекты

Информационная система базируется на:

- алгоритмах компьютерного зрения на основе технологий машинного и глубокого обучения;
- современных ИТ-решениях к хранению, обработке и визуализации данных;

Информационная система позволит:

- ускорить и упростить работу с экспериментальными данными различным группам исследователей
- упростить и ускорить диагностику патологий ЦНС, и в частном случае, выработку эффективных методов профилактики и защиты от ионизирующего излучения.



Исследования проводятся с использованием всех возможностей Гетерогенной платформы HybriLIT ОИЯИ.





Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Цель - ускорить и упростить работу с экспериментальными данными

Разрабатываемая **информационная система (ИС)** позволит объединить и структурировать данные, полученные в проводимых в ЛРБ экспериментах, различных типов в единое информационное пространство, способное предоставить как удобство хранения и доступа к данным, так и набор передовых (актуальных) алгоритмических процедур автоматизации анализа данных для решения вопросов, связанных с диагностикой различных патологий ЦНС, которые остаются глобально открытыми и недостаточно исследованными, в том числе на текущем этапе развития медицины.



Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Разрабатываемая ИС будет реализована в виде веб-приложения и иметь клиент-серверную архитектуру. В роли сервера будет выступать Node.js, а клиент написан при помощи библиотеки React.

- Разрабатываемая ИС берёт на себя все необходимые манипуляции с базой данных (БД), файловыми хранилищами, Batch-системой и прочими компонентами платформы HybriLIT;
- ИС обеспечивает удобный интерфейс для: хранения, изменения и добавления данных эксперимента;

Плюсы веб-приложения.

- Не требует установки или обновления;
- Обеспечит доступ к ИС с разных платформ, где есть веб-браузер и Интернет (ПК, Смартфоны/Планшеты);
- Все расчеты будут происходить на удаленном выделенном сервере.



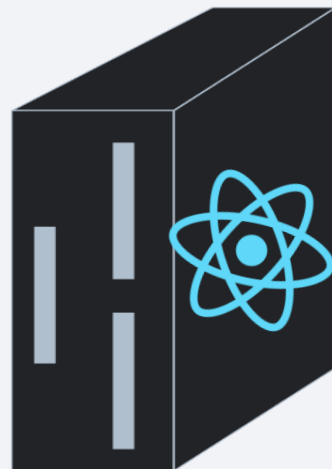
Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Клиенты



Frontend



Backend





Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



1

Ограниченный доступ

Пользователи получают доступ к сервису после аутентификации

2

Работа с данными экспериментов

Сохранение различных данных в БД, а также возможность их изменения и удаления

3

Загрузка и хранение файлов

Сохранение фото и видеоматериалов с экспериментов в файловой системе сервера

4

Представление результатов

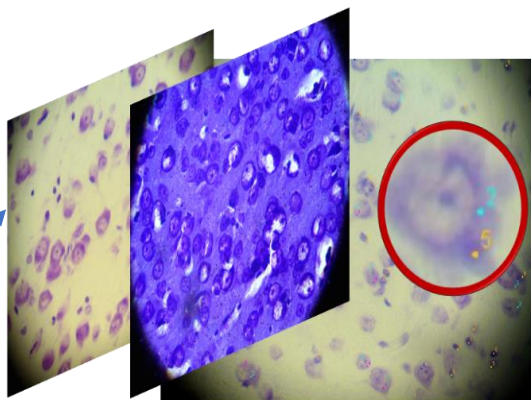
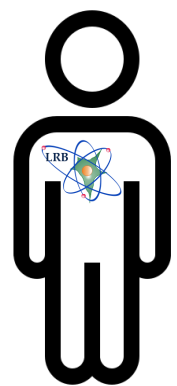
Статистика, диаграммы и графики



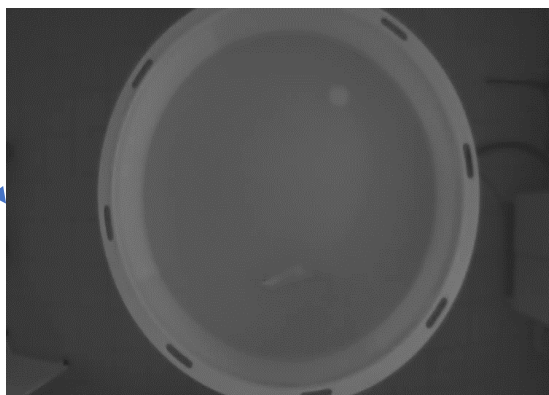
Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



