



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

## Цифровой блок НИУ ВШЭ

# О СУПЕРКОМПЬЮТЕРНОМ КОМПЛЕКСЕ CHARISMA – ЗАГРУЗКА, ОБСЛУЖИВАНИЕ, РАСШИРЕНИЕ, О ВЫПОЛНЯЕМЫХ ИССЛЕДОВАНИЯХ, О МОНИТОРИНГЕ ЭФФЕКТИВНОСТИ ЗАДАЧ ПОЛЬЗОВАТЕЛЕЙ

Начальник отдела суперкомпьютерного моделирования:  
Костенецкий Павел Сергеевич, к.ф.-м.н., доцент.

Конференция пользователей суперкомпьютерного комплекса НИУ ВШЭ  
Москва, 05.10.2021.



# ХАРАКТЕРИСТИКИ СУПЕРКОМПЬЮТЕРА «сHARISMA»

## (COMPUTER OF HSE FOR ARTIFICIAL INTELLIGENCE AND SUPERCOMPUTER MODELLING)

- 6 место в ТОП 50 СНГ (upgrade)
- Пиковая производительность: 2 Петафлопс (2 квадриллиона операций в секунду над числами с двойной точностью).
- LINPACK-производительность: 927.4 Терафлопс.
- 46 вычислительных узлов
  - 6 узлов с 1 ТБ ОЗУ, 8 GPU A100 80 ГБ,
  - 10 узлов с 1,5 ТБ ОЗУ, 4 GPU V100 32 ГБ,
  - 19 узлов с 768 ГБ ОЗУ, 4 GPU V100 32 ГБ,
  - 11 узлов с 384 ГБ ОЗУ для расчётов на CPU.
- 2 управляющих узла,
- 1 сервер резервного копирования 182ТБ
- 48 графических процессоров NVIDIA Tesla A100 80 ГБ
- 116 графических процессора NVIDIA Tesla V100 32 ГБ
- 2584 ядер центральных процессоров
- Оперативная память: 40,3 ТБ
- Дисковая память: 968 ТБ
  - параллельная СХД на базе Lustre: 840 ТБ
- Коммуникационная сеть: 2 x InfiniBand EDR (2x100 Гбит/с, топология Fat Tree)





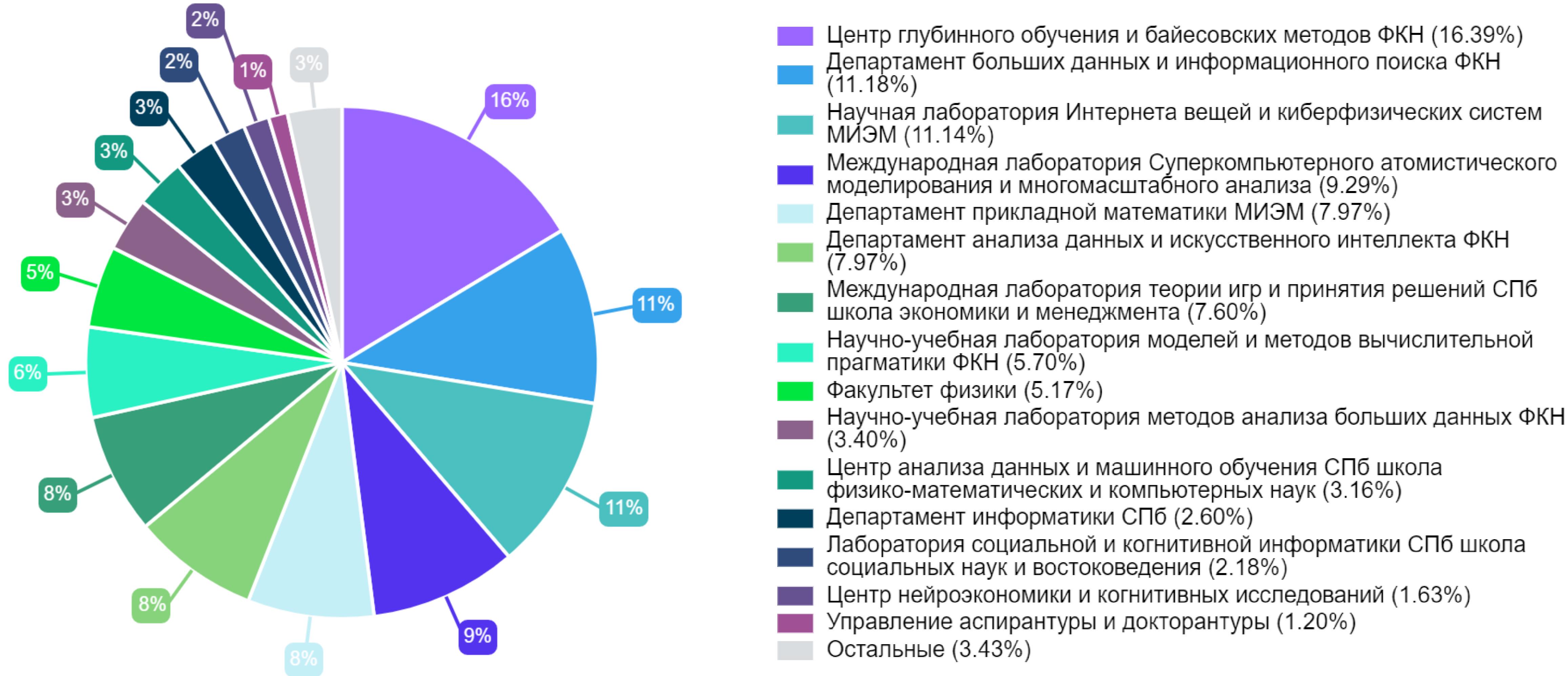
# МОЩНЕЙШИЕ КОМПЬЮТЕРЫ СНГ



№	Место установки / Название	Linpack (Тфлоп/с)
1	SberCloud / "Кристофари"	6669
2	МГУ / "Ломоносов-2"	2478
3	new МТС / "GROM"	2258
4	▽ ФГБУ "ГВЦ Росгидромета", Москва	1200
5	upgrade СПбПУ, Санкт-Петербург / "Политехник - РСК Торнадо"	971↑(910)
6	upgrade НИУ Высшая школа экономики / "сHARISMa"	927.4↑(653)
7	upgrade МСЦ РАН / "МВС-10П ОП2"	759↑(572)
8	▽ НИЦ "Курчатовский институт"	755
9	▽ Сколтех / "ZHORES CDISE Cluster"	496
10	▽ Компьютерные Экосистемы / "PetaNode 1.2 Cluster"	420
11	▽ Тинькофф Банк / "Колмогоров"	419
12	▽ МСЦ РАН / "МВС-10П"	383
13	▽ ОИЯИ, Дубна / "Имени Н.Н. Говоруна сегмент SKYLAKE"	313
14	▽ ННГУ, Нижний Новгород / "Лобачевский"	290
15	▽ ЮУрГУ, Челябинск / "РСК Торнадо ЮУрГУ"	288



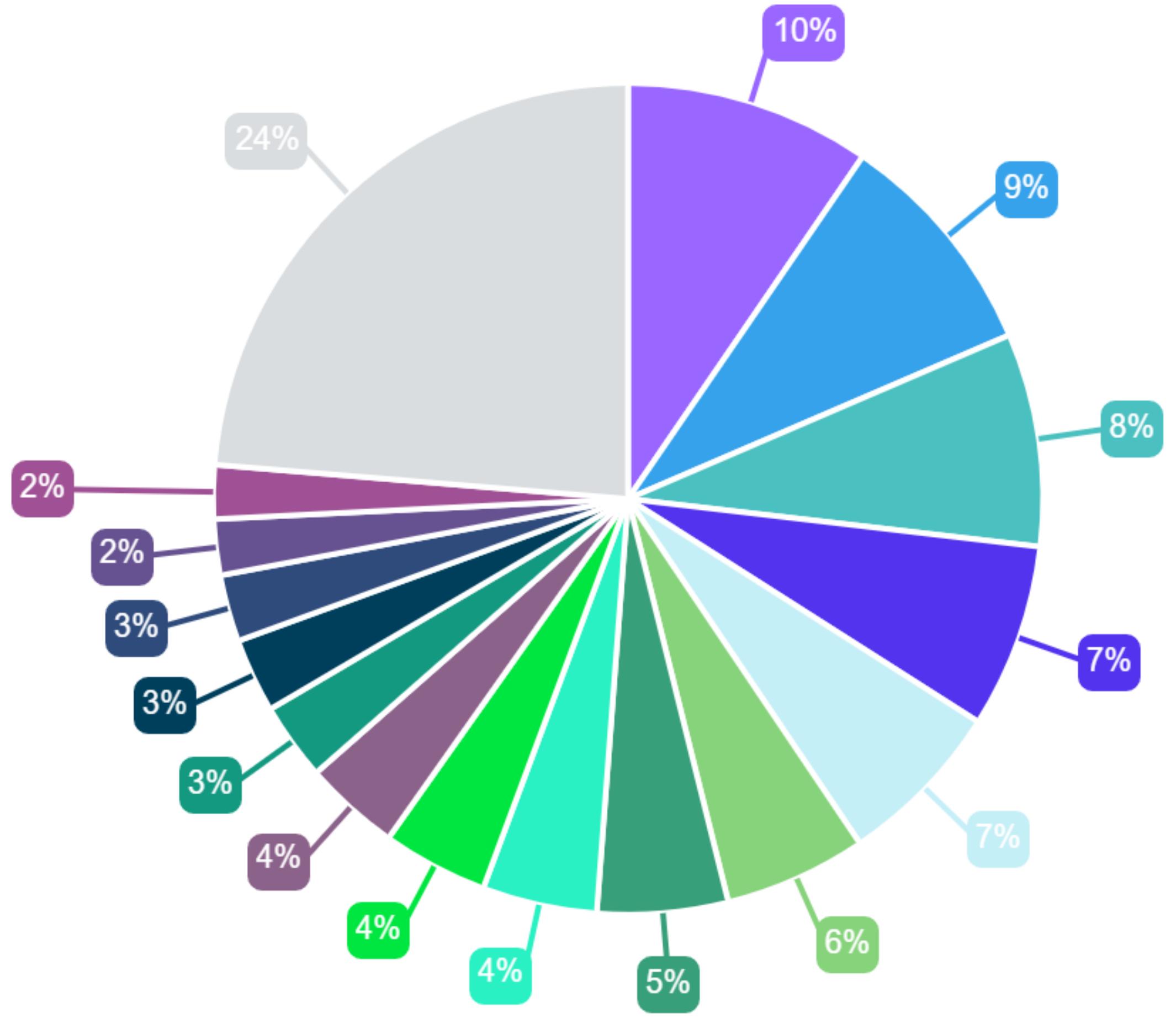
# ИСПОЛЬЗОВАНИЕ МАШИННОГО ВРЕМЕНИ ПОДРАЗДЕЛЕНИЯМИ ЗА 1 ГОД (ТОР-15)



Период: октябрь 2020 - октябрь 2021 г.

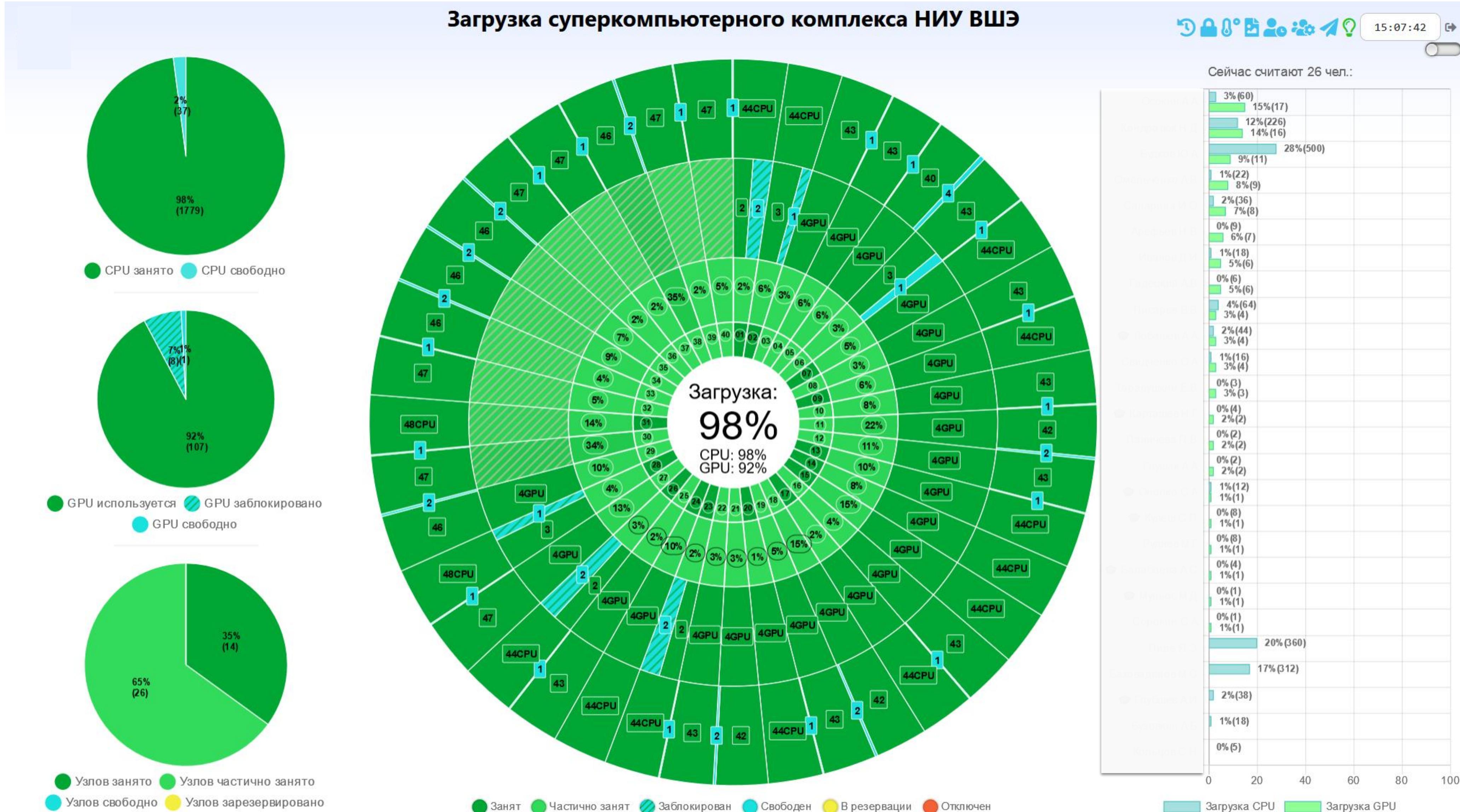


# ИСПОЛЬЗОВАНИЕ МАШИННОГО ВРЕМЕНИ ПРОЕКТАМИ ЗА 1 ГОД (ТОР-15)



- Методы глубинного обучения для семантического парсинга текста и синтеза программ - Рук.: Осокин А.А. (9.57%)
- Мульти尺度ное моделирование ион-молекулярных систем - Рук.: Будков Ю.А. (8.97%)
- Изучение оценок неопределенности глубоких нейронных сетей - Рук.: Ашуха А.П. (8.28%)
- Improving Upside Down Reinforcement Learning - Рук.: Иванов Д.И. (7.17%)
- Адаптация архитектуры нейронной сети Трансформер для анализа исходных кодов программ - Рук.: Чиркова Н.А. (6.54%)
- Параллельные алгоритмы и атомистическое многомасштабное моделирование материалов - Рук.: Стегайлов В.В. (5.55%)
- Автоматическая обработка текстов - Рук.: Артемова Е.Л. (5.09%)
- Изучение явления многочастичной локализации в одномерных цепочках методом точной диагонализации - Рук.: Баховадинов М.С. (4.47%)
- Моделирование случайных блужданий с объемным и поверхностным подкреплением - Рук.: Тамм М.В. (4.12%)
- Атомистическое и континуальное моделирование в задачах физики - Рук.: Писарев В.В. (3.74%)
- Языковое моделирование текстов онлайн социальных сетей - Рук.: Карпов И.А. (3.01%)
- Экзотические фазы сильно коррелированных фермионов в низких размерностях - Рук.: Буровский Е.А. (2.89%)
- Методы предобработки текстовых данных для обучения языковых моделей - Рук.: Омельченко А.В. (2.60%)
- Машинное обучение и статистическая физика - Рук.: Щур Л.Н. (2.17%)
- Машинное обучение на графах - Рук.: Игнатов Д.И. (2.09%)
- Остальные (23.67%)

# СОБСТВЕННАЯ СИСТЕМА МОНИТОРИНГА



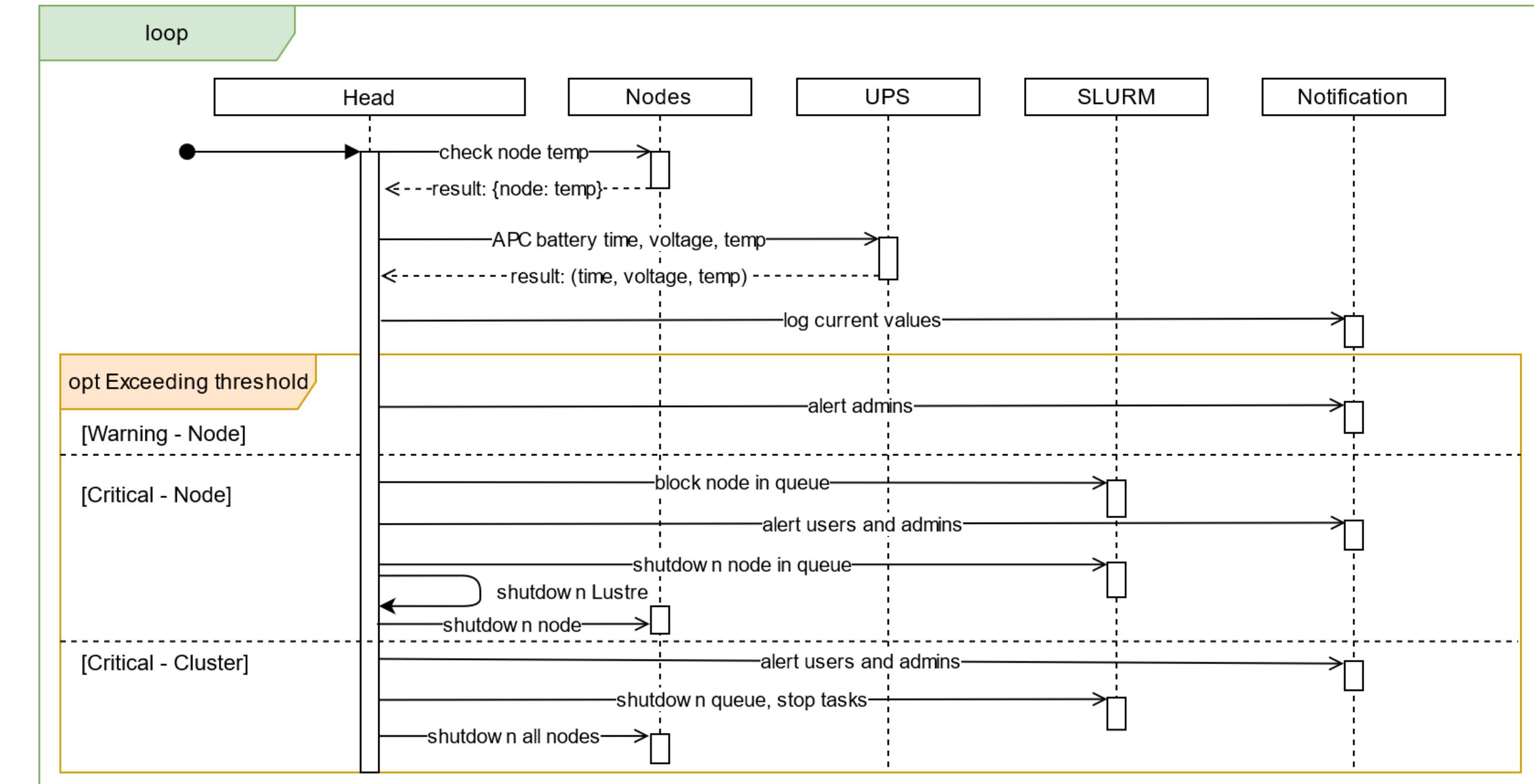
\*При многопользовательском режиме работы, 100% загрузка недостижима. Как видно на диаграмме, всегда есть фрагменты вычислительных ресурсов, которые недостаточны для запуска следующей задачи в очереди.



# СИСТЕМА АВАРИЙНОГО ОТКЛЮЧЕНИЯ СУПЕРКОМПЬЮТЕРА

События, требующие незамедлительного реагирования:

- Выход из строя системы охлаждения вычислительного узла,
- Выход из строя системы кондиционирования воздуха в помещении,
- Переключение ИБП на режим работы от аккумуляторных батарей,
- Низкий заряд аккумуляторных батарей ИБП.



Система сделана на базе простых скриптов.



# РЕГИСТРАЦИЯ НА СУПЕРКОМПЬЮТЕРЕ ЧЕРЕЗ ЕЛК

Вышка Digital | Smart Point Сервисы / Научные сервисы /

Научные сервисы

Поиск по сервисам

- Антиплагиат-НИР**  
Отправить отчет по научно-исследовательским работам для проверки на наличие неоригинального текста (технического заимствования), заимствования и/или необоснованного самоцитирования
- Конкурсы научного фонда**  
Конкурсы научного фонда
- Мониторинг конкурсов**  
Подписка на рассылку мониторинга объявленных конкурсов
- Суперкомпьютер** ★  
Система управления доступом к Суперкомпьютерному комплексу НИУ ВШЭ



Быстрая регистрация проектов и пользователей на суперкомпьютере через **Единый личный кабинет НИУ ВШЭ**.



# ГЕНЕРАТОР ПАКЕТНЫХ ФАЙЛОВ ДЛЯ СУПЕРКОМПЬЮТЕРА CHARISMA

## ПАРАМЕТРЫ ЗАПУСКА

ИНФОРМАЦИЯ О ЗАДАЧЕ

JOB INFORMATION

ВЫЧИСЛИТЕЛЬНЫЕ РЕСУРСЫ

COMPUTING RESOURCES

РЕЖИМ МАССИВА

JOB ARRAY

ПОДКЛЮЧИТЬ МОДУЛИ

LOAD MODULES

ВРЕМЯ РАБОТЫ

JOB DURATION

ВЫВОД

OUTPUT

## SBATCH СКРИПТ ЗАПУСКА

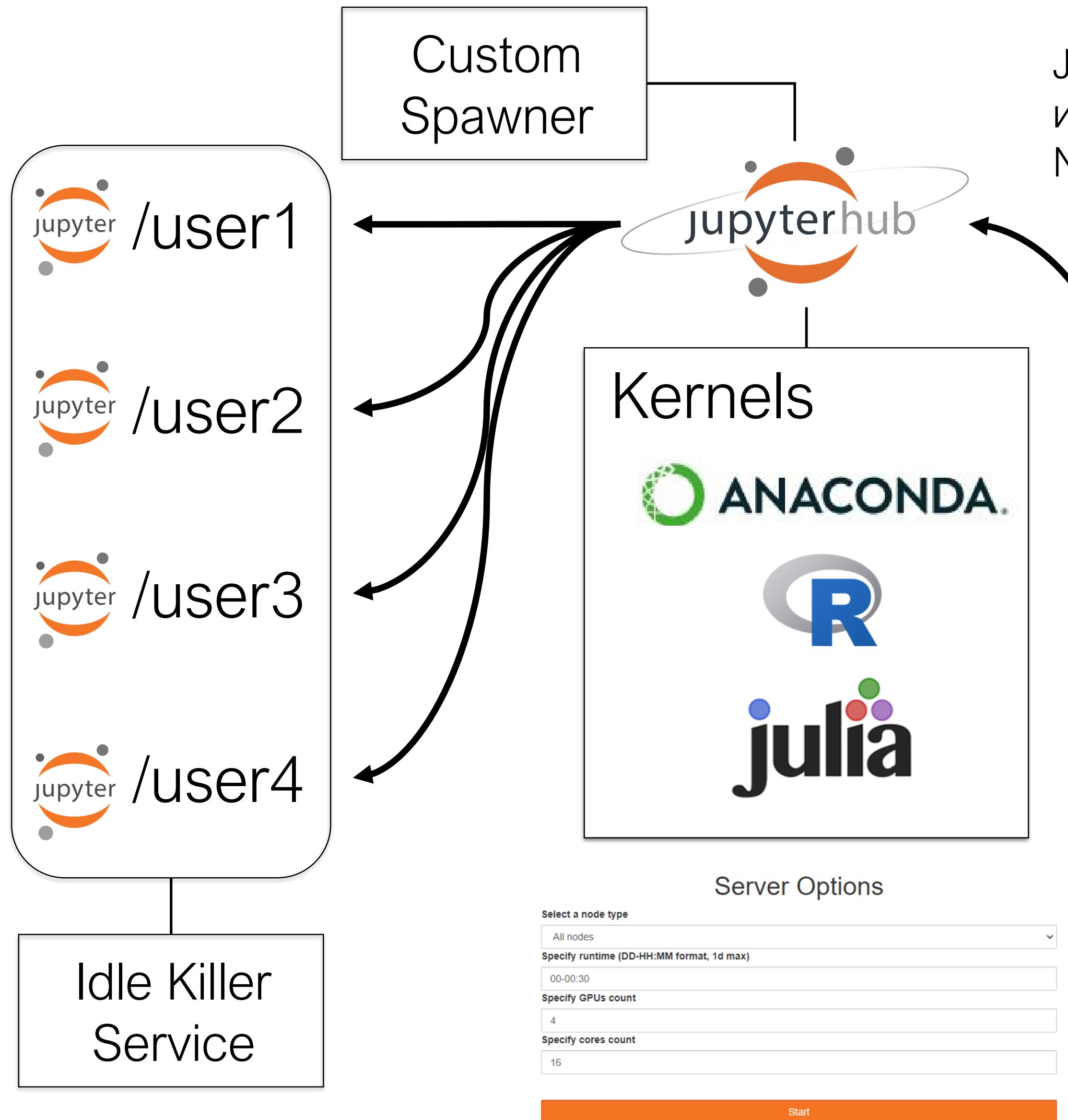
```
#!/bin/bash
#SBATCH --time=1-01:00
#SBATCH --gres=gpu:2
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --cpus-per-task=2
#SBATCH --gpus-per-node=2
#SBATCH --ntasks=1
```

COPY

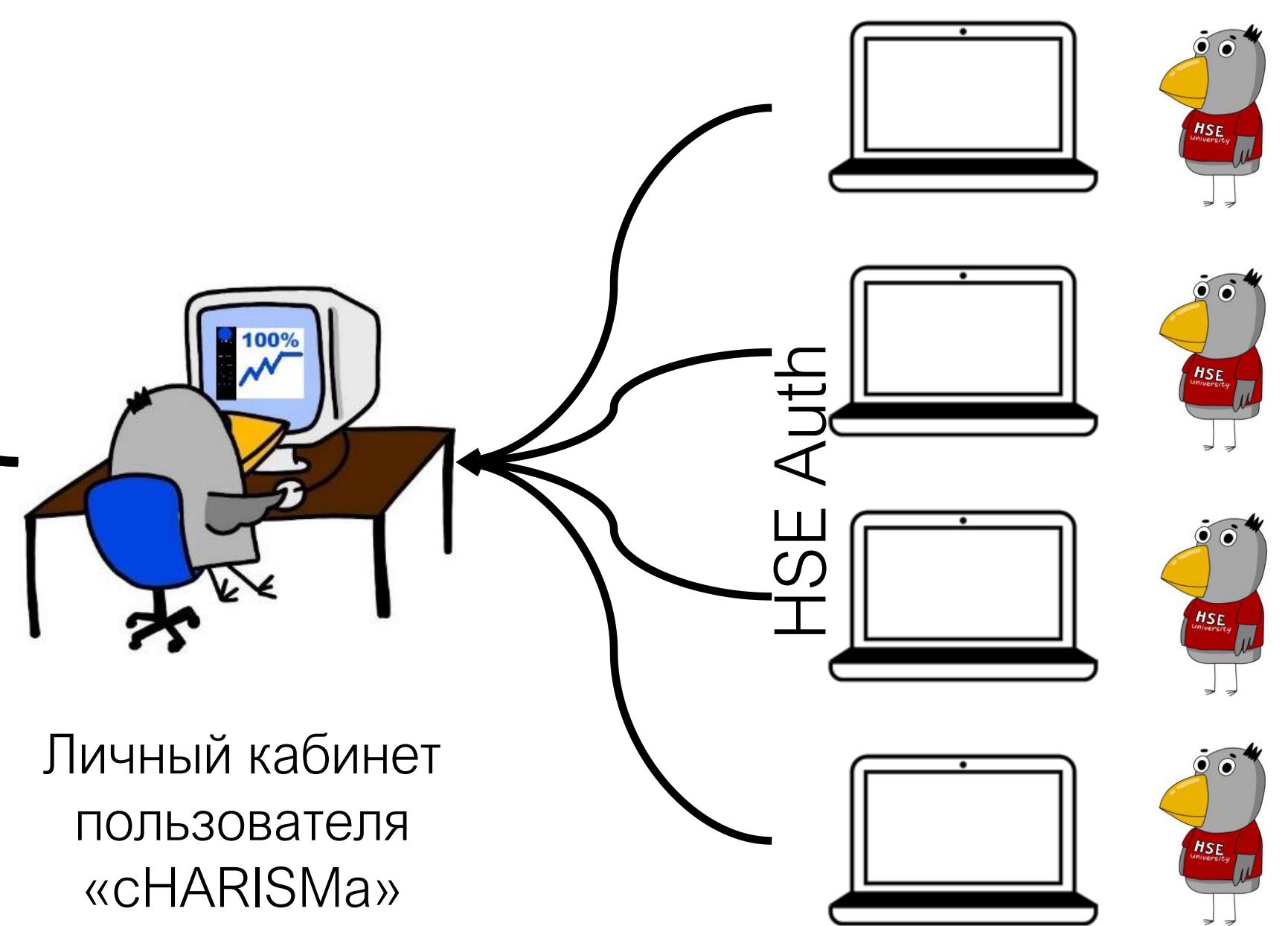
<https://lk.hpc.hse.ru/sbatch>



# JUPYTER HUB



Jupyter Notebook – функциональный и популярный инструмент для выполнения расчётов (Data Science, Neural Networks) и визуализации полученных результатов





# ЛИЧНЫЙ КАБИНЕТ ПОЛЬЗОВАТЕЛЯ СУПЕРКОМПЬЮТЕРА

Сводка данных

Личный кабинет пользователя суперкомпьютерного комплекса

Сводка данных

Логин-сервер: ОК

Очередь задач: ОК

Файловое хранилище: ОК

Задачи пользователя  
Завершено: 2  
Выполняется: 0  
[Статистика](#)

Задачи в очереди SLURM  
Выполняется: 197  
Ожидает: 23  
Запрошено: 184 CPU, 46 GPU

Срок действия аккаунта  
до 31.12.2099  
[Управление проектами](#)

Квоты пользователя на вычислительные ресурсы

Использование CPU  
0.0%  
Используется 0 ядер CPU из 300

Использование GPU  
0.0%  
Используется 0 GPU из 20

Использование СХД  
0.0%  
Занято 15.66 MB из 1.33 TB

[lk.hpc.hse.ru](http://lk.hpc.hse.ru)

© Отдел суперкомпьютерного моделирования



# HPC TASKMASTER – СИСТЕМА ДЛЯ ОБНАРУЖЕНИЯ НЕЭФФЕКТИВНЫХ И НЕКОРРЕКТНО ЗАПУЩЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ ЗАДАЧ

---

1. Шамсутдинов А.Б., Костенецкий П.С. Разработка системы мониторинга эффективности задач на суперкомпьютере сCHARISMa // Параллельные вычислительные технологии ПаВТ'2021, 30 марта - 1 апреля 2021, г. Волгоград
2. Костенецкий П.С., Шамсутдинов А.Б., Чулкевич Р.А., Козырев В.И. HPC TaskMaster – система мониторинга эффективности задач суперкомпьютера // Суперкомпьютерные дни в России: труды международной конференции (27-28 сентября 2021 г., г. Москва). Москва: Издательство МГУ, 2021. В печати.

## Аналоги

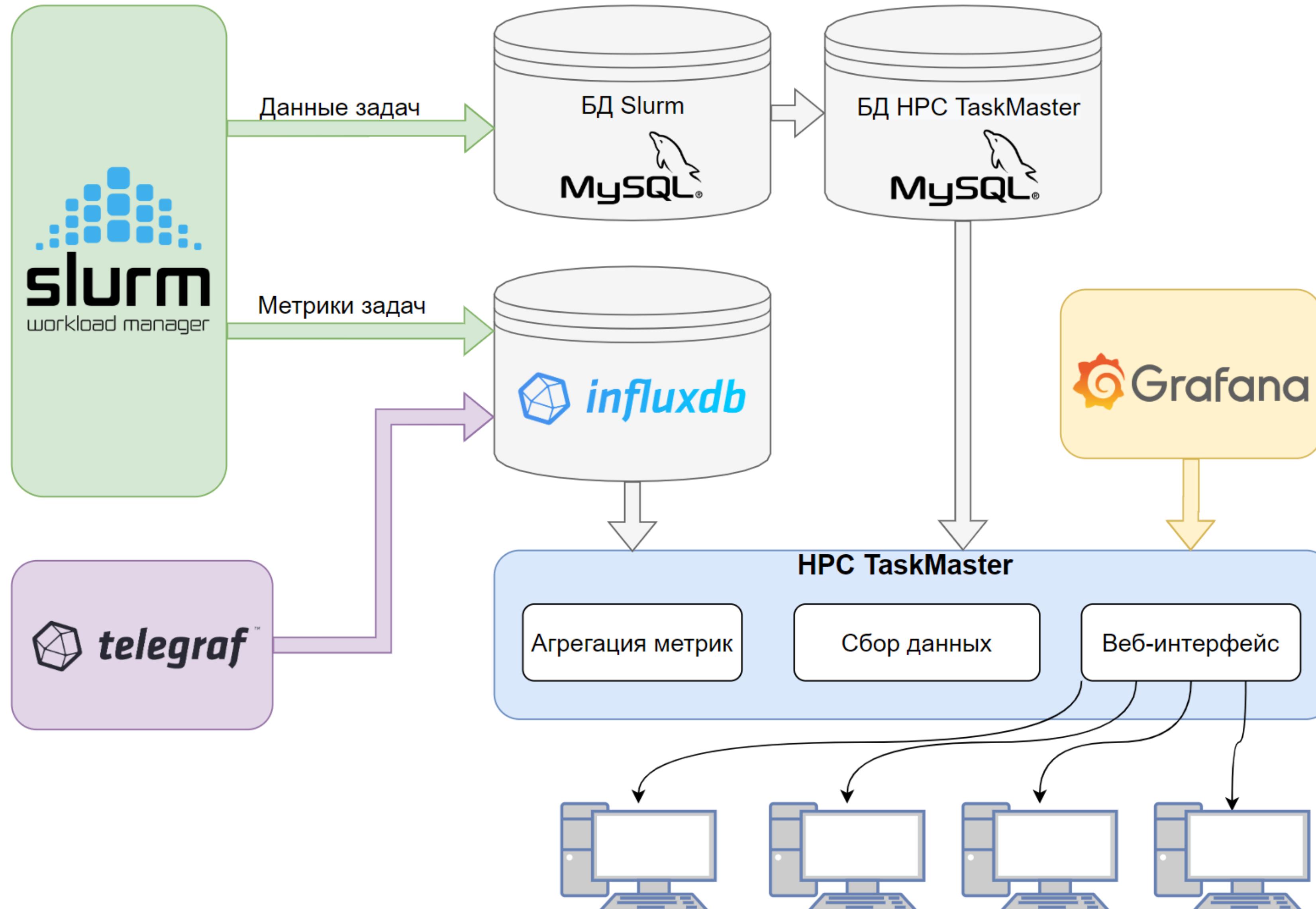
1. Chan N. A Resource Utilization Analytics Platform using Grafana and Telegraf for the Savio Supercluster. In ACM Int. conf. proc. series. 2019 <https://doi.org/10.1145/3332186.3333053>
2. Nikitenko D. et al. JobDigest - Detailed System Monitoring-Based Supercomputer Application Behavior Analysis // Communications in Computer and Information Science. Springer Verlag. 2017. Vol. 793. P. 516–529. DOI:10.1007/978-3-319-71255-0\_42

Система позволит экономить до 30% вычислительных ресурсов.

Система интегрирована в личный кабинет пользователя суперкомпьютера.



# ПРИНЦИП РАБОТЫ НРС TASKMASTER



Открытый исходный код

Система собирает  
информацию о задачах,  
а не об узлах кластера

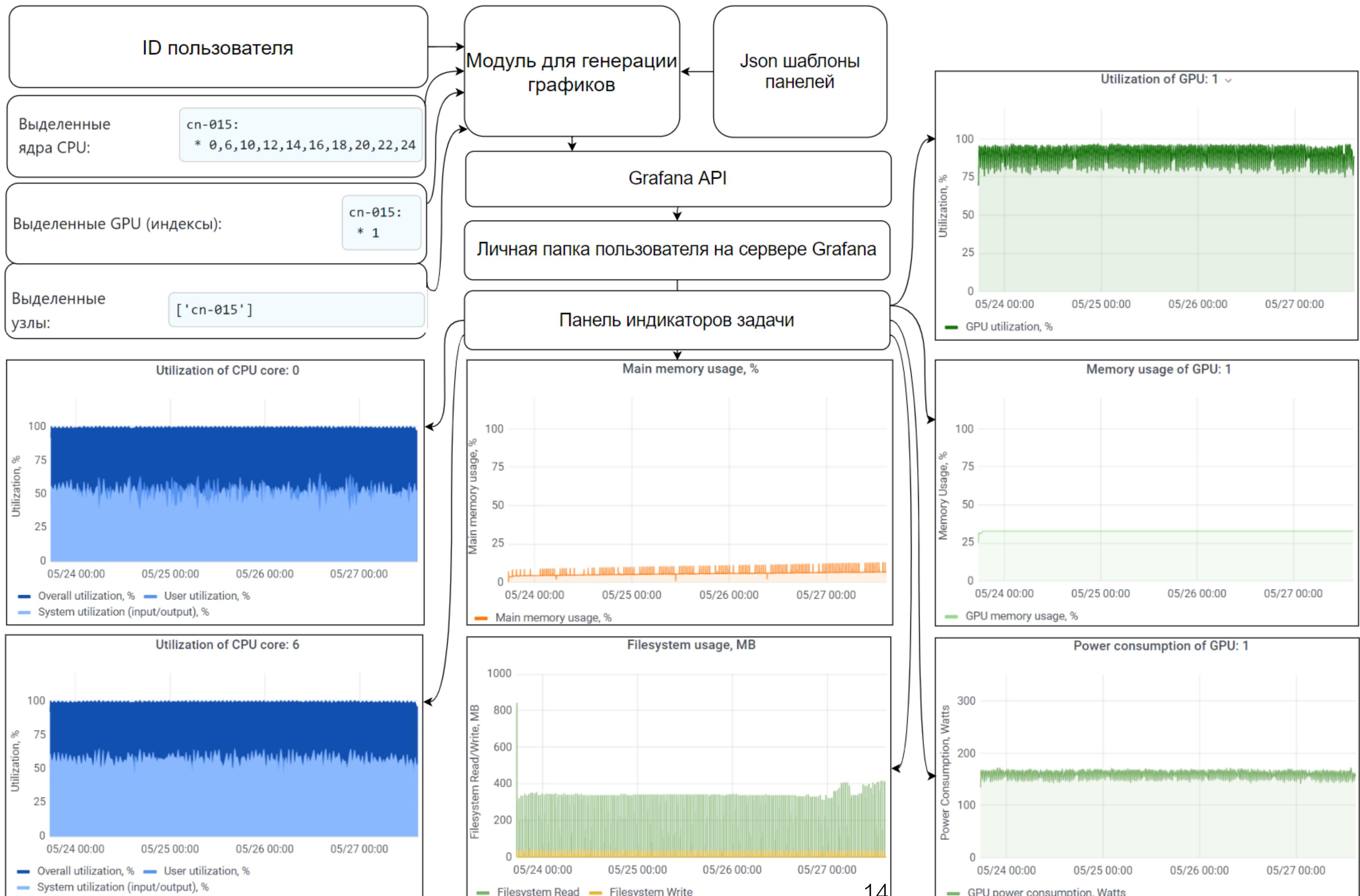
Метрики задач  
автоматически  
анализируются на  
наличие проблем

Для каждой задачи  
формируется вывод

Строятся интерактивные  
графики при помощи  
Grafana



# ВИЗУАЛИЗАЦИЯ ВЫПОЛНЕНИЯ ЗАДАЧИ В НРС TASKMASTER



Для создания графиков собираются ID ядер CPU и GPU, выделенных конкретной задаче

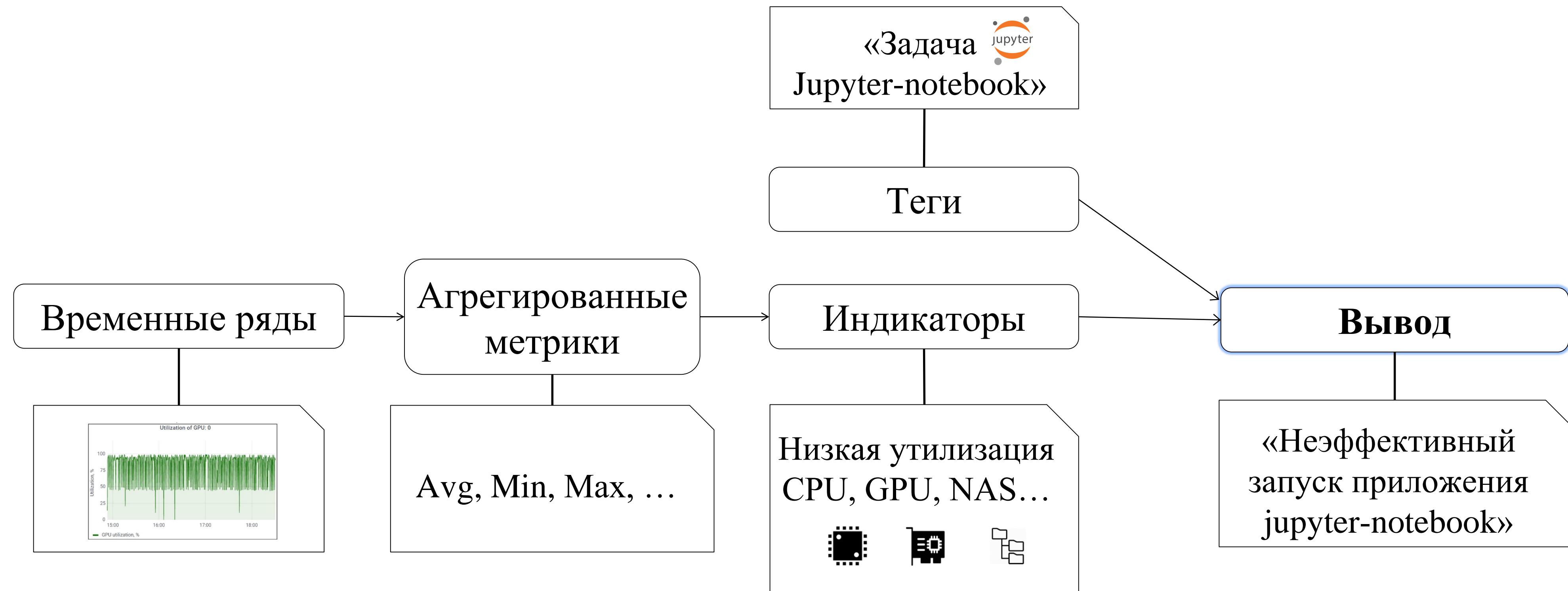
Также, собирается список вычислительных узлов, на которых запущена задача

Модуль генерирует json файл для Grafana из шаблонов и загружает его на сервер при помощи API

Для каждого пользователя создается личная папка на сервере Grafana

Отображение графиков в личном кабинете происходит при помощи технологии iframe

# ЭТАПЫ ОБРАБОТКИ ДАННЫХ В НРС TASKMASTER



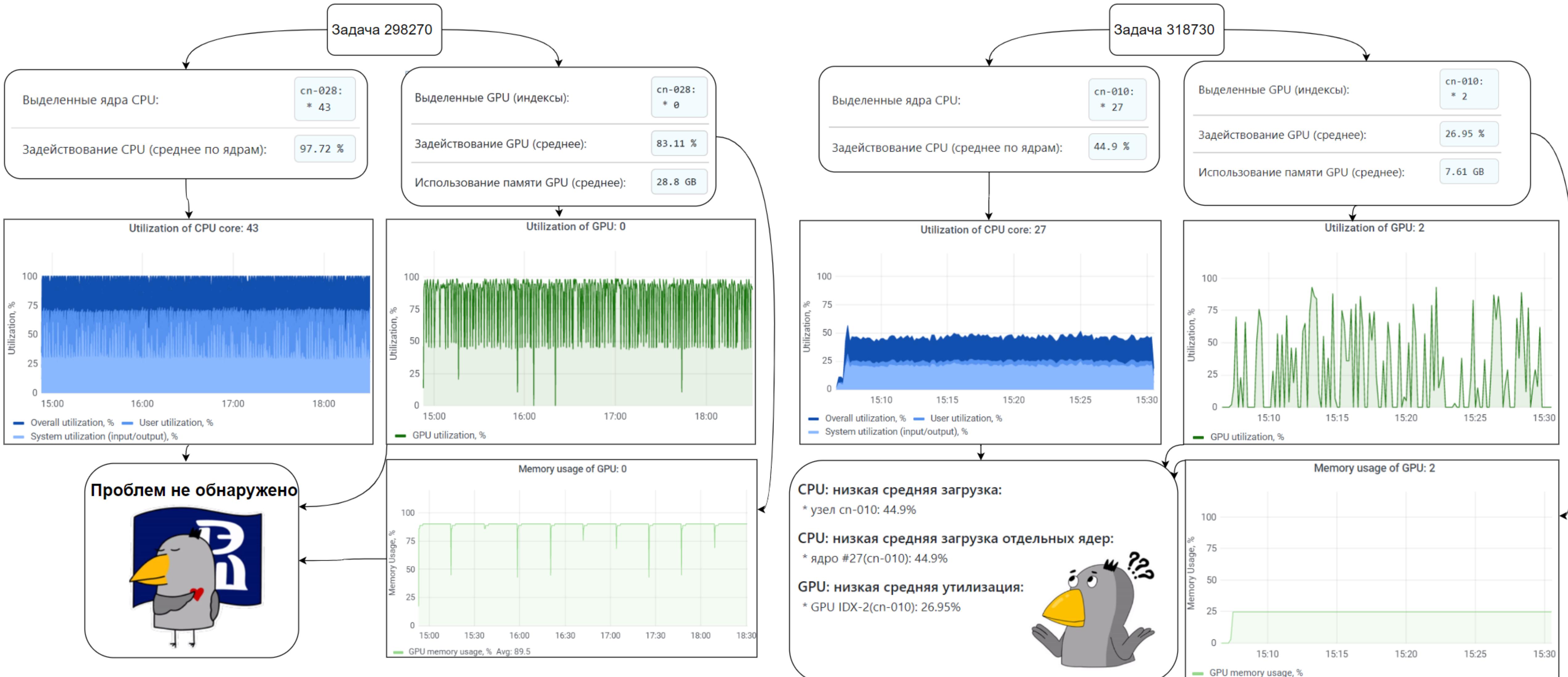
- На входе есть N временных рядов (результаты мониторинга выполнения задач на кластере).
- По ним вычисляются метрики (средние значения, максимумы и минимумы и т.д.)
- Для каждой задачи получается вектор метрик
- Вектор метрик обрабатывается набором функций, каждая из которых выдает один индикатор с указанием его веса (от 0 до 1)
- Собираем все результаты функций в вектор индикаторов для задачи (например, «Малое использование памяти»).
- Далее вектор показателей поступает на вход булевым функциям. На выходе получаются выводы о задаче (результат анализа, например: «Непараллельная задача запущена на нескольких ядрах»)





# ИНДИКАТОРЫ ЗАДАЧИ В НРС TASKMASTER

Примеры индикаторов: низкая средняя утилизация ядер CPU и GPU, низкая утилизация отдельных CPU, низкое использование видеопамяти и т.д.



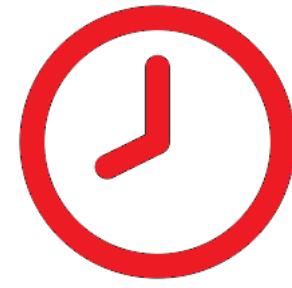


# ТЕГИ ЗАДАЧИ В НРС TASKMASTER

## Временные теги



Задача завершилась очень быстро

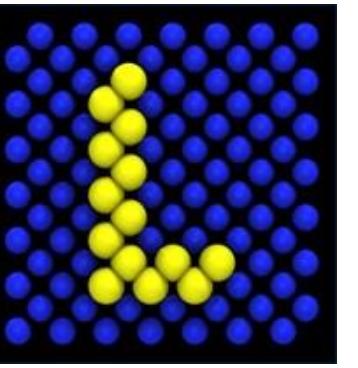


Задача работает аномально долго

## Теги типов задач



Jupyter Notebook



LAMMPS



VASP

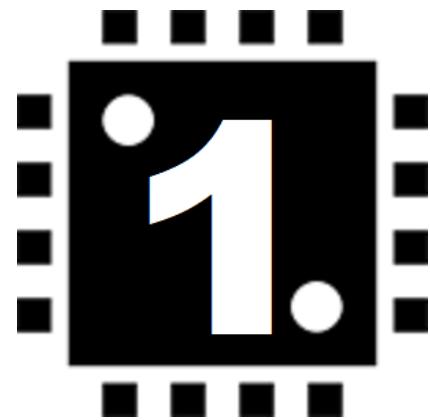


GROMACS

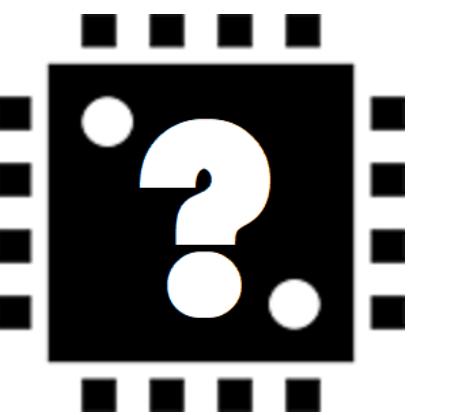
## Прочие теги



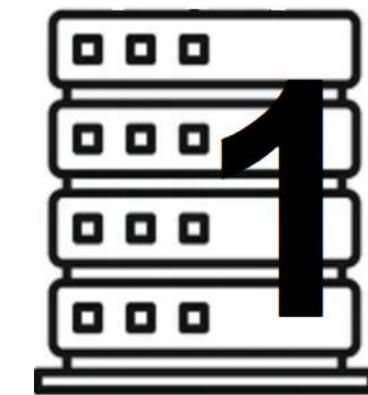
Задача завершена с ошибкой



Задача запущена на одном ядре



Выделено меньше ядер CPU, чем выделено GPU



Задача запущена на одном узле

Теги задачи – метки, передающие одно из свойств задачи

Тегами можно обозначить такие свойства, как длительность, тип задачи, обнаруженные ошибки

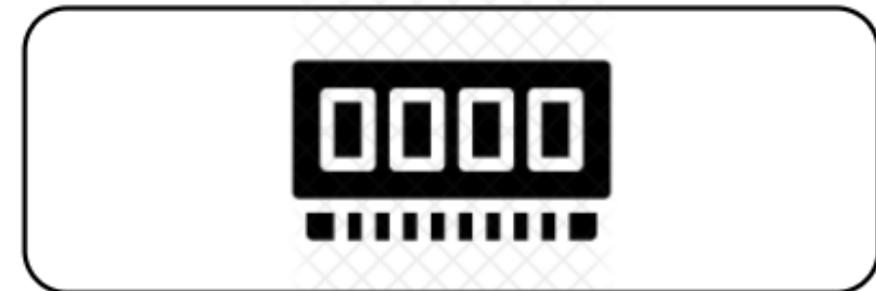
В отличие от индикаторов, теги не хранят в себе значение уровня

Теги нужны, чтобы делать более точные выводы о работе задачи

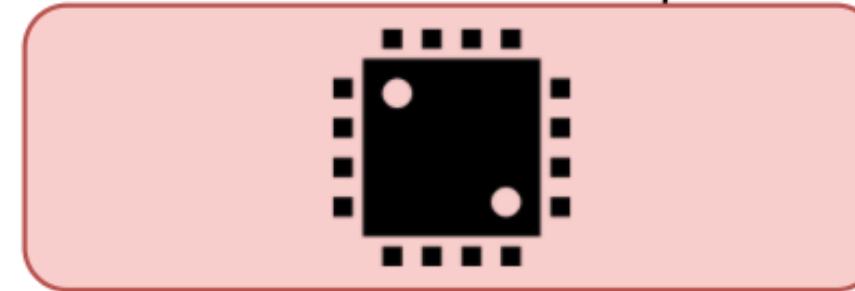
С помощью тегов можно с легкостью внедрять новые свойства в систему выводов

# ПРИМЕР ВЫВОДА НА БАЗЕ ИНДИКАТОРОВ И ТЕГОВ

Низкое использование ОЗУ

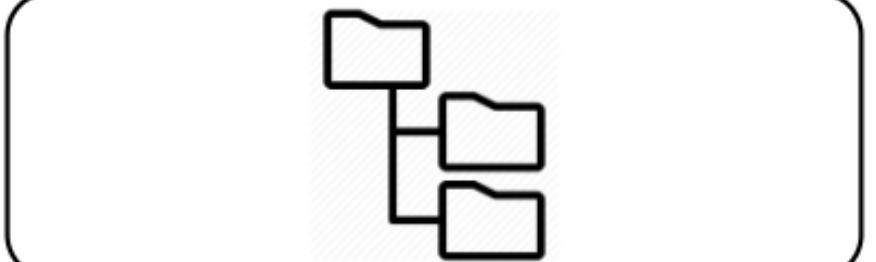


Тег задачи



Низкая средняя загрузка ядра CPU

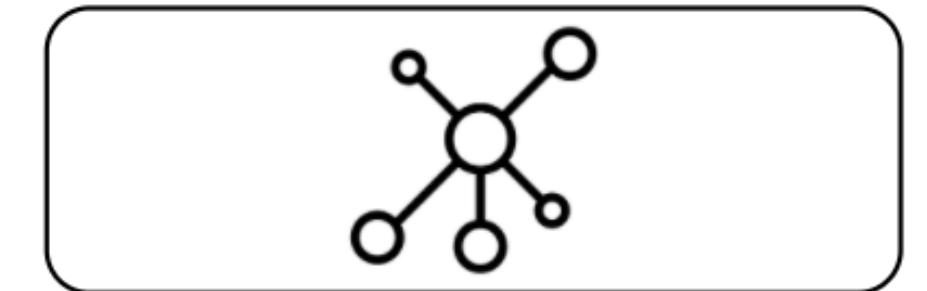
Низкое использование ФС



Вывод:

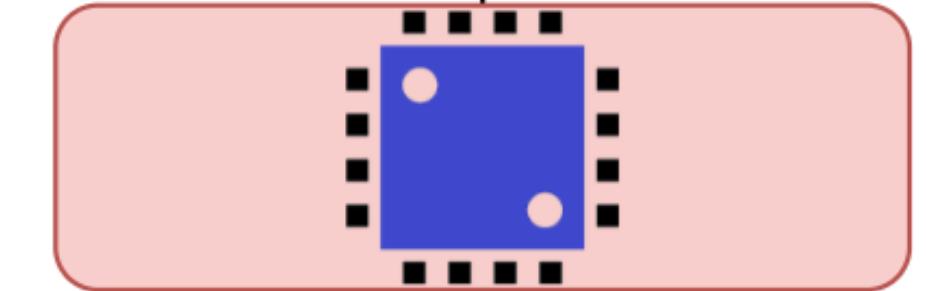
Неэффективное использование Jupyter Notebook  
Низкая средняя загрузка ядра CPU  
Низкая средняя загрузка CPU  
Низкая средняя загрузка GPU

Низкое использование сети InfiniBand

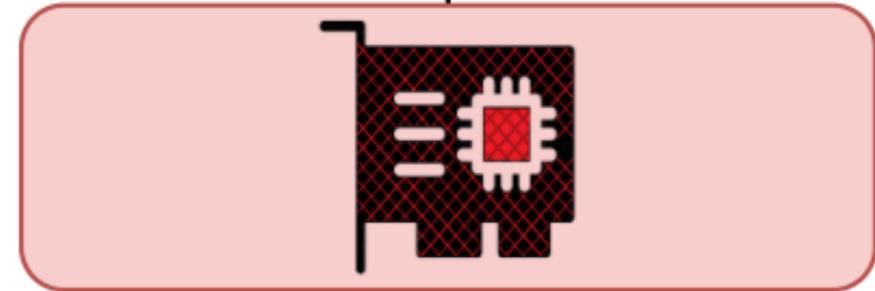


<10 min

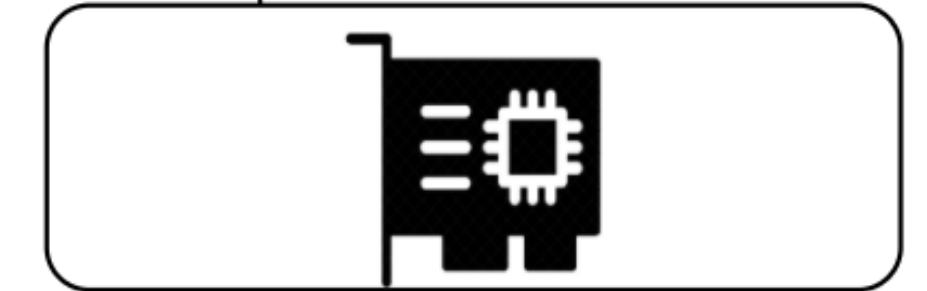
Длительность задачи



Низкая средняя загрузка CPU



Низкая средняя загрузка GPU



Низкое использование памяти GPU

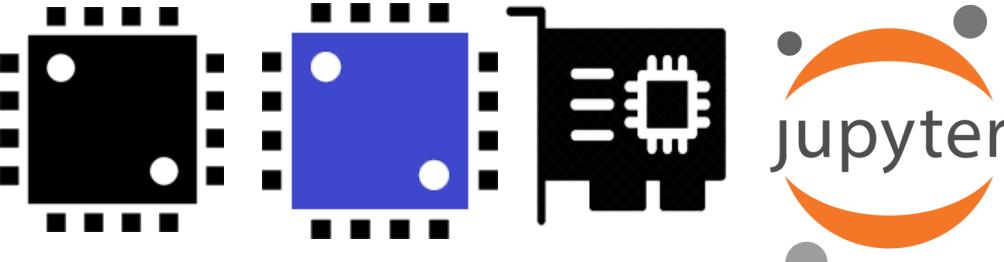


# ИЕРАРХИЯ ВЫВОДОВ В НРС TASKMASTER

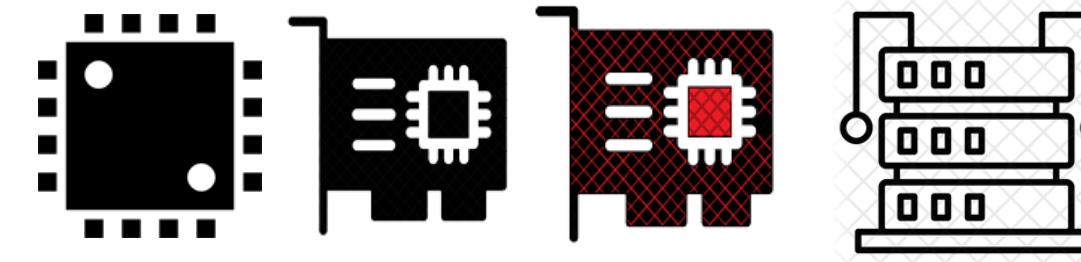
Приоритет вывода

**Точные:** данными выводами являются те, для которых определен сам тип задачи, что позволяет сделать точный вывод на основе индикаторов и тегов

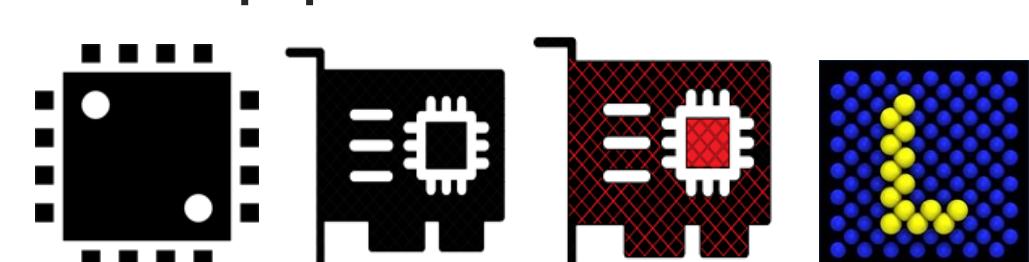
- Jupyter Notebook работает неэффективно



- Неэффективное использование srun/salloc



- Lammps работает неэффективно

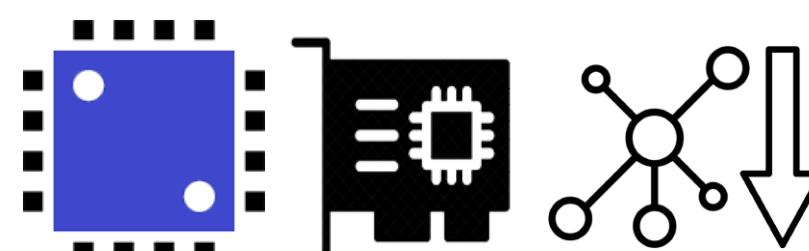


**Расширенные:** более глубокий вывод создается на основании нескольких индикаторов и тегов

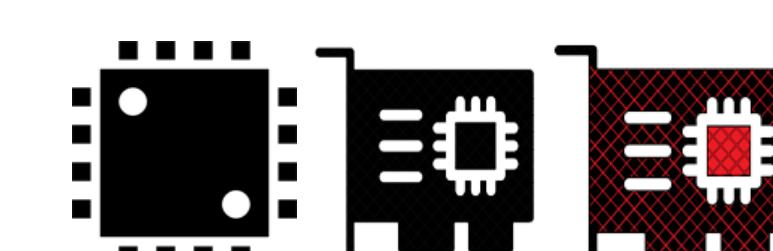
- Ошибка в параметрах запуска задачи



- Непараллельная задача запущена на нескольких узлах

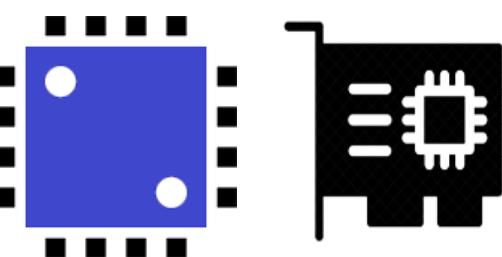


- На узле запущена непараллельная задача

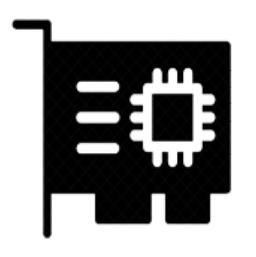


**Базовые:** вывод представляет собой перечисление выявленных индикаторов проблем. Такие выводы можно сделать на основании одного индикатора

- Задача не использует CPU и GPU



- Задача не использует GPU



- Задача практически не использует память GPU

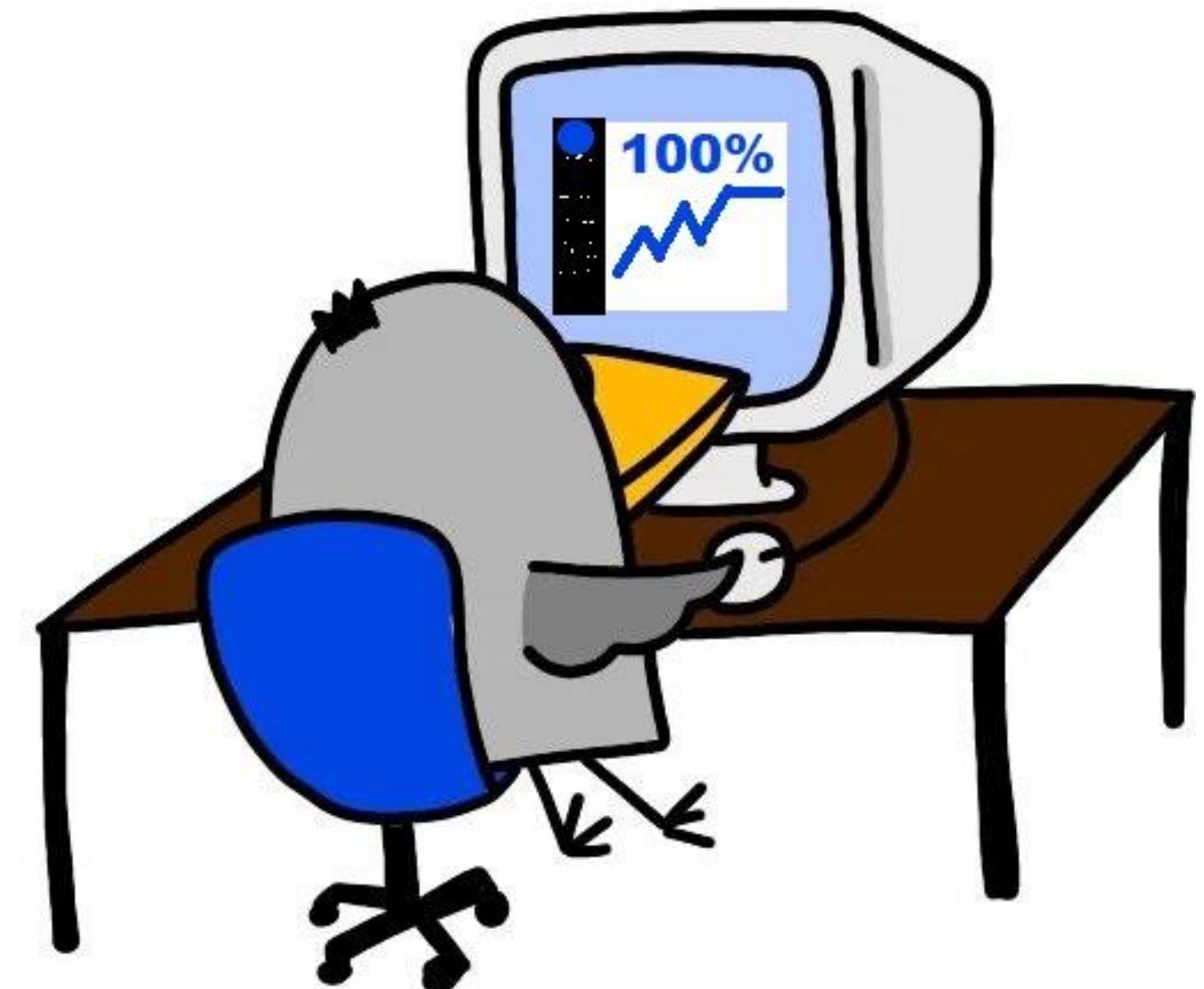


- И т. д.



# ДАЛЬНЕЙШЕЕ РАЗВИТИЕ НРС TASKMASTER

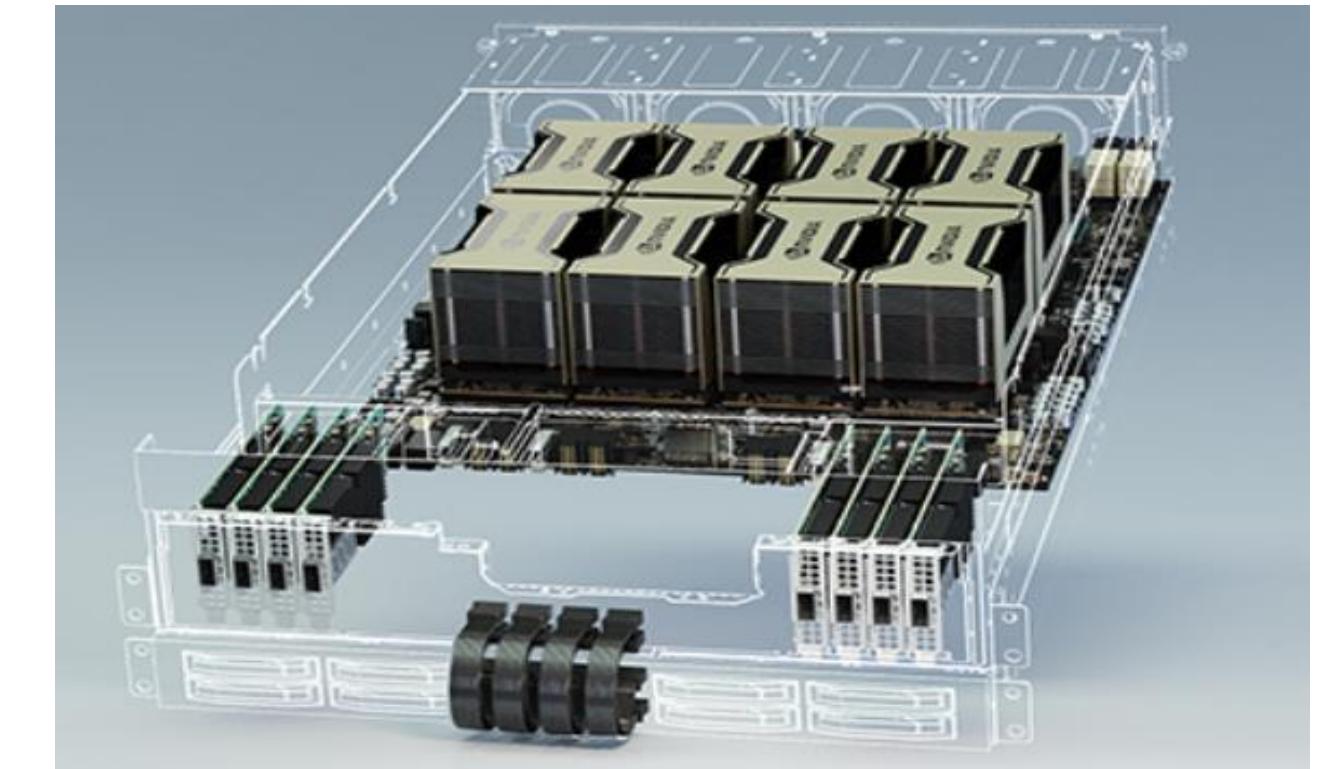
- Разработка способа определения типа задач пользователей за счет анализа временных рядов при помощи математических методов и/или машинного обучения
- Добавление новых типов индикаторов и тегов для формирования новых выводов
- Разработка системы оповещений пользователей о запуске ими неэффективных задач
- Подготовка документации и предоставление открытого доступа к системе





# РАСШИРЕНИЕ СУПЕРКОМПЬЮТЕРА 2021

6 вычислительных узлов типа Е (HPE XL675d Gen10+)	
CPU	2 x AMD <a href="#">EPYC 7702</a> 2-3.35 ГГц (по 64 ядра)
GPU	8 x NVIDIA <a href="#">A100 80GB SXM</a> (NVLink)
Платформа GPU	NVIDIA <a href="#">HGX A100 8-GPU</a>
Оперативная память	1 ТБ
Энергонезависимая память	2 x SSD 960 ГБ (RAID 1)
Вычислительная сеть InfiniBand	2 x Mellanox, HDR(EDR) 200 Гбит/с



NVIDIA HGX A100 8-GPU

# LINPACK НА ГЕТЕРОГЕННОМ КЛАСТЕРЕ

Суперкомпьютер НИУ ВШЭ состоит из узлов разных типов: для обучения искусственного интеллекта, для классический математического моделирования и для задач, не требующих GPU. Неоднородный кластер – наиболее неудобный вариант для тестирования на базе общепринятого Linpack, необходимого для участия в ТОП50.

Не участвовать в ТОП50 суперкомпьютеров нельзя, поэтому на базе GPU-узлов была собрана тестовая виртуальная конфигурация кластера и единый Linpack удалось запустить.

Что потребовалось для теста:

- Обновление микропрограмм старых и новых узлов Dell/HPE.
- Поэтапное обновление микропрограмм коммутаторов.
- Обновление микропрограмм сетевых адаптеров.
- Устранение несовместимости различных версий firmware.
- Обновление драйверов.
- Обновление планировщика задач
- Обновление CUDA.
- Сборка универсального образа для старых и новых вычислительных узлов A100 / V100 / Intel / AMD / HP/ Dell.
- Промежуточное тестирование оборудования.
- Синхронизация частот CPU, шины и GPU.
- Подбор размерности матрицы для теста.
- И, наконец, разделение A100 на две виртуальные части по технологии MIG.

Результат: 927.4 Терафлопс.

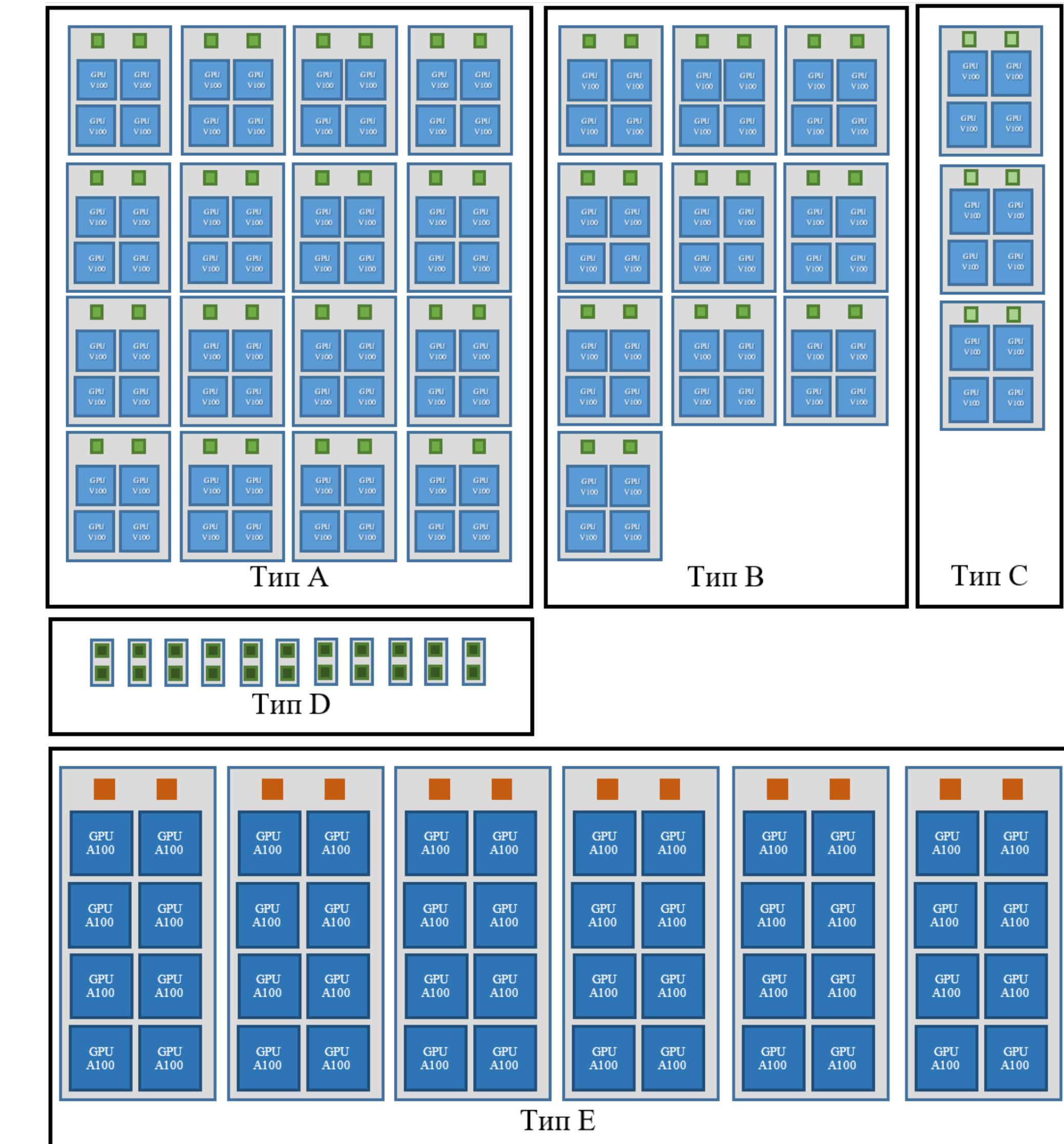


Рис. 1. Схематическое представление вычислительных узлов суперкомпьютера (размер пропорционален производительности)

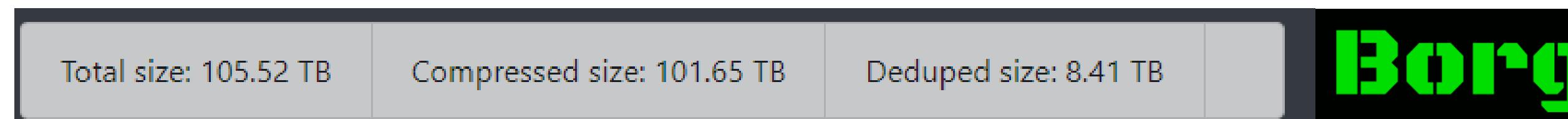
# СЕРВЕР РЕЗЕРВНОГО КОПИРОВАНИЯ

Общее дисковое пространство – 182 ТБ (RAW).

Доступное дисковое пространство (RAID10) – 76 ТБ.

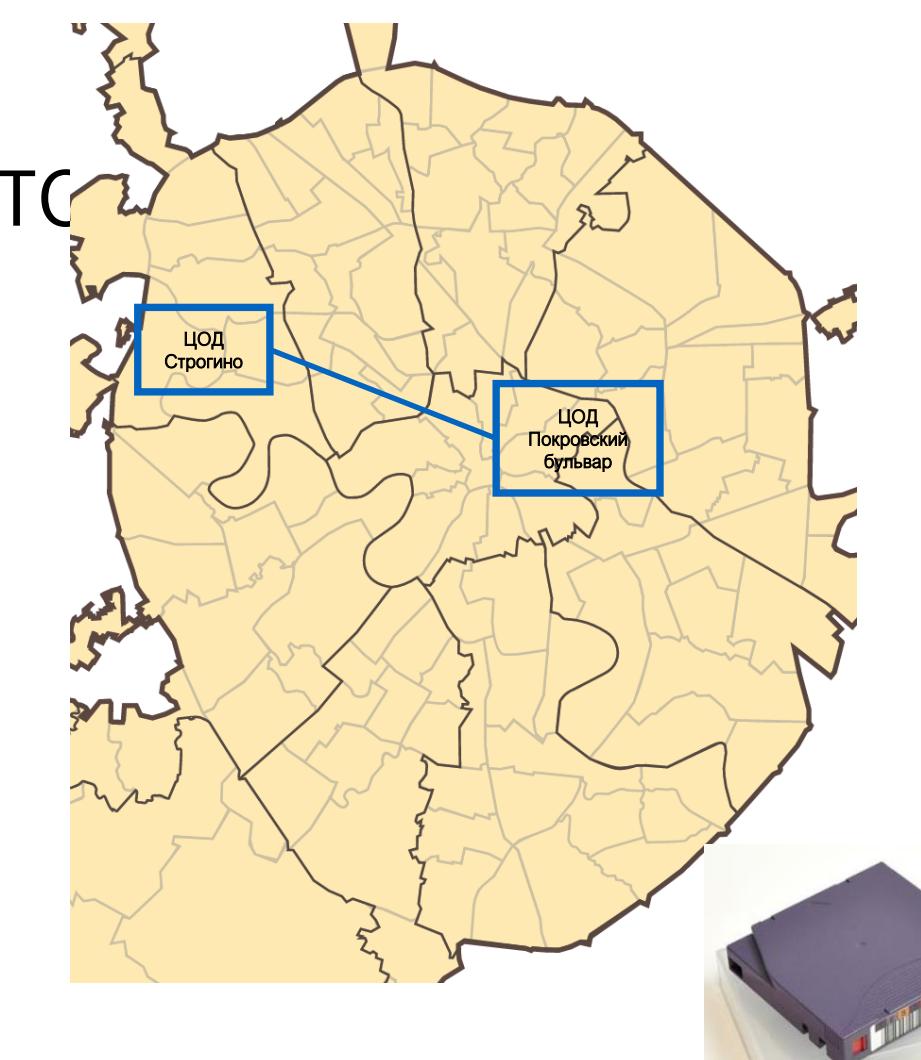
Для резервного копирования только критичной информации (не директорий пользователей).

На сервере нами настроено открытое ПО *Borg Backup* для сжатия, дедупликации и шифрование данных.



HPE Apollo 4200 Gen10

Пользовательские директории раз в квартал сжимаются и копируются на ленточную библиотеку в резервный ЦОД на Покровском бульваре.



Dell PowerVault ML3  
дит



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

# ПРОЕКТЫ, ВЫПОЛНЯЮЩИЕСЯ НА СУПЕРКОМПЬЮТЕРЕ (СО СТАТЬЯМИ ИЗ Q1)



# НАУЧНЫЕ ИССЛЕДОВАНИЯ НА СУПЕРКОМПЬЮТЕРЕ

---

**За 2020 г. общее финансирование работ, выполнение которых требовало суперкомпьютерных вычислений, составило более 570 млн. руб.**



# ИЗБРАННЫЕ ПУБЛИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ СУПЕРКОМПЬЮТЕРНОГО КОМПЛЕКСА

Опубликовано более 60 статей со ссылкой на использование ресурсов суперкомпьютерного комплекса.

Год	ВСЕГО	SCOPUS					CORE A*	Другие
		Всего в SCOPUS	Q1	Q2	Q3	Q4		
2021 (данные на июнь)	27	22+	4	1	16	1	0	5
2020	34	24	11	3	4	6	3	7

В соответствии с решением НТС СКК, авторы в своих публикациях должны ссылаться на использование ресурсов СК, для чего необходимо в разделе Acknowledgments использовать следующие формулировки: «This research was supported in part through computational resources of HPC facilities at HSE University [1]» или «Исследование выполнено с использованием суперкомпьютерного комплекса НИУ ВШЭ [1]».

[1] Kostenetskiy P.S., Chulkevich R.A., Kozyrev V.I. HPC Resources of the Higher School of Economics // Journal of Physics: Conference Series. 2021. Т. 1740, № 1. Р. 012050

\* Ссылка необходима для автоматического учета через Scopus исследований, выполненных на суперкомпьютере.  
Мы не получаем какие-либо надбавки за ваши цитирования, т.к. работаем в учебно-вспомогательном подразделении.



# АТОМИСТИЧЕСКОЕ И КОНТИНУАЛЬНОЕ МОДЕЛИРОВАНИЕ В ЗАДАЧАХ ФИЗИКИ

## Заказчики:

Программа фундаментальных исследований НИУ ВШЭ

## Исполнитель:

Международная лаборатория суперкомпьютерного атомистического моделирования и многомасштабного анализа НИУ ВШЭ

**Руководитель проекта:** Писарев Василий Вячеславович

## Цель исследования:

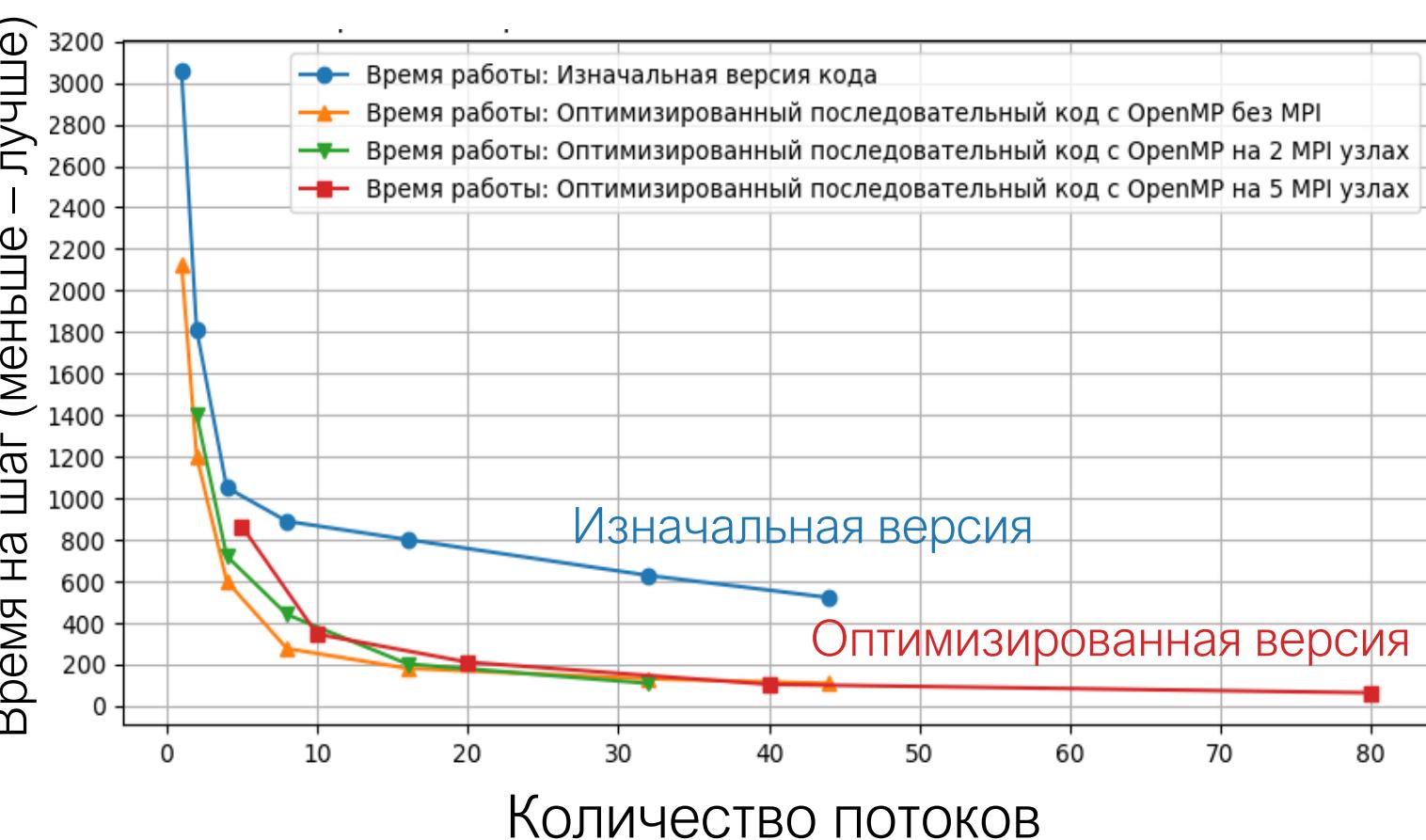
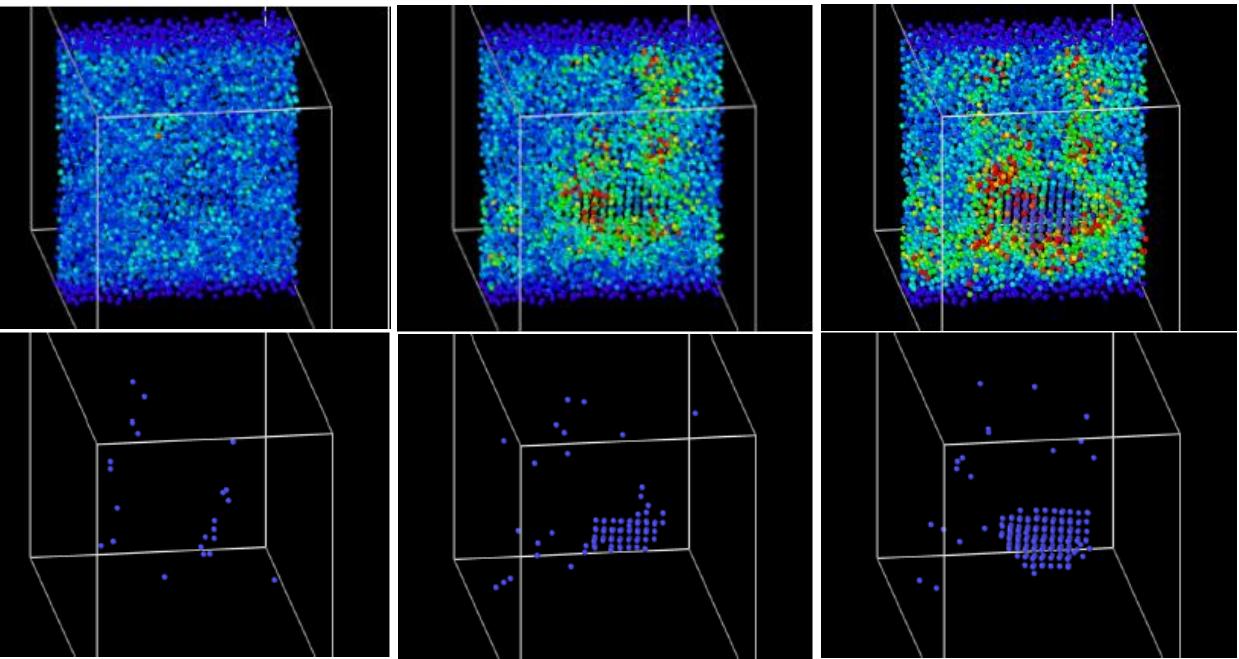
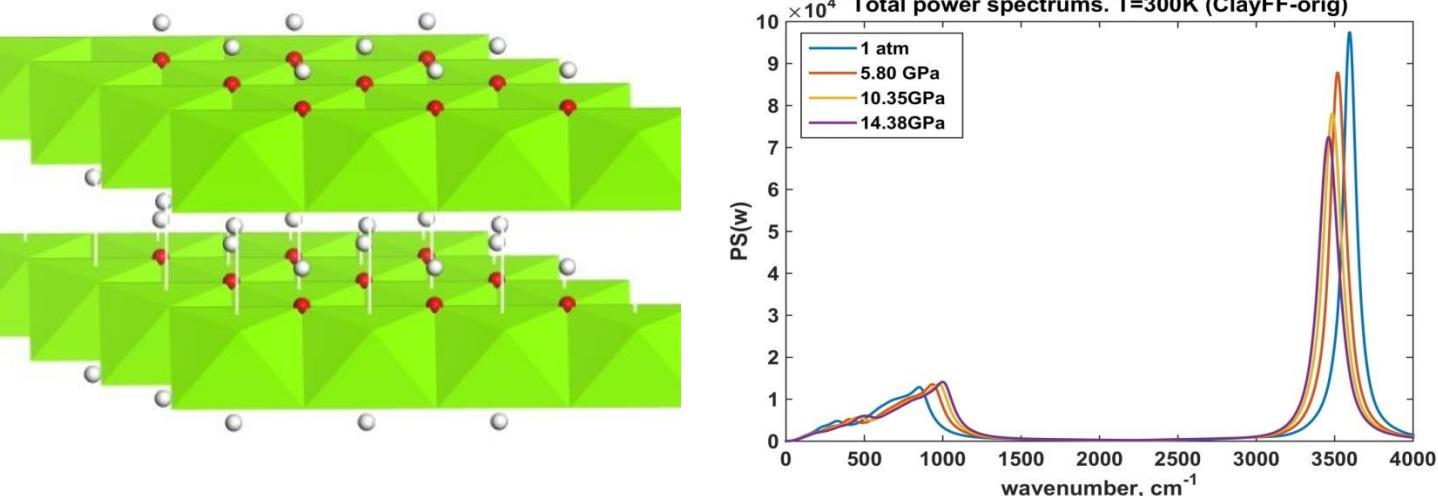
Исследование свойств материалов методами суперкомпьютерного моделирования, разработка вычислительных методов и их оптимизация под архитектуру СК ВШЭ

## Результат:

- 1) Методом МД рассчитаны свойства материалов – уравнение состояния и свойства брусита  $Mg(OH)_2$  при высоких давлениях, коэффициенты переноса углеводородных флюидов, механизмы кристаллизации жидкотекущих пленок.
- 2) Программа для моделирования плазменного потока оптимизирована и распараллелена с использованием технологий OpenMP и MPI, что привело к ускорению расчёта в 48 раз. Исходные коды выложены в открытый доступ: <https://github.com/kolotinsky1998/Ionwake/tree/develop>

## Публикации:

- Kondratyuk N., Lenev D., Pisarev V. Transport coefficients of model lubricants up to 400 MPa from molecular dynamics // J. Chem. Phys., 2020 **(Q1)**
- Kirova E., Pisarev V. The morphological aspect of crystal nucleation in wall-confined supercooled metallic melt // J. Phys.: Condensed Matter **(Q1)**
- Kondratyuk N., Pisarev V., Ewen J. P. Probing the high-pressure viscosity of hydrocarbon mixtures using molecular dynamics simulations // J. Chem. Phys. 2020 **(Q1)**





# РАЗРАБОТКА АЛГОРИТМОВ МЮОННОЙ ИДЕНТИФИКАЦИИ ДЛЯ LHCb

## Заказчики:

Коллаборация LHCb



## Исполнитель:

Научно-учебная лаборатория методов анализа больших данных ФКН НИУ ВШЭ

**Руководитель проекта:** Казеев Никита Александрович

## Цель исследования:

Разработка быстрых методов для идентификации мюонов в детекторе LHCb

## Результат:

Разработана и внедрена модель машинного обучения на основе CatBoost, исследованы качество и скорость её работы. Исследование представлено во внутреннем отчёте коллаборации и журнальной статье.

## Публикации:

*Anderlini L. et al.* Muon identification for LHCb Run 3 // Journal of Instrumentation, 2020 (Q1)

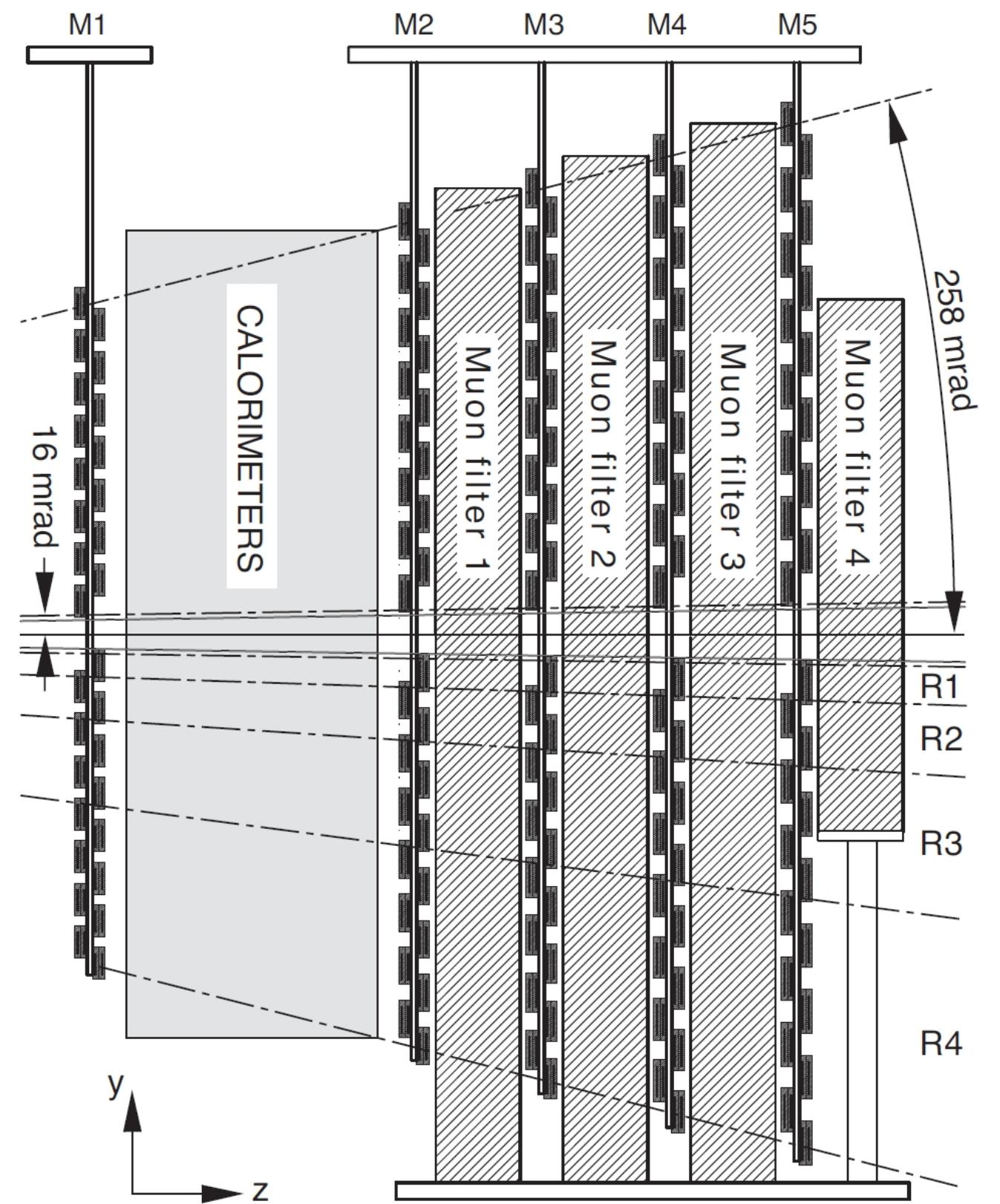


Image: side view of the LHCb muon detector, reproduced from  
A Augusto Alves Jr et al. "Performance of the LHCb muon  
system".

In: Journal of Instrumentation 8.02 (2013), P02022.



# БЕЙСЛАЙНЫ ТРАНСФЕРНОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ТОНАЛЬНОСТИ РУССКОЯЗЫЧНЫХ ТЕКСТОВ

## Заказчики:

Департамент бизнес-информатики Высшей школы бизнеса НИУ ВШЭ

## Исполнитель:

С.И. Сметанин, аспирантская школа по компьютерным наукам НИУ ВШЭ

Руководитель проекта: Комаров Михаил Михайлович

## Цель исследования:

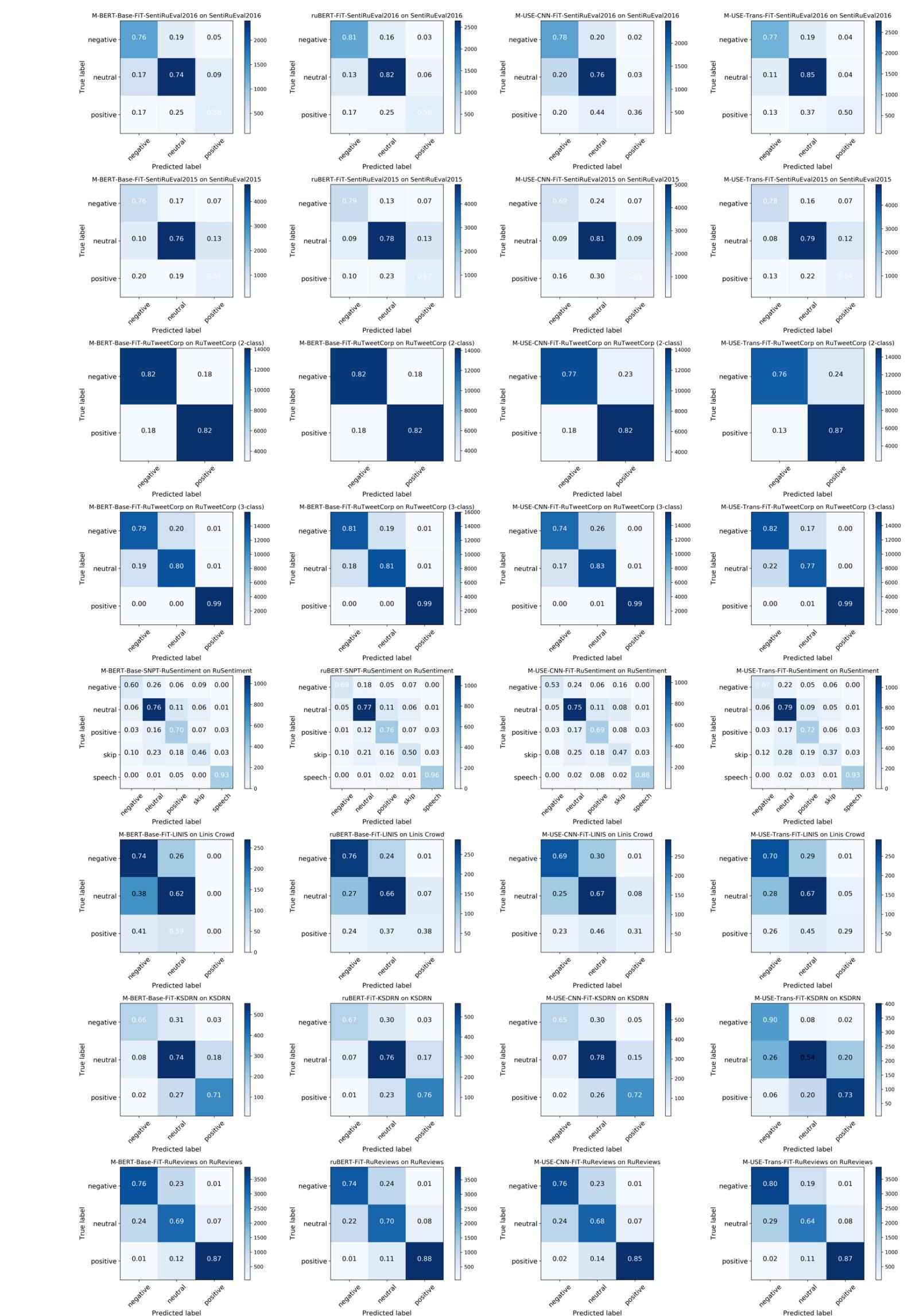
Определение бейслайнов трансферного обучения для анализа тональности русскоязычных текстов.

## Результат:

Были выявлены наиболее часто используемые публичные наборы данных для анализа тональности на русском языке и наиболее актуальные языковые модели, которые официально поддерживают русский язык. Была произведена тонкая настройка языковых моделей (Multilingual BERT, RuBERT и Multilingual USE) получены высоки и в некоторых случаях лидирующие результаты классификации на семи наборах данных: SentiRuEval-2016, SentiRuEval-2015, RuTweetCorp, RuSentiment, LINIS Crowd, Kaggle Russian News Dataset, а также RuReviews. Обученные модели были опубликованы в открытом доступе для исследовательского сообщества.

## Публикации:

*Smetanin, S., Komarov, M. Deep transfer learning baselines for sentiment analysis in Russian // Information Processing & Management, 2021 (Q1)*





# РАСЧЕТ СВОБОДНОЙ ЭНЕРГИИ СОЛЬВАТАЦИИ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ В скCO<sub>2</sub>

**Заказчики:** РФФИ

**Исполнитель:**

- Департамент прикладной математики МИЭМ НИУ ВШЭ
- Институт химии растворов им. Г.А.Крестова РАН

**Руководитель проекта:** Будков Юрий Алексеевич

**Цель исследования:**

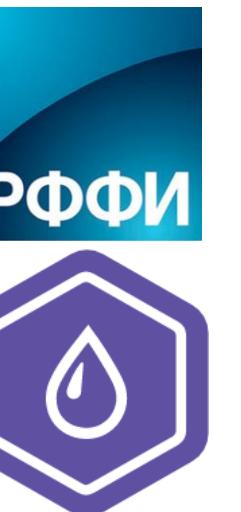
Расчет значений свободной энергии сольватации лекарственных соединений в сверхкритическом диоксиде углерода методами полно-атомной и грубо-структурной молекулярной динамики.

**Результат:**

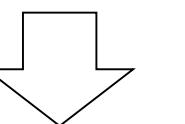
Обнаружено, что использование грубо-структурных методов для моделирования слаборастворимых лекарственных соединений в сверхкритическом диоксиде углерода дает количественные результаты, находящиеся в хорошем согласии с экспериментальными данными и результатами полно-атомного молекулярно-динамического моделирования.

**Публикации:**

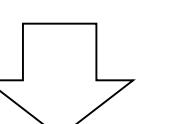
- Gurina D., Budkov, Y., Kiselev M. Impregnation of Poly (methyl methacrylate) with Carbamazepine in Supercritical Carbon Dioxide: Molecular Dynamics Simulation // The Journal of Physical Chemistry B, 2020 (Q1)
- Kalikin N., Kurskaya M., Ivlev D., Krestyaninov M., Oparin R., Kolesnikov A., Budkov Y., Idrissi A., Kiselev M. Carbamazepine solubility in supercritical CO<sub>2</sub>: a comprehensive study // Journal of Molecular Liquids, 2020 (Q1)



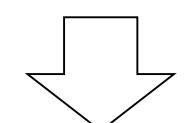
Критические параметры  
соединения и растворителя



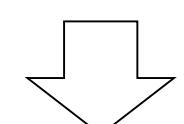
Параметры потенциала  
межмолекулярного  
взаимодействия



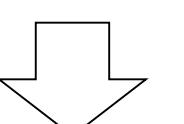
Молекулярно-динамическое  
моделирование леннард-  
джонсовского флюида



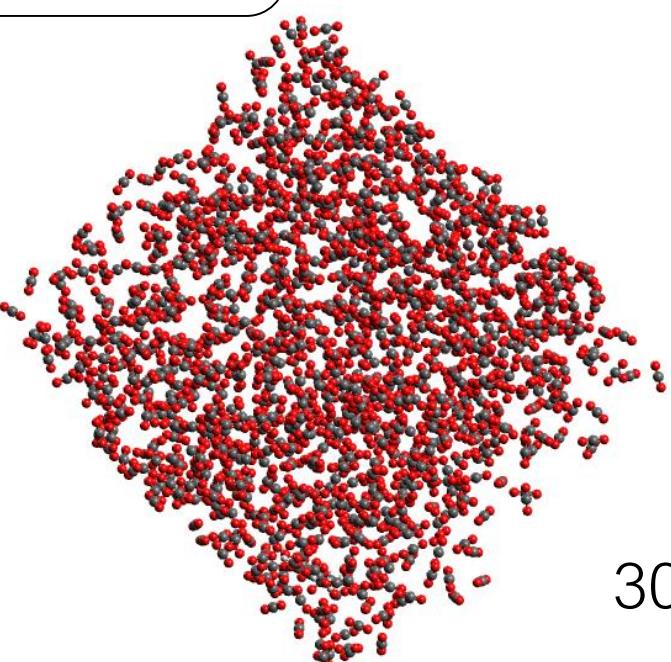
Расчет парциальных  
зарядов соединения



Полно-атомное  
молекулярно-динамическое  
моделирование



Расчет свободной энергии  
сольватации





# РАЗРАБОТКА НОВЫХ МЕТОДОВ ПРЕДСКАЗАНИЯ И АНАЛИЗА ХАОТИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

## Исполнитель:

Департамент анализа данных и искусственного интеллекта ФКН

**Руководитель проекта:** д.ф.-м.н. В.А. Громов

## Цель исследования:

Разработка новых методов предсказания и анализа хаотических временных рядов.

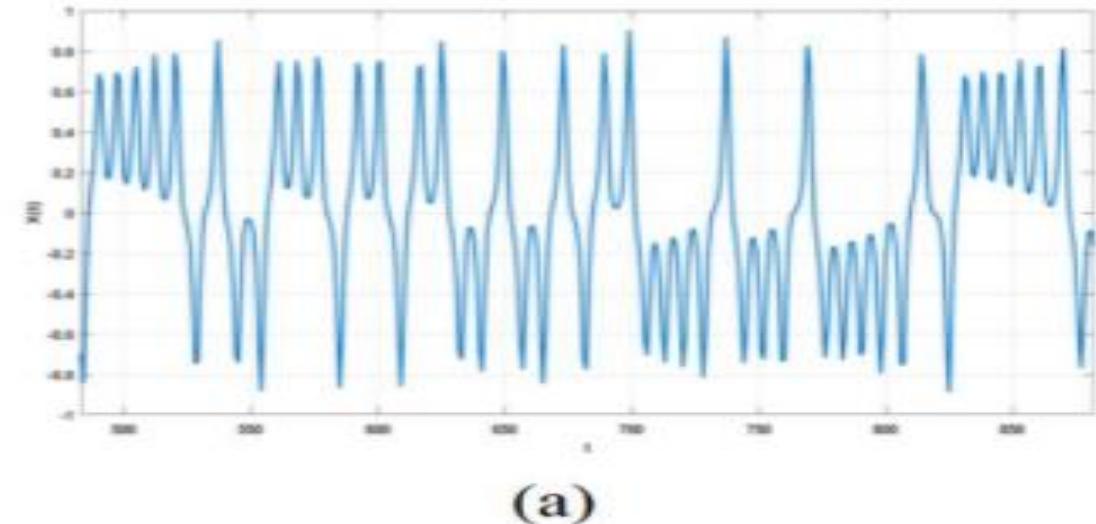
## Результат:

Для решения задачи разработаны различные методы генерации, прогнозирования и сравнения хаотических временных рядов. Методы были протестированы на различных наборах реальных данных, таких как количество потребленной электроэнергии, курс золота, колебания температуры и показали хорошие результаты.

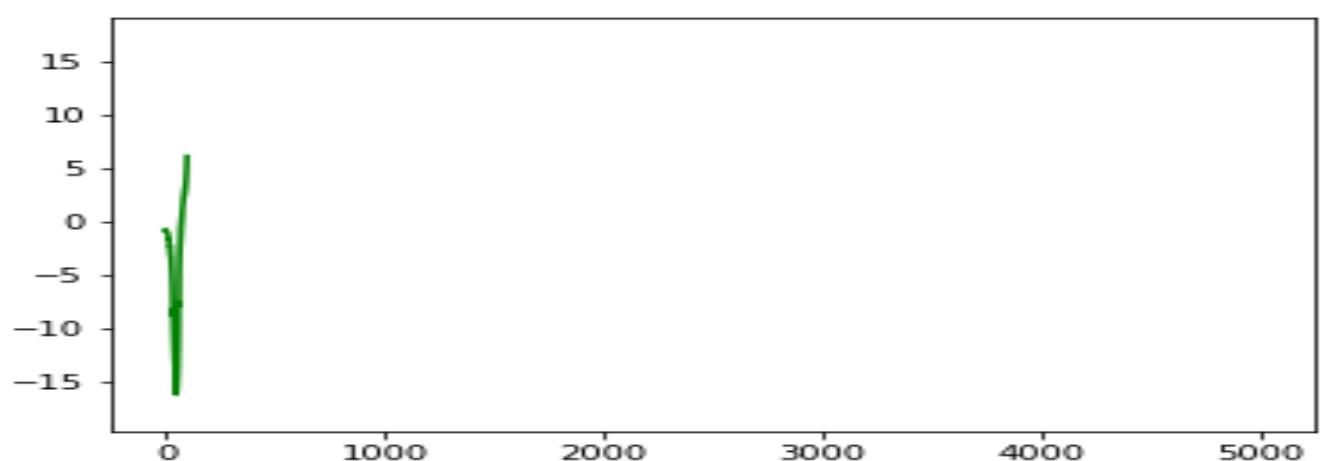
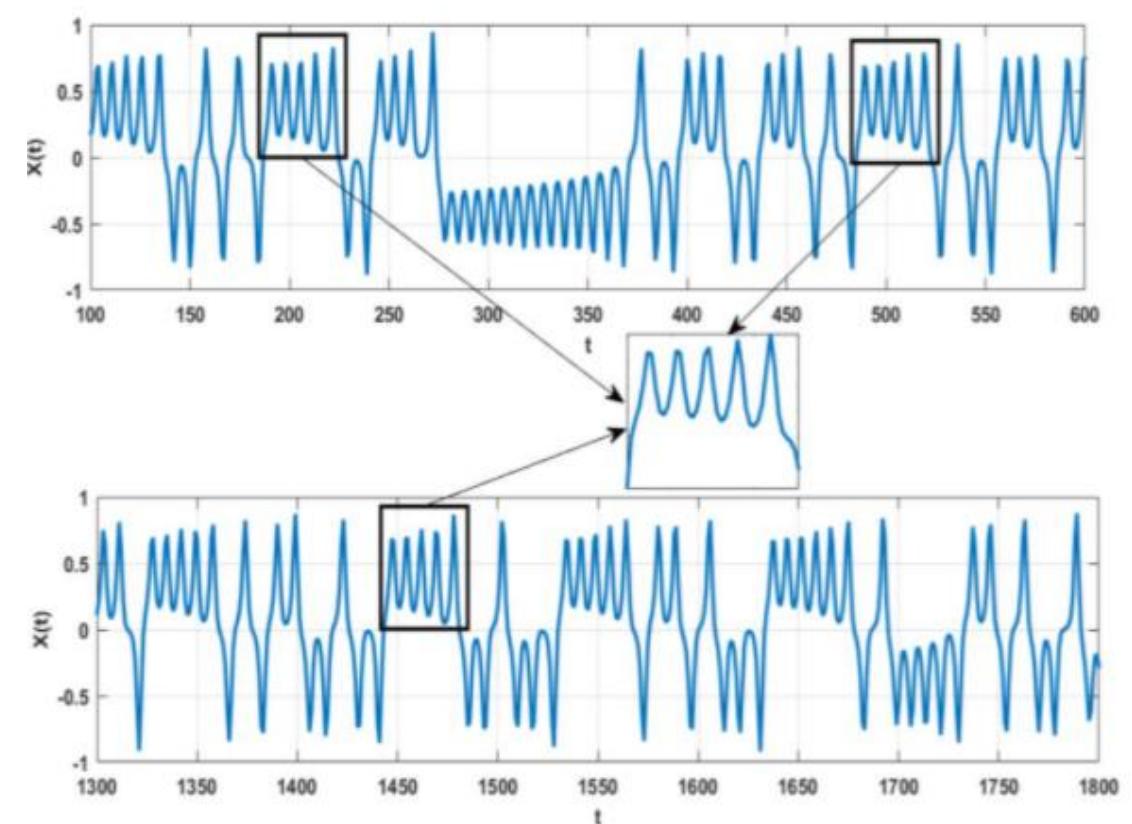
Методы реализованы в рамках библиотеки `pintur` и `scipy`.

## Публикации:

Gromov V., Necheporenko A., Gaisin A., Volkov I., Diner S. Generalized Relational Tensors for Irregularly Sampled Time Series: Methods to Store, Re-generate, and Predict // Chaos: An Interdisciplinary Journal of Nonlinear Science, подано в печать (Q1)

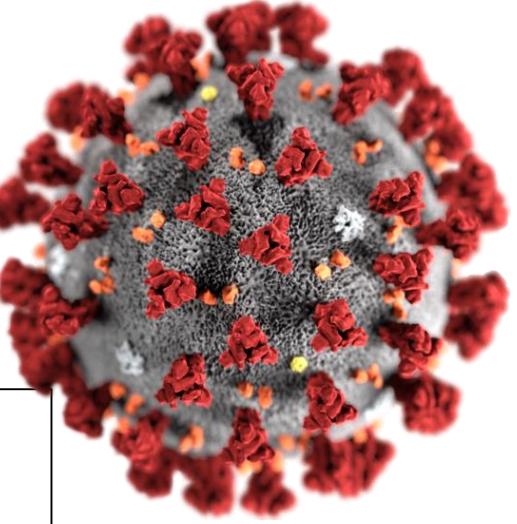


$$P(I_s = I_s^*, p_s = p_s^*) = \frac{T(I_1^*, \dots, I_s^*, p_1^*, \dots, p_s^*)}{\sum_{(I_s, p_s)} T(I_1^*, \dots, I_{s-1}^*, I_s, p_1^*, \dots, p_{s-1}^*, p_s)},$$





# АНАЛИЗ ЭПИДЕМИИ КОРОНАВИРУСА В РОССИИ



**Заказчик:** консорциум CoRGi



**Исполнитель:**

Международная лаборатория статистической и вычислительной геномики

**Руководитель проекта:** PhD Щур Владимир Львович

**Партнеры:**

Сколтех (Москва), НИИ гриппа (Санкт-Петербург), ИППИ (Москва)

**Цель исследования:**

Анализ эпидемии коронавируса SARS-CoV-2 в России.

**Результаты:**

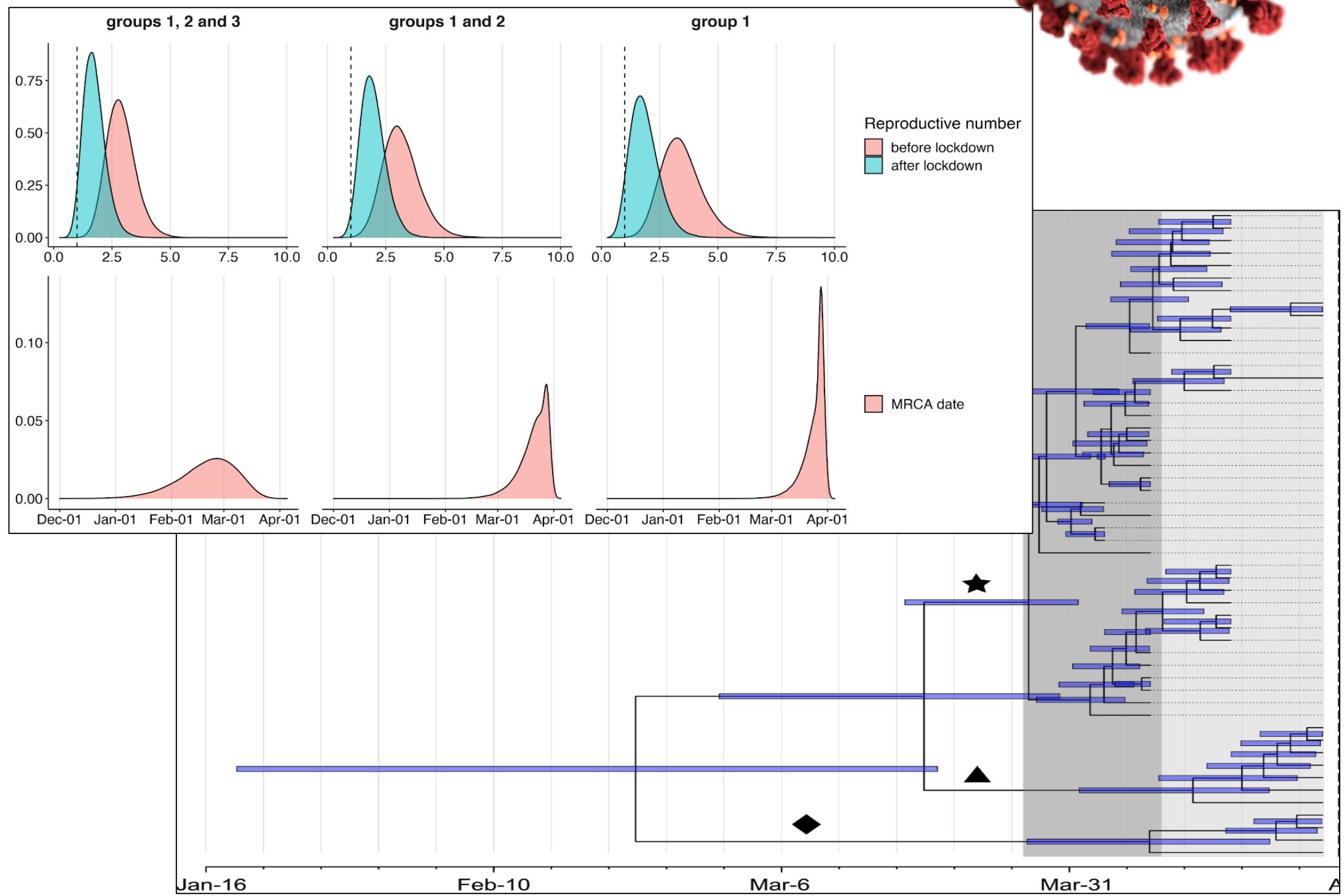
- Определено, что в Россию коронавирус завозился несколько десятков раз.
- Большинство завозов случились в конце февраля - первой половине марта из Европы.
- Вирус был завезен в Россию не менее 67 раз.
- Следов завозов инфекции из Китая нет.
- Выполнен байесовский анализ филодинамики вспышки Covid-19 в НИИ им. Вредена (Санкт-Петербург) с использованием пакета BEAST2.
  - Определен эффект карантина,
  - Обнаружено от двух до трёх независимых заносов инфекции в больницу.

**Перспективы:**

- Изучение внутрипациентской изменчивости коронавируса SARS-CoV-2:
  - Разработка нового метода для определения вариантов вируса, циркулирующих в популяции,
  - Планируется разработка моделей, учитывающих передачу нескольких вариантов вируса между пациентами.

**Публикации:**

Komissarov et al. Genomic epidemiology of the early stages of SARS-CoV-2 outbreak in Russia // Nature Communications, 2021 (Q1)



**Упоминания в новостях:**

<https://iq.hse.ru/news/381056189.html>

<https://www.kommersant.ru/doc/4441864>

<https://www.vesti.ru/nauka/article/2431960> и др.



# ОПТИМИЗАЦИЯ ЧЁРНОГО ЯЩИКА С ПОМОЩЬЮ ЛОКАЛЬНЫХ ГЕНЕРАТИВНЫХ СУРРОГАТОВ

**Исполнитель:** Лаборатория методов анализа больших данных ФКН НИУ ВШЭ

**Руководитель проекта:** к.ф.-м.н. Устюжанин Андрей Евгеньевич

**Цель:** Разработка новых методов для оптимизации недифференцируемых симуляторов для оптимизации калориметра LHCb

## Описание:

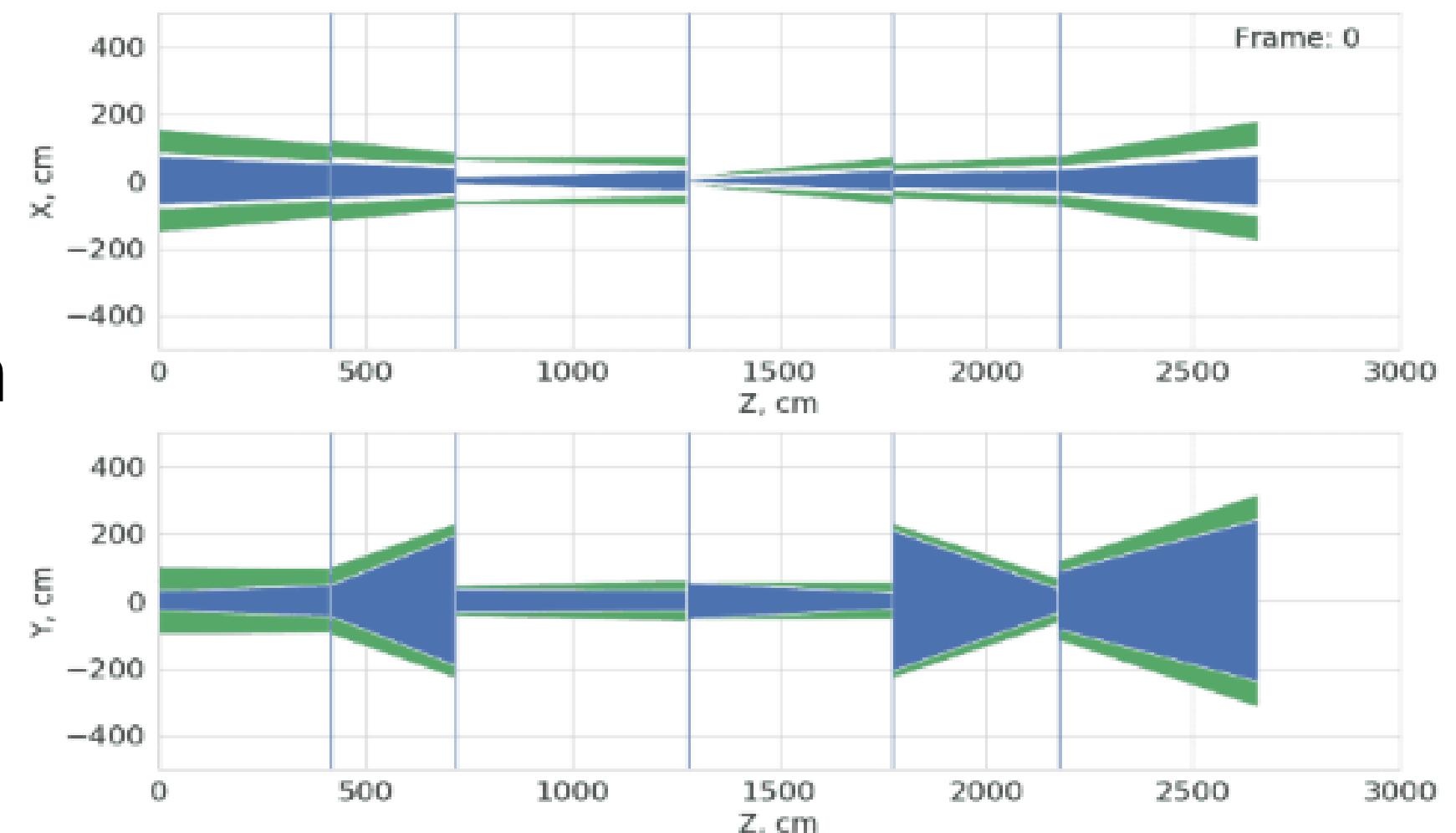
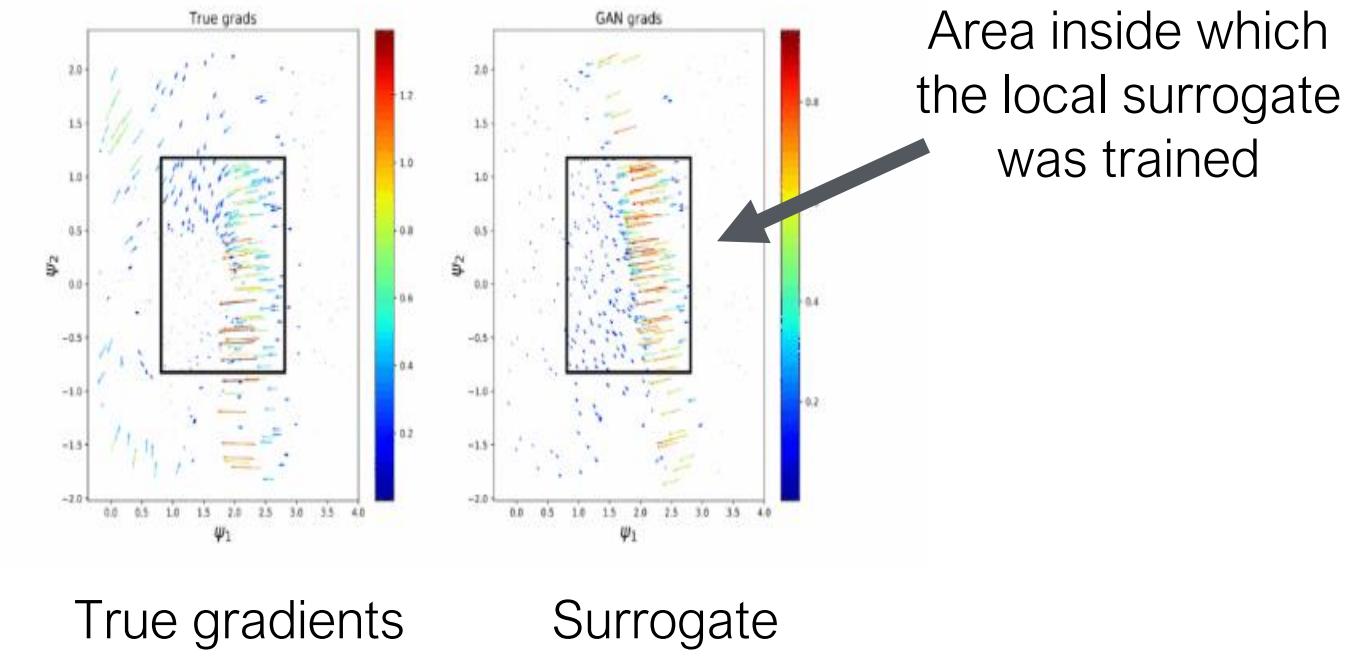
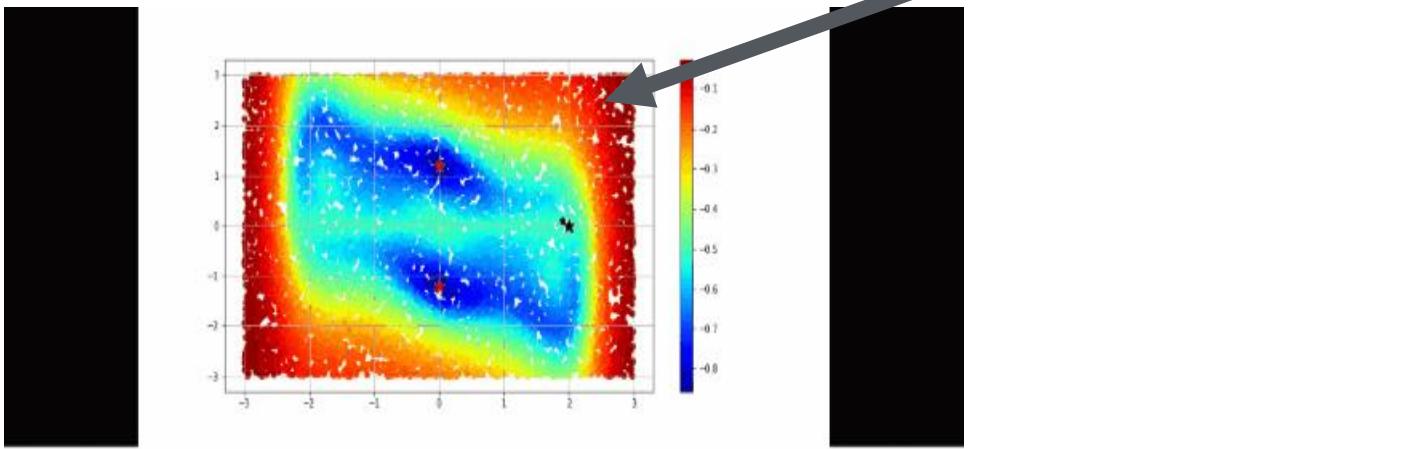
Мы используем глубокие генеративные модели для итерационной аппроксимации симулятора в локальных окрестностях пространства параметров.

Мы показываем, что эти локальные суррогаты могут использоваться для аппроксимации градиента симулятора и, таким образом, позволяют оптимизировать параметры симулятора с помощью градиентных методов оптимизации.

Метод реализован в рамках библиотеки pytorch, все исходные коды находятся в открытом доступе: <https://github.com/shir994/L-GSO>

## Публикации:

Shirobokov S., Belavin V., Ustyuzhanin A., Kagan M., Baydin G. Black-Box Optimization with Local Generative Surrogates // Advances in Neural Information Processing Systems 33, 2020 (A\*)





# ПАРАЛЛЕЛЬНЫЕ АЛГОРИТМЫ И АТОМИСТИЧЕСКОЕ МНОГОМАСШТАБНОЕ МОДЕЛИРОВАНИЕ МАТЕРИАЛОВ

## Заказчики:

Программа фундаментальных исследований НИУ ВШЭ

## Исполнитель:

Международная лаборатория суперкомпьютерного атомистического моделирования и многомасштабного анализа НИУ ВШЭ

**Руководитель проекта:** Стегайлов Владимир Владимирович

## Цель исследования:

Разработка и применение параллельных алгоритмов классических и квантовых молекулярно-динамических расчетов для многомасштабного моделирования материалов.

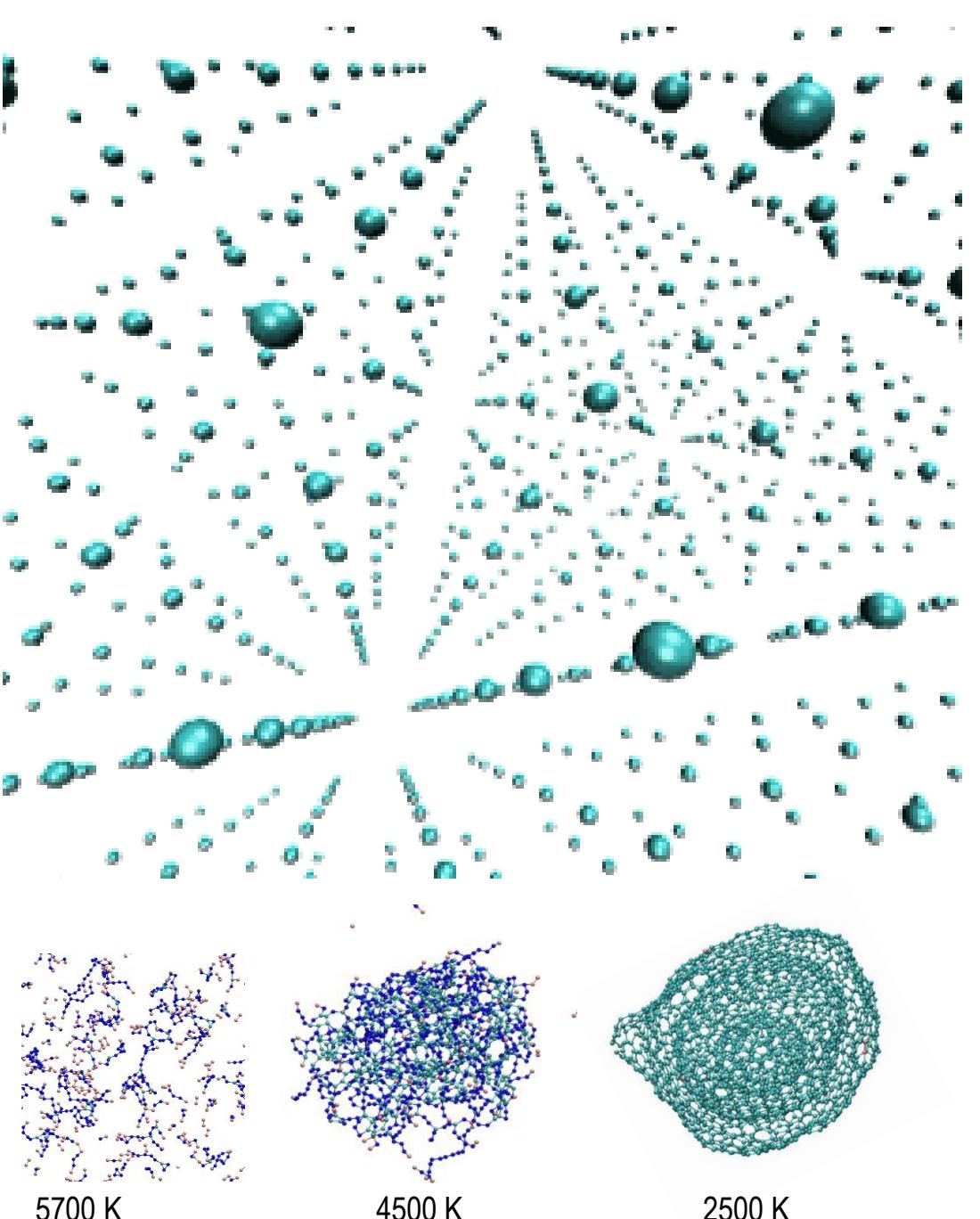
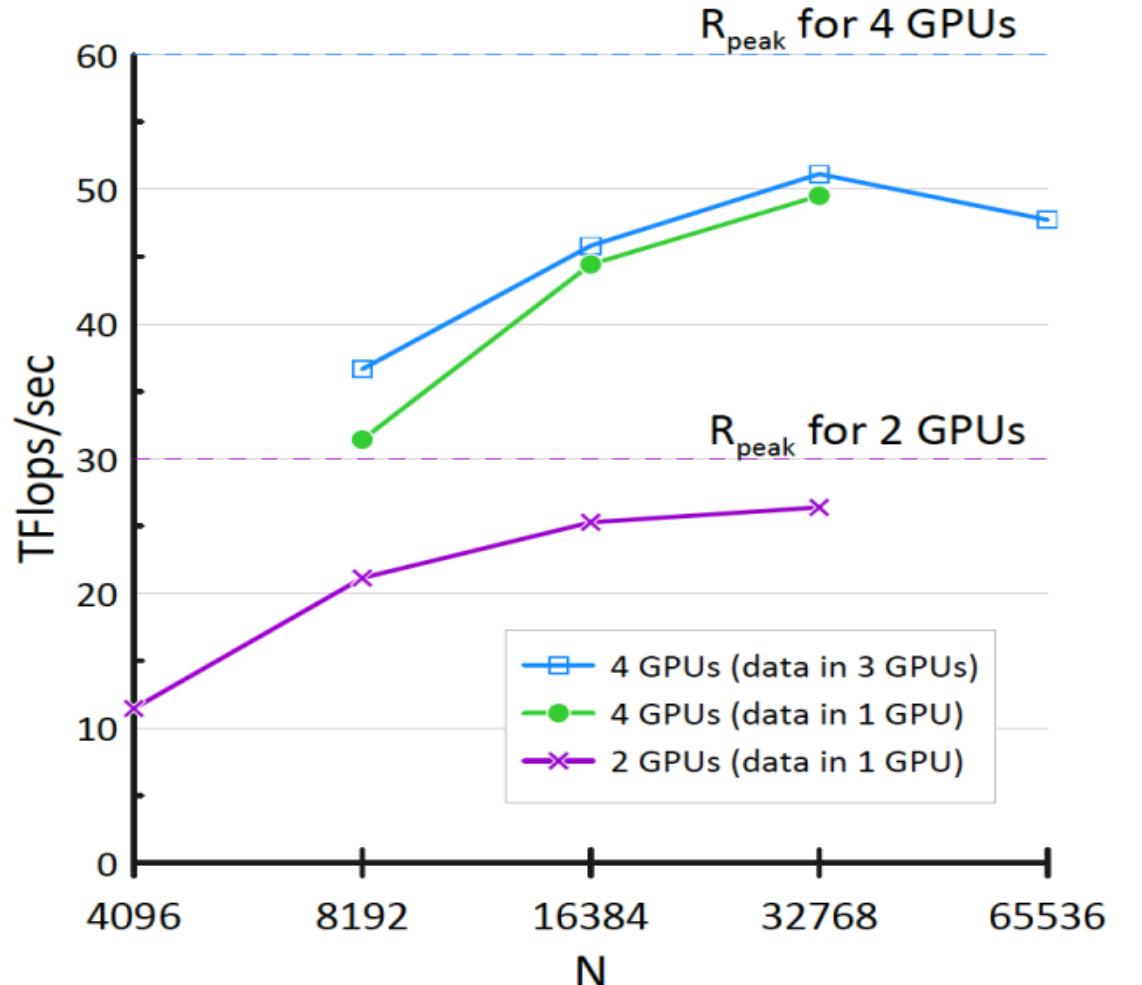
## Результат:

Разработан алгоритм матричного умножения для систем с несколькими GPU-ускорителями, объединенными высокоскоростной шиной передачи данных NVLink.

Проведено моделирование процесса образования фуллереноподобных углеродных наноструктур и определены наилучшие параметризации реакционного потенциала межатомного взаимодействия для углерода ReaxFF.

## Статьи:

- Orekhov N., Ostroumova G., Stegailov V. High temperature pure carbon nanoparticle formation: Validation of AIREBO and ReaxFF reactive molecular dynamics // Carbon. V.170, 2020 (Q1)
- Choi Y. R., Nikolskiy V., Stegailov V. Matrix-Matrix Multiplication Using Multiple GPUs Connected by Nvlink // Proceedings of IEEE: 2020 Global Smart Industry Conference (GloSIC), 2020





# ИЗУЧЕНИЕ И УСОВЕРШЕНСТВОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ АНСАМБЛИРОВАНИЯ НЕЙРОННЫХ СЕТЕЙ

## Заказчики:

Samsung Research, Samsung Electronics

## Исполнитель:

Центр глубинного обучения и байесовских методов,  
лаборатория компании Самсунг ФКН НИУ ВШЭ

**Руководитель проекта:** Ашуха Арсений Павлович

## Цель исследования:

Изучение и усовершенствование современных методов ансамблирования  
нейронных сетей.

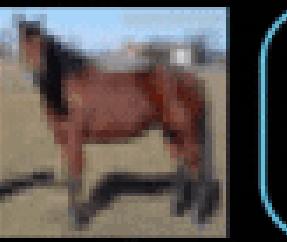
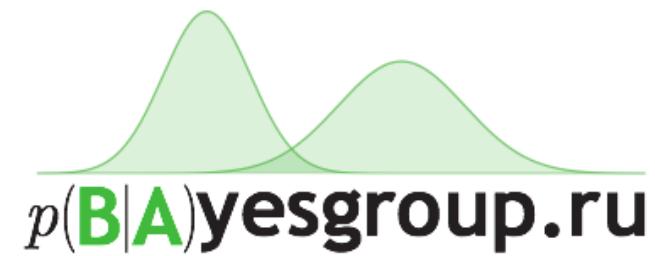
## Результат:

- Проведено сравнение методов ансамблирования на реалистичных изображениях (CIFAR-10, CIFAR-100, ImageNet)
- Показана несостоятельность широко используемых метрик качества ансамблей, предложены метрики лишенные ряда недостатков;
- Предложены новые методы ансамблирования, которые могут быть скомбинированы с другими подходами
- Методы реализован в рамках библиотеки pytorch, все исходные коды выложены в открытый доступ: [github.com/bayesgroup/pytorch-ensembles](https://github.com/bayesgroup/pytorch-ensembles), [github.com/bayesgroup/gps-augment](https://github.com/bayesgroup/gps-augment).

## Публикации:

- Molchanov D., Lyzhov A., Molchanova Y., Ashukha A., Vetrov D. Greedy Policy Search: A Simple Baseline for Learnable Test-Time Augmentation // Conference on Uncertainty in Artificial Intelligence 2020 (A\*)
- Ashukha A., Lyzhov A., Molchanov D., Vetrov D. Pitfalls of In-Domain Uncertainty Estimation and Ensembling in Deep Learning // International Conference on Learning Representations, 2020

SAMSUNG  
Research



DNN

$p(y | x)$



# ЗАДАЧИ МОДЕЛИРОВАНИЯ МОДЕЛЕЙ СТАТИСТИЧЕСКОЙ ФИЗИКИ

**Заказчики:** РФФИ, аспирантские исследования

**Исполнитель:**

Кафедра ПИКСиС МИЭМ НИУ ВШЭ

**Руководитель проекта:** Щур Лев Николаевич

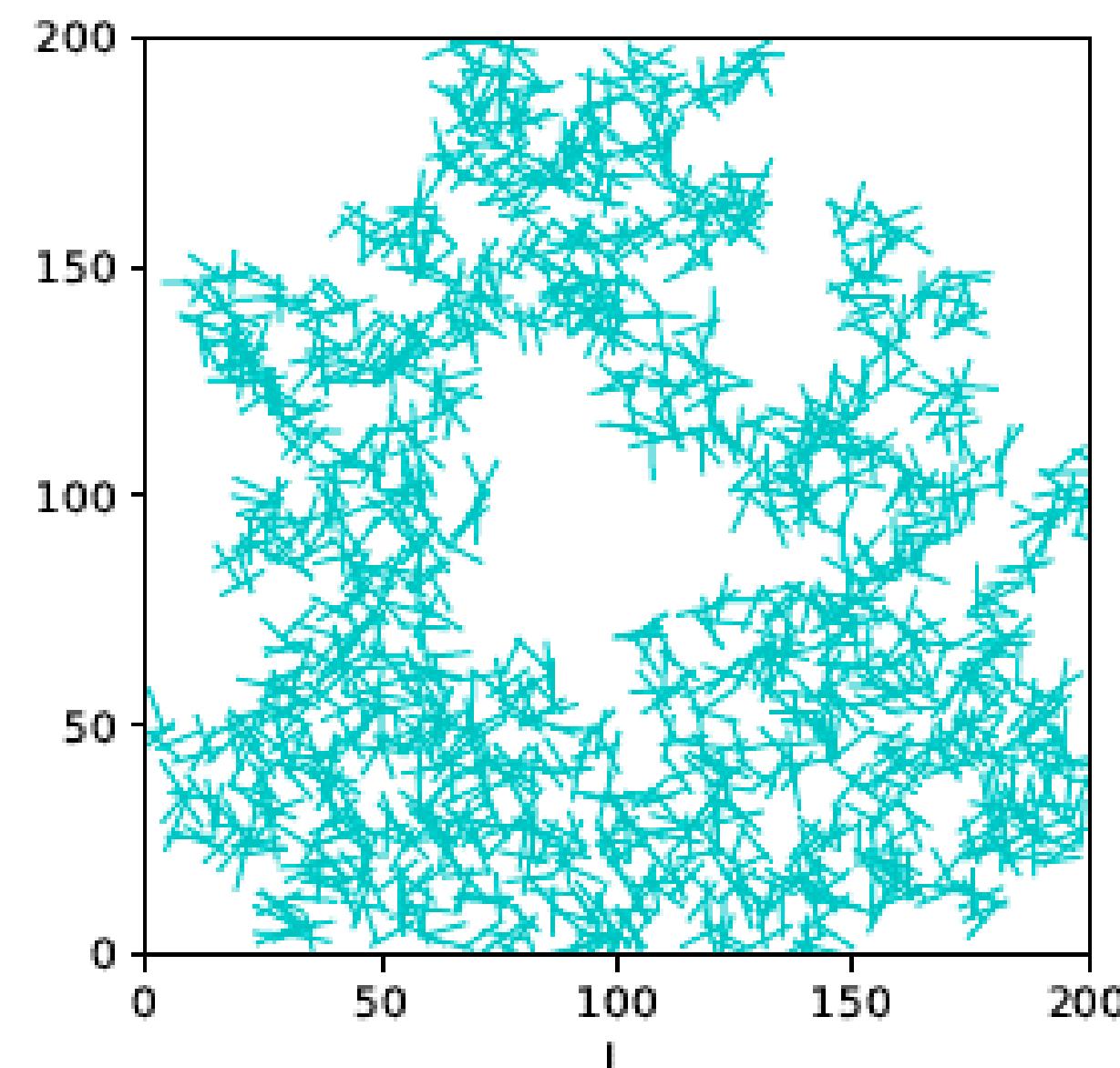
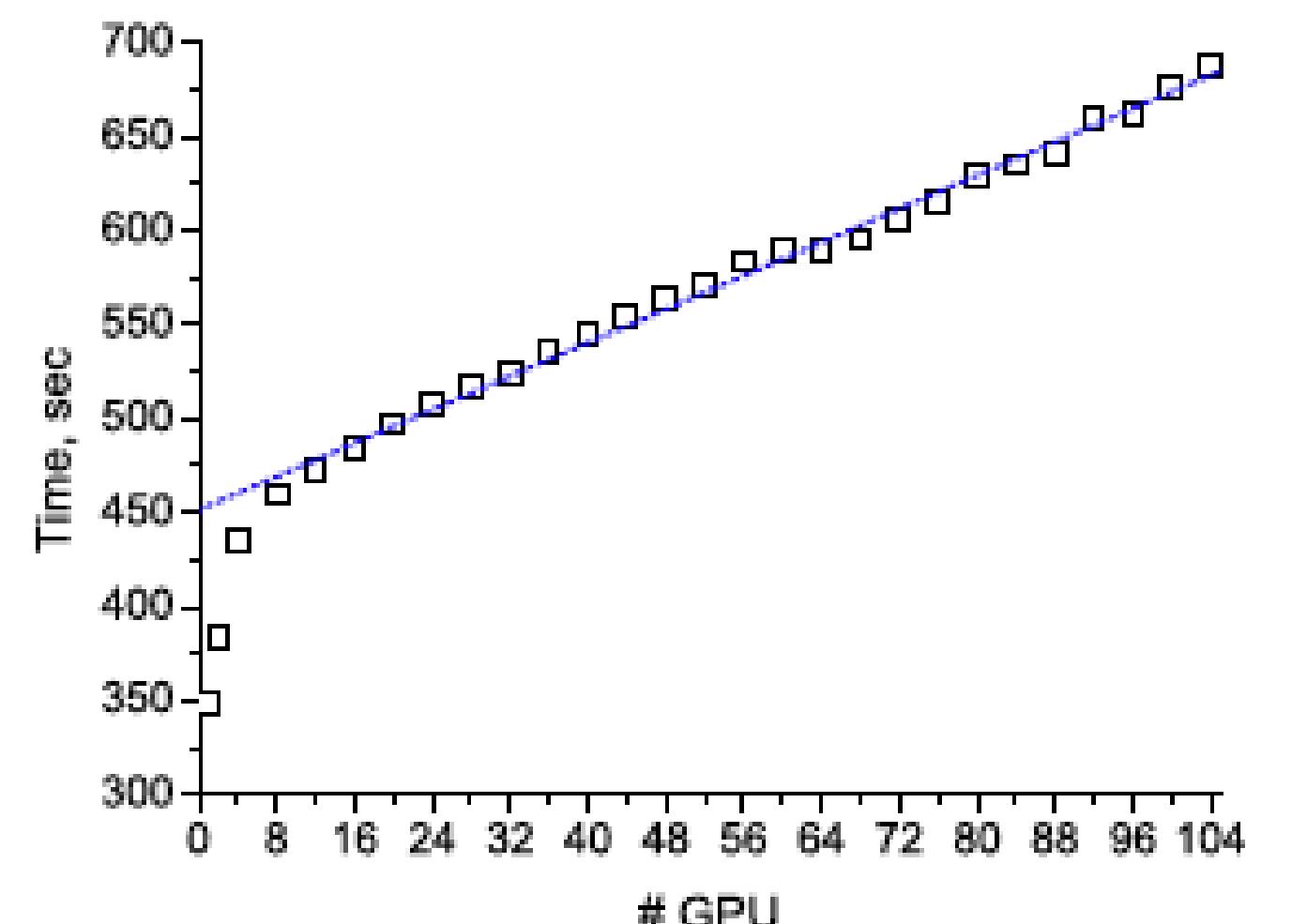
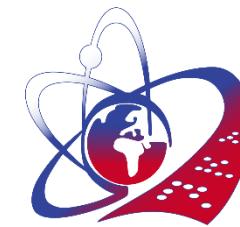
**Проект:** РФФИ 20-37-90084, РФФИ 20-37-90085

**Цель исследования:** Разработка методов моделирования задач статистической механики

**Результаты:** Разработан алгоритм реализации метода отжига популяции для гибридной архитектуры, проведено моделирование более двух миллионов реплик на кластере сCHARISMA с использование 26 узлов и 104 GPU. Разработан алгоритм генерации пенных структур в плоскости. Разработана методика исследования синхронизации в модели персональных устройств.

**Публикации:**

- Russkov A., Chulkevich R., Shchur L. Algorithm for replica redistribution in an implementation of the population annealing method on a hybrid supercomputer architecture // Computer Physics Communications, 2021 **(Q1)**
- Ziganurova L., Shchur L. Local time evolution in Personal Communication Service model // Journal of Physics: Conference Series, 2021 **(Q3)**
- Klimenkova O., Shchur L. Algorithm for foam generation in plane // Journal of Physics: Conference Series, 2021 **(Q3)**





# СТЕПЕННЫЕ ЗАКОНЫ В АНСАМБЛЯХ НЕЙРОННЫХ СЕТЕЙ

## Заказчики:

Samsung Research, Samsung Electronics,  
Российский научный фонд

## Исполнитель:

Центр глубинного обучения и байесовских методов,  
лаборатория компании Самсунг ФКН НИУ ВШЭ  
Лобачева Е.М., Чиркова Н.А., Кодрян М.С., Ветров Д.П.

**Руководитель проекта:** Ветров Дмитрий Петрович

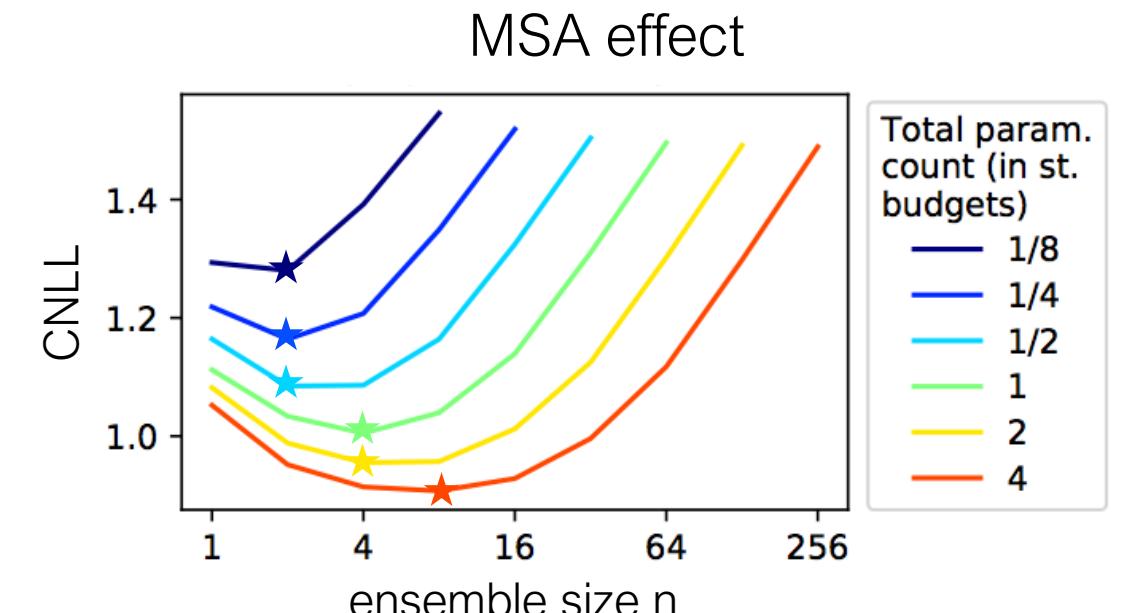
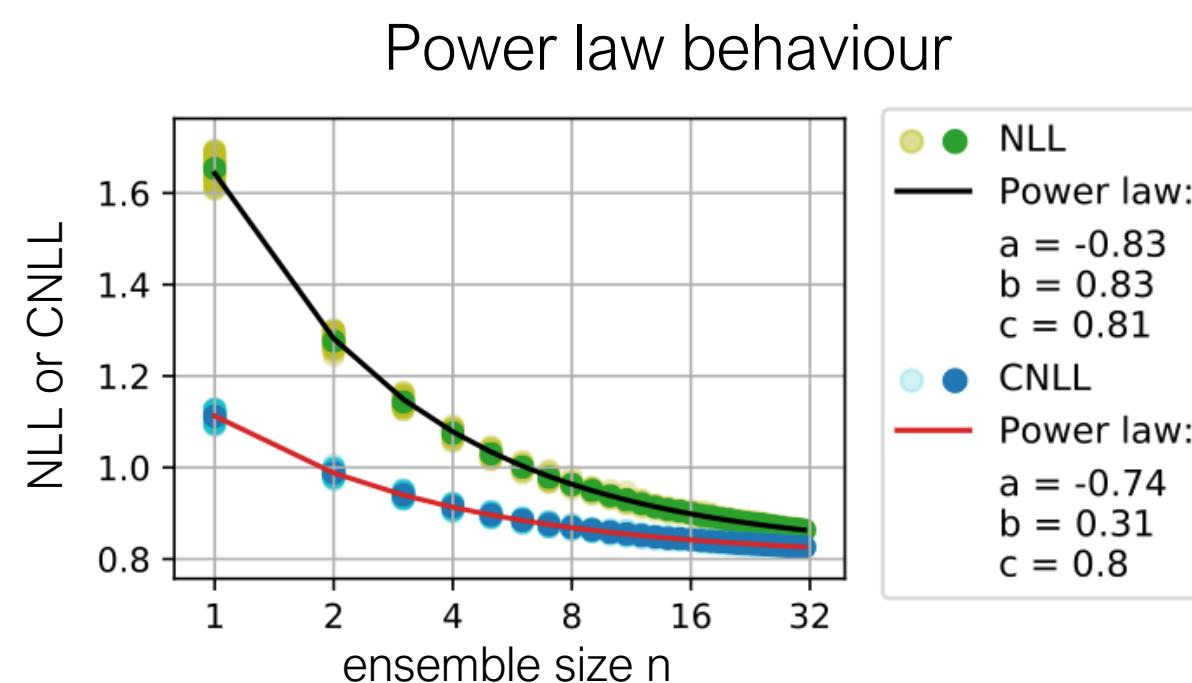
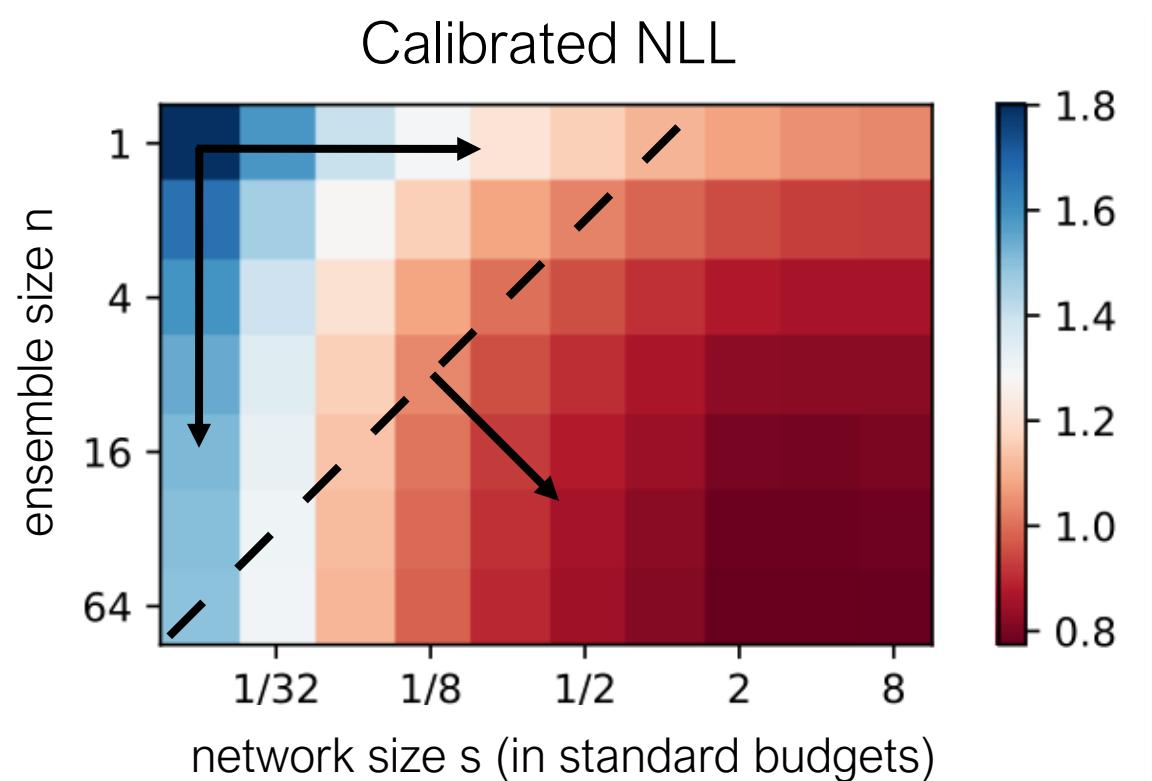
**Цель исследования:** Исследование поведения качества ансамбля нейронных сетей в зависимости от числа сетей в ансамбле и размеров этих сетей.

## Результат:

Выведены условия, при которых качество ансамбля нейронных сетей ведет себя как степенной закон. Кроме того, обнаружен эффект преимущества разделения памяти: при фиксированном бюджете памяти ансамбль из нескольких сетей среднего размера превосходит по качеству работы одну большую сеть. Используя обнаруженные степенные законы, мы можем на основе небольшого количества обученных сетей предсказывать возможный выигрыш от ансамблирования сетей с заданной структурой и оптимальное разделение памяти при заданном бюджете памяти.

## Публикации:

- Lobacheva E. et al. On Power Laws in Deep Ensembles // NeurIPS, 2020 (A\*)
- Chirkova N. Et al. Deep Ensembles on a Fixed Memory Budget: One Wide Network or Several Thinner Ones? // arXiv:2005.07292, 2020





# САЙТ ОТДЕЛА СУПЕРКОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

RU EN

## Отдел суперкомпьютерного моделирования

Национальный исследовательский университет «Высшая школа экономики» → Отдел суперкомпьютерного моделирования

Отдел суперкомпьютерного моделирования НИУ ВШЭ создан в 2019 году для администрирования суперкомпьютера НИУ ВШЭ и методической поддержки пользователей по части применения суперкомпьютерных вычислений.

**Пользователям**

S U P R O J E C T

Регистрация  
Правила  
Инструкции  
Заявка на техническую поддержку

**О нас**

DELL EMC

Об отделе  
Вычислительные ресурсы  
История изменений конфигурации НРС-клUSTERа

### Новости

**Суперкомпьютер НИУ ВШЭ удвоил свою производительность**



**КОНФЕРЕНЦИИ**

5 октября, 17:00  
Конференция пользователей Суперкомпьютерного комплекса НИУ ВШЭ

[Все анонсы >](#)

**Поддержка пользователей**



✓  
Наши ошибки?  
Выделите её, нажмите Ctrl+Enter и отправьте нам уведомление. Спасибо за участие!  
Сервис предназначен только для отправки сообщений об орфографических и пунктуационных ошибках.

<https://hpc.hse.ru>