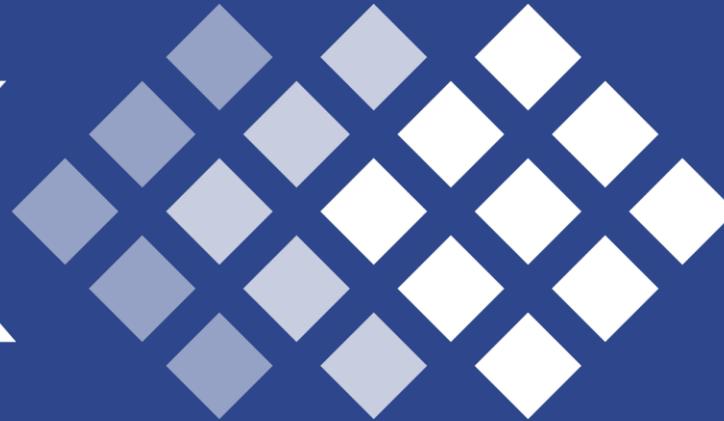


РСК

Группа компаний



**Развитие компонентных решений РСК
для высокопроизводительных систем и ЦОД**

15 лет опыта

создания и эксплуатации
высокопроизводительных систем

Разработка инновационных, энергоэффективных,
высокопроизводительных и высокоплотных
вычислительных систем для решения уникальных задач

Области применения



Искусственный интеллект
(ИИ)



Машинное обучение (ML)



Супер-ресурсоемкие расчеты



Моделирование сложных
объектов



Цифровые двойники



Обработка больших данных
и др.

О группе компаний PCK

Ведущий российский разработчик и интегратор инновационных суперкомпьютерных решений



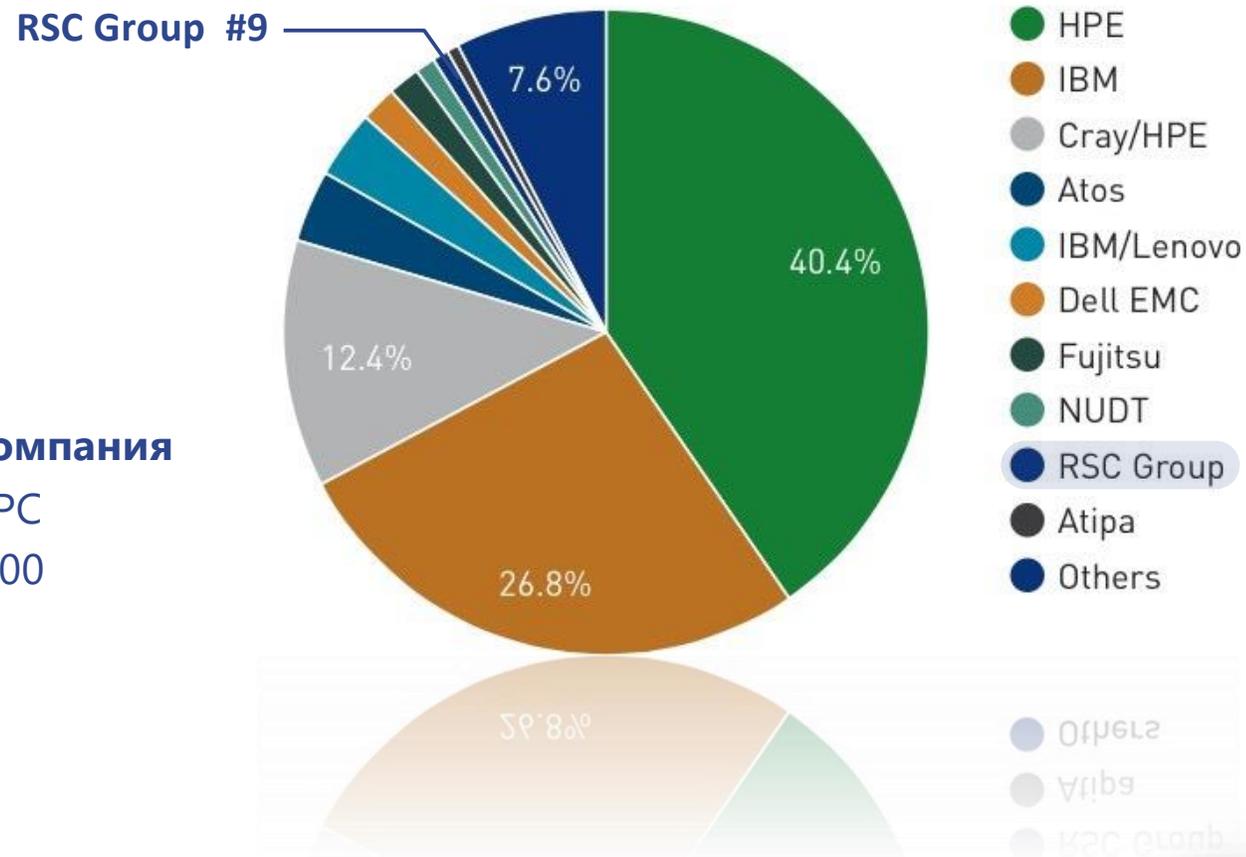
Russian DC Awards 2020 в номинации «Лучшее ИТ-решение для ЦОДа»



Единственная российская компания в мировом рейтинге Top10 HPC Vendors System Share by Top500 (Ноябрь 2014)*

* Топ 10 поставщиков по объему рынка
www.top500.org/statistics/list/

Vendors System Share



Ведущий российский разработчик и интегратор инновационных суперкомпьютерных решений



Joint Institute for Nuclear Research

Наиболее энергоэффективная система в России



Больше 70% всех российских систем **в мировом рейтинге HPCG**



3 системы PCK – единственные представители России **в мировом рейтинге IO500**



24% доля в российском рейтинге Top50

Решения РСК для развития вычислительной ИИ-инфраструктуры

Сервер «РСК Экзастрим ИИ» с 8-ю GPU

Решение «РСК Экзастрим ИИ»



Сервер «РСК Экзастрим ИИ» 2U: 208/408 TFLOPS (FP64/TF64) до 2 ТБ ОЗУ, до 128 ТБ SSD
21 сервер в шкафу «РСК Экзастрим» 42U: 4,368/8,568 PFLOPS (FP64/TF64), 115 кВт на шкаф

Параметр	Значение
Поддержка ЦПУ	Intel Xeon Scalable 4-го или 5-го поколения
Оперативная память	DDR5, 32 слота, до 2 ТБ
Поддержка ГПУ	До 8-ми NVidia H100/H200 Pairwise NVLink
Постоянная память	До 8-ми SSD рулеров в форм-факторе EDSFF 1.S, суммарно 128 ТБ
Коммутация	До 4-х портов Infiniband HDR/NDR
Сеть	10 ГБ Ethernet
Система охлаждения	100% жидкостное охлаждение, температура 40-50 °С
Блок питания	Разработка РСК, прямое жидкостное охлаждение

Суперкомпьютер «Говорун» – модернизация 5 этап

2024-2025 гг.



- 2 новых узла
«РСК Экзастрим ИИ»
(8 карт Nvidia H100 в каждом)
- 2 новых узла хранения
RSC Tornado AFS (2x1 ПБ)
- Прирост производительности
416 Тфлопс (37%)
- **2,2 ПФлопс** - суммарная
производительность системы

Конфигурация серверов «РСК Экзастрим ИИ», установленных в ОИЯИ:

- высота узла **2U**,
- **два процессора Intel Xeon Platinum 8468** (4-го поколения, 48 ядер, тактовая частота 2,1-3,8 ГГц, объем кэш-памяти 105 МБ),
- **8 графических ускорителей NVidia H100** (PCIe, 80 ГБ),
- **1 ТБ** оперативной памяти,
- **16 ТБ** емкости хранения данных на базе SSD-дисков с интерфейсом NVMe,
- **4 блока питания** производства РСК,
- **система прямого жидкостного охлаждения** разработки РСК



Внешний массив PCIe-коммутации RSC ScaleStream-C для GPU/TPU-карт

Решение RSC ScaleStream-C (JBOG)

Внешний массив PCIe-ускорителей оптимизирован для установки самых производительных GPU/TPU-карт и предназначен для расширения возможностей серверов с целью решения сложных задач в областях высокопроизводительных вычислений, машинного обучения и ИИ.

Массив поддерживает до **10-ти PCIe x16 ускорителей** с возможностью использования карт разной ширины и объединения их мостами, а также подключения до **4-х серверов** через внешние кабели на базе стандарта **PCIe x16 Gen4**.

Достигаемая производительность:

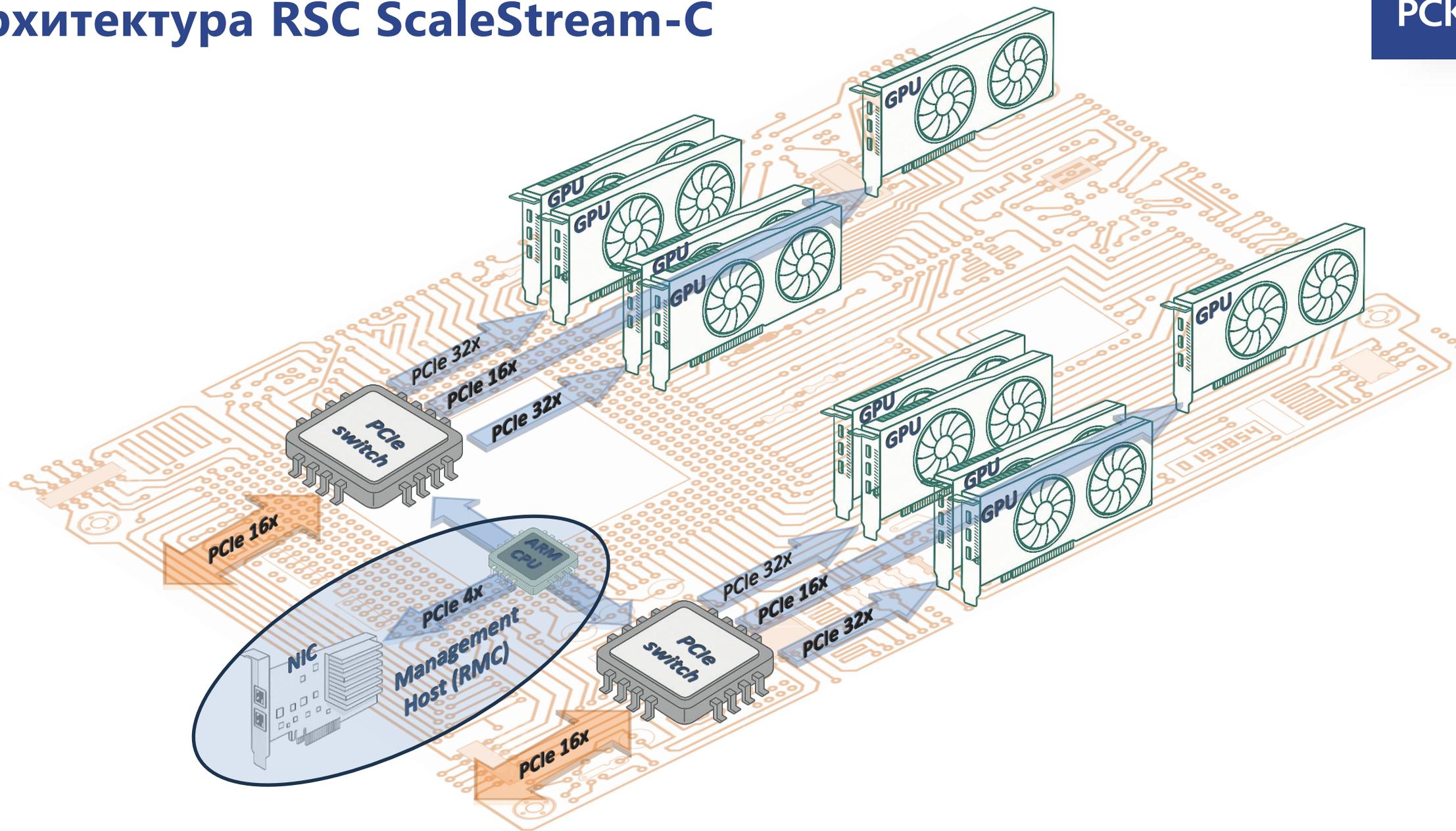
До **300 ТФлопс (FP64)** на массив из 10-ти карт NVIDIA H200.
До **960 TOPS (int8)** на массив с 10-ю картами LinQ HPQ.



Демонстрация работы массивов RSC ScaleStream-C в составе ПАК LinQ НРС и с ИИ-ускорителями LinQ НРQ (совместно с «ХайТэк») на «Микроэлектроника 2025»



Архитектура RSC ScaleStream-C



Варианты использования

Модернизация : доустановка GPU к серверам

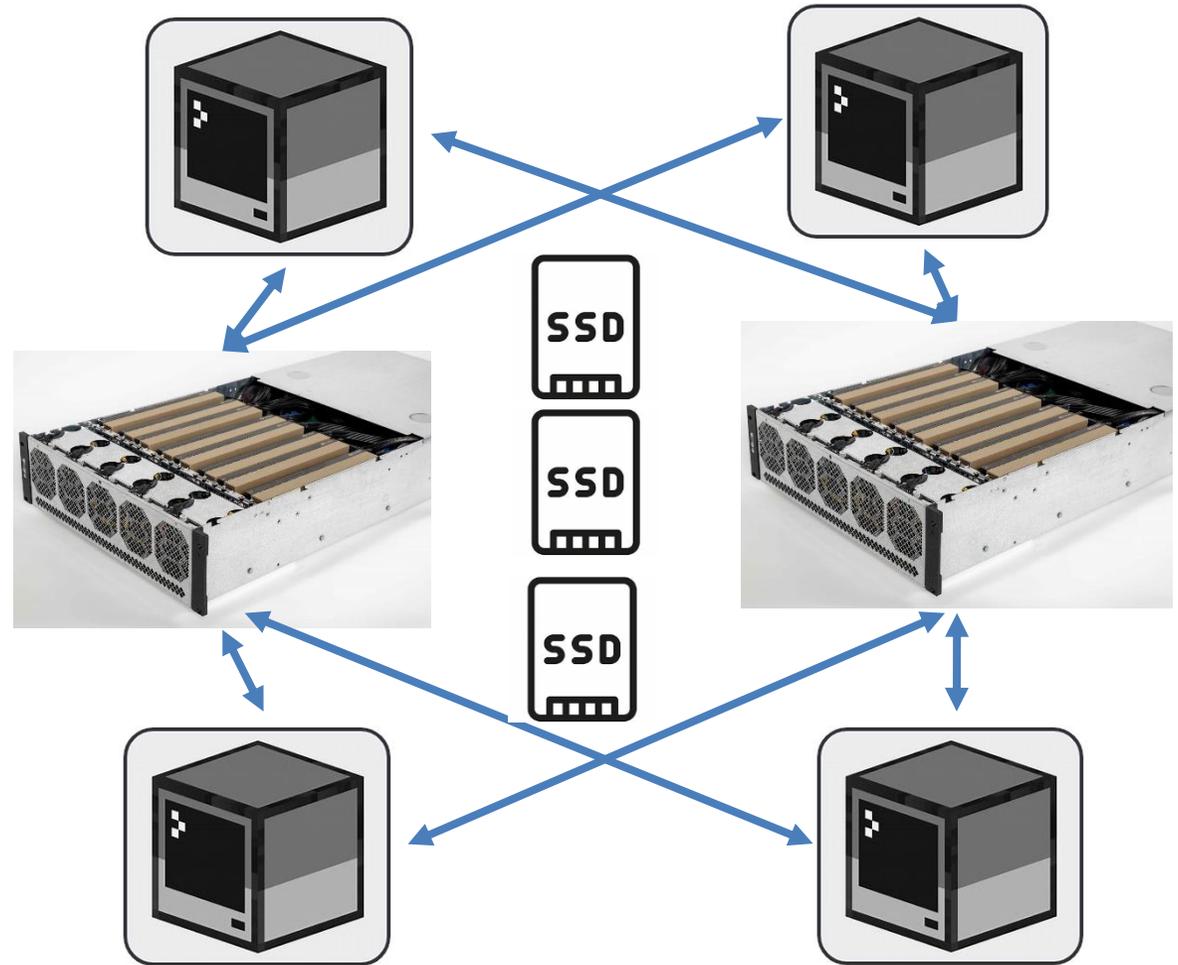
Производительность: наращивание числа GPU/TPU на сервер, если нагрузка позволяет (например, перебор паролей - 16 GPU на сервер).

Отказоустойчивость: возможность подключения GPU-ресурсов к нескольким серверам.

Гибкость: возможность подключения GPU-ресурсов к нескольким серверам динамическое переключение.

Избегание «**Задушенных GPU**» - задаче требуется 2 GPU, она занимает сервер с 4 GPU.

Сочетание с **NVMe SSD** - **ПСК БазИС**



От хаоса к прозрачности: GraphQL для унификации управления НРС-инфраструктурой

В рамках этого мастер-класса мы расскажем о том, как использовать GraphQL для унификации управления НРС-инфраструктурой.

Мы покажем