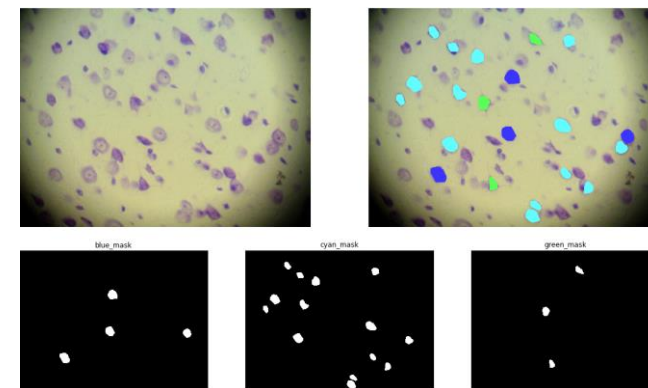
Концептуальная схема работы сервиса



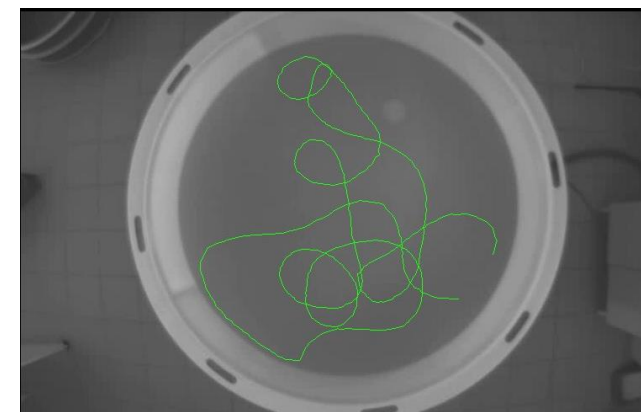
Данные
экспериментов



Анализ



Результат

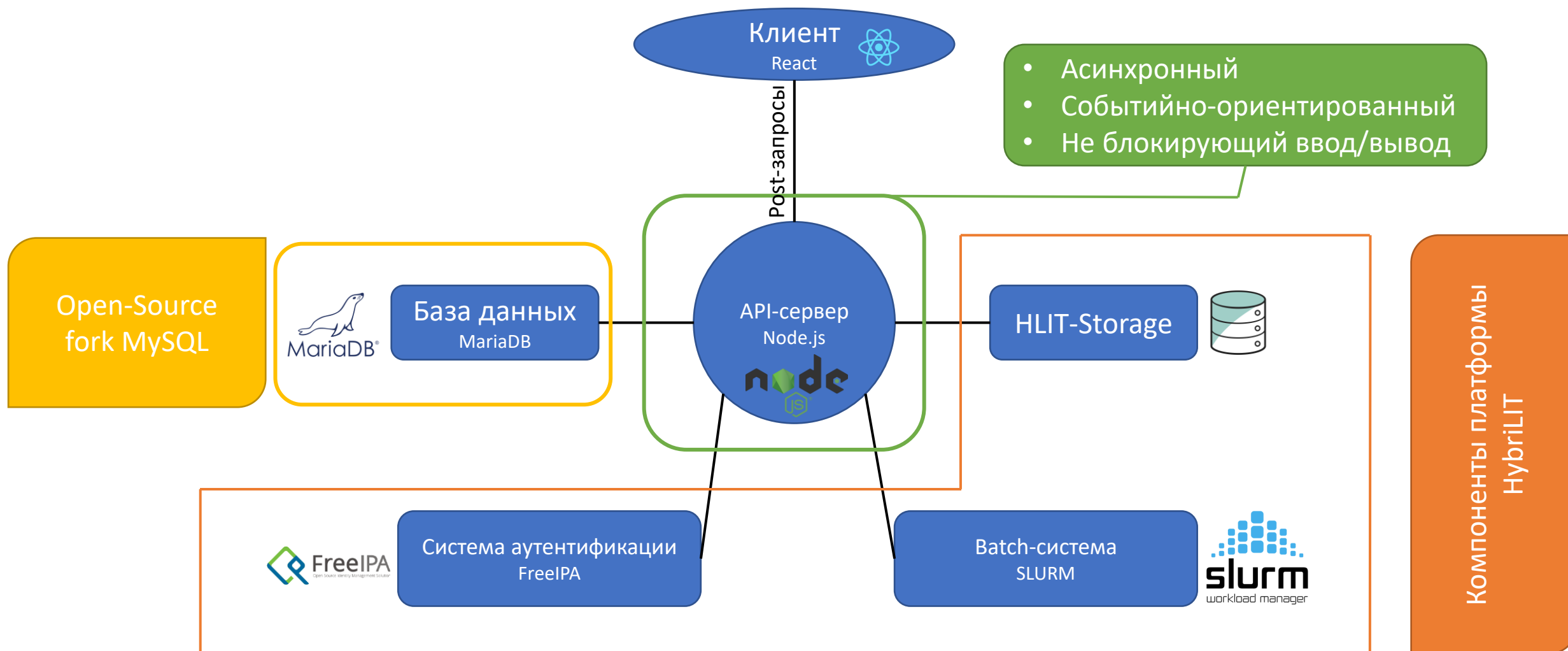




Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Общая схема проекта

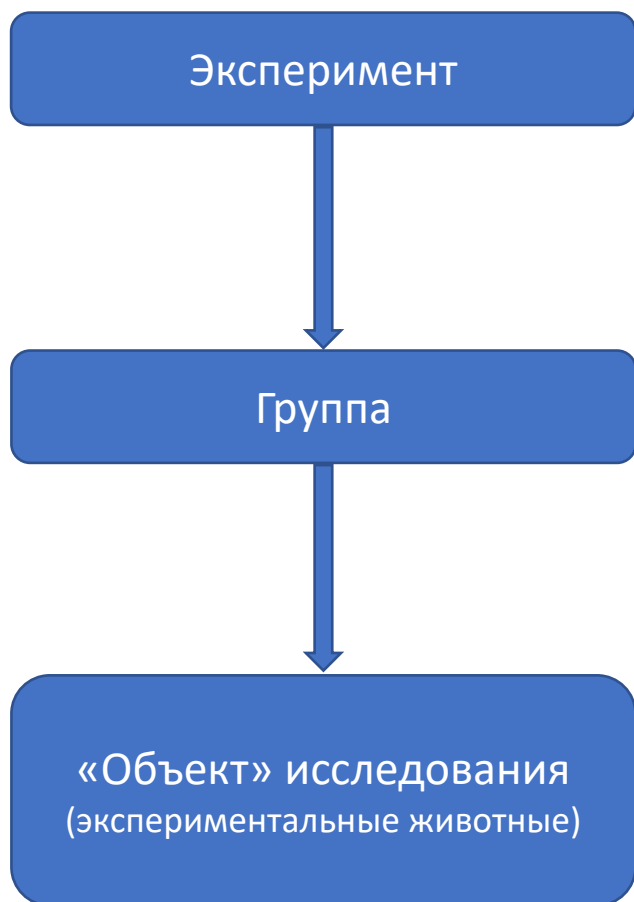




Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Иерархическая структура данных ИС



Главный объект ИС внутри которого будет храниться все содержимое эксперимента. Имеет стандартный набор атрибутов: название, имя создателя, дата создания, описание; и особый атрибут – воздействие (пример гамма-излучение).

Вспомогательный объект ИС связанный с определенным **экспериментом**, внутри которого будут храниться данные об «объектах» исследования (экспериментальные животные). У группы помимо стандартных атрибутов существуют особые: исследуемый орган, краситель, увеличение микроскопа, препарат.

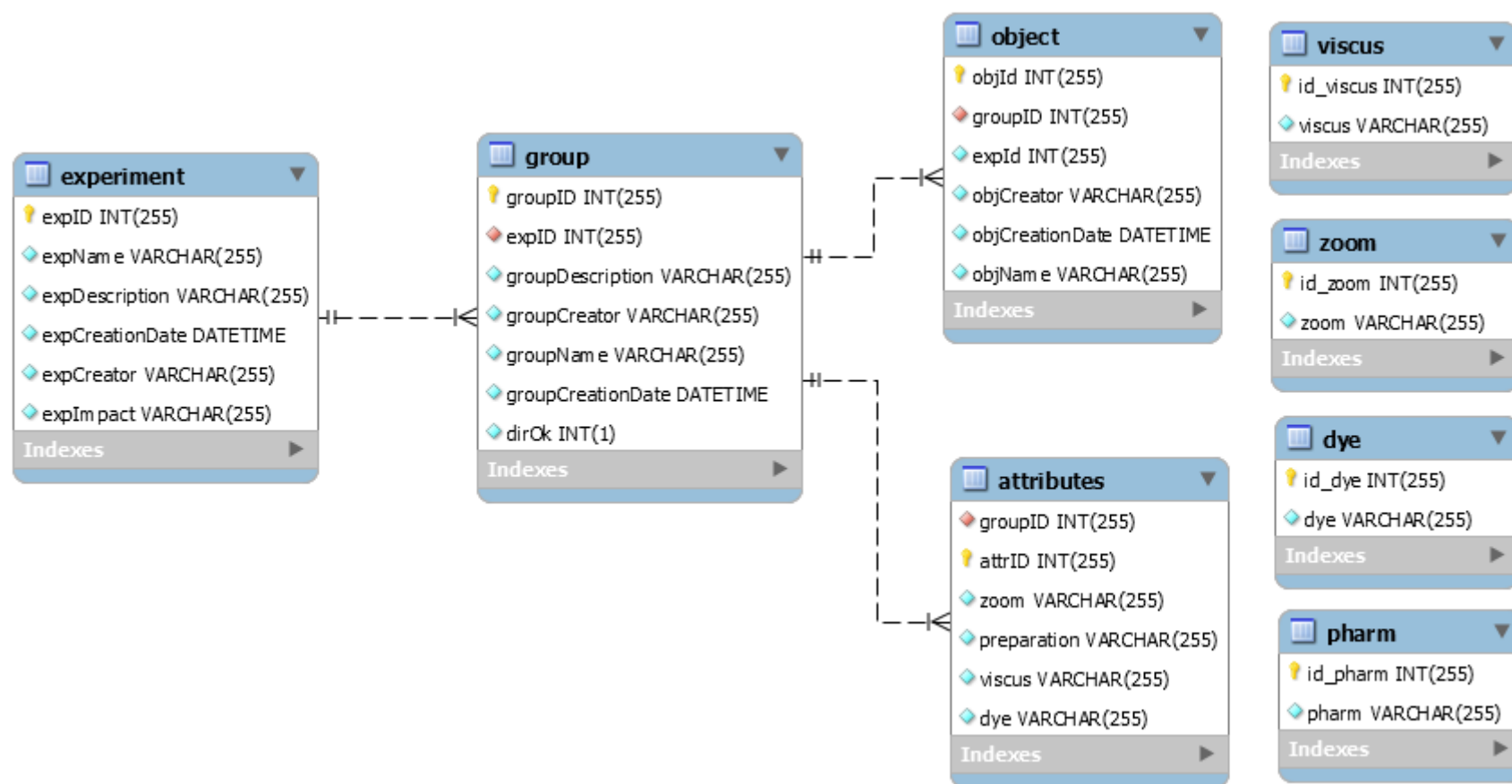
Вспомогательная объект ИС связанный с определенной **группой**, внутри будут храниться мета-данные «объектов» эксперимента (фотографий и видеофайлов)



Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Схема базы данных

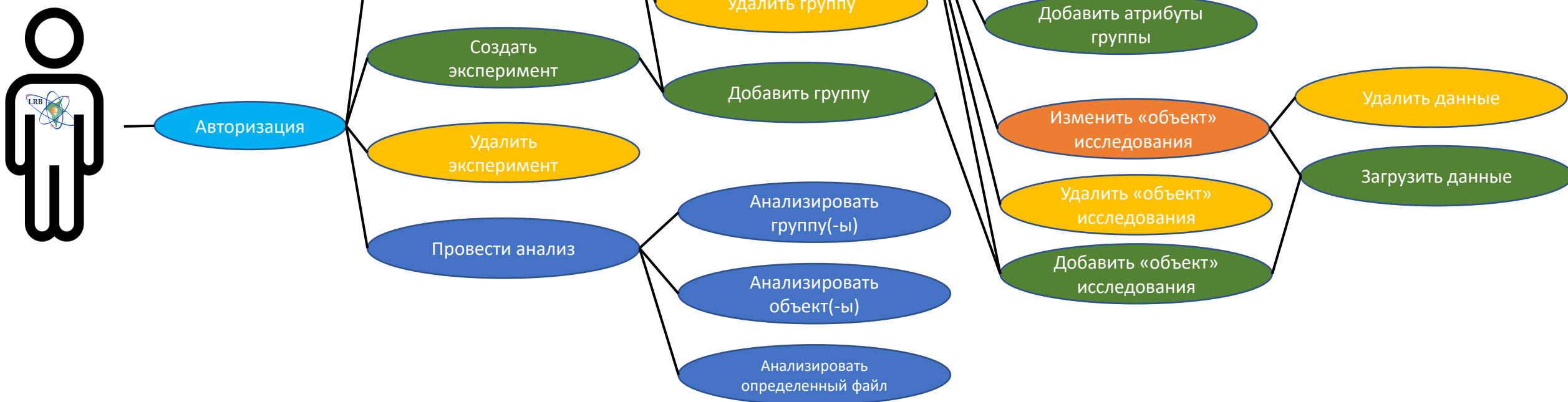




Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Основной сценарий использования ИС

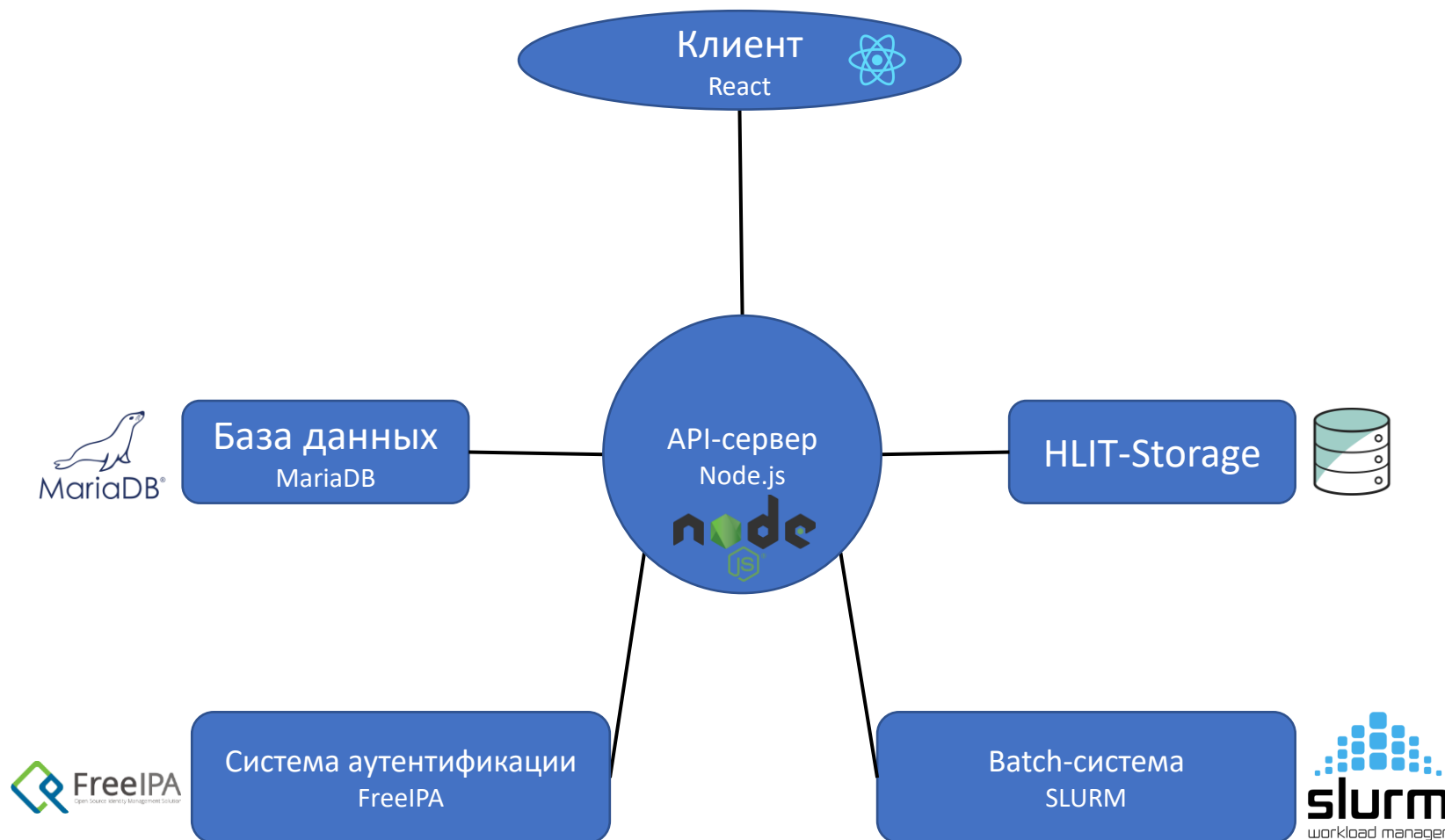




Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Пример работы API-сервера. Авторизация

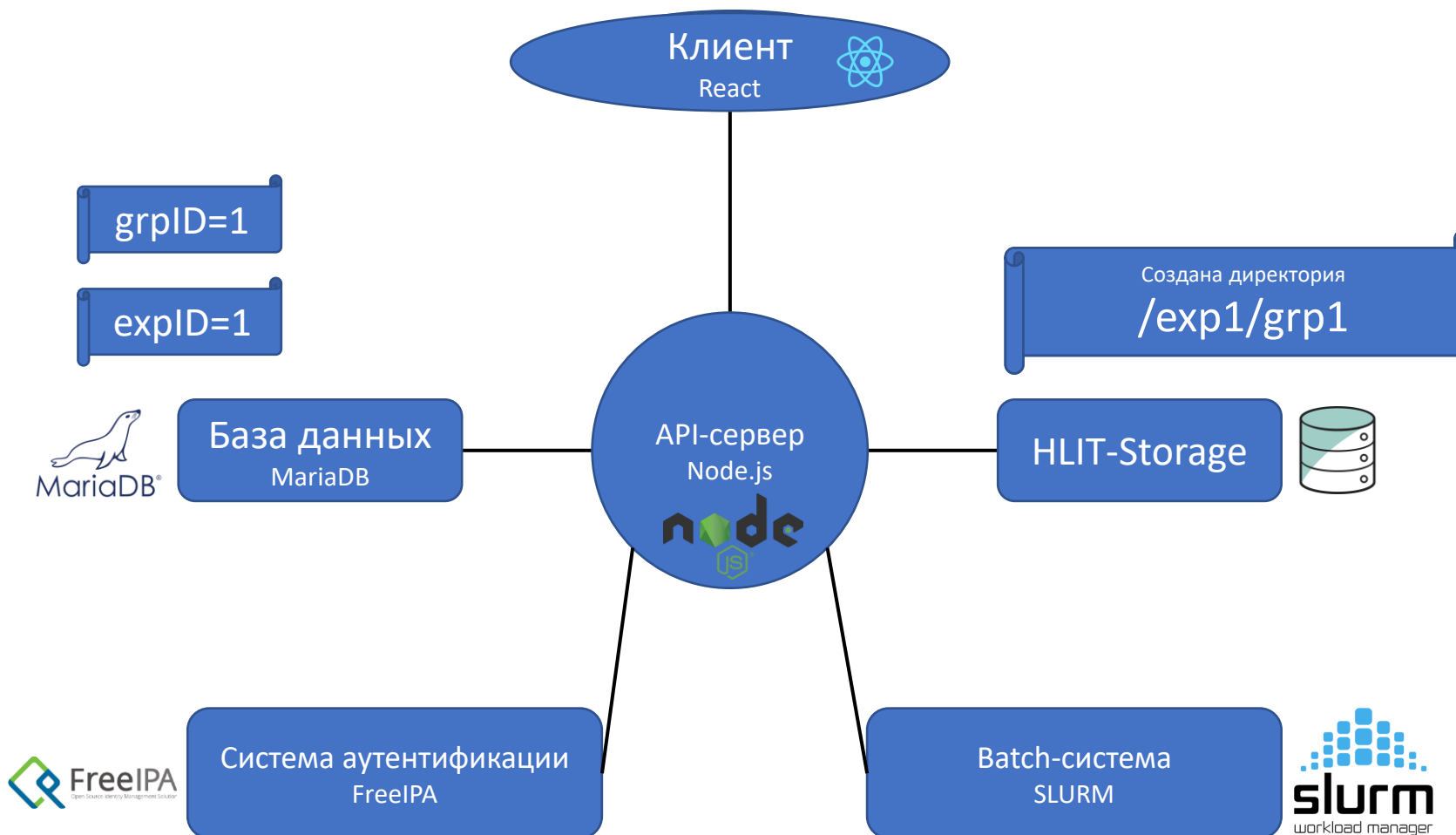




Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Пример работы API-сервера. Добавление эксперимента/группы





Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Заключение

Разработан прототип серверной части ИС, который взаимодействует с:

- Базой данных MariaDB;
- Системой по управлению идентификацией FreeIPA;
- Системой хранения HybriLIT;
- SLURM workload manager.

Разработанный прототип работает на выделенной виртуальной машине и на его основе ведется разработка клиентской части ИС на библиотеке React.



Информационная система для задач радиационной биологии (Совместный проект ЛИТ и ЛРБ)



Спасибо за внимание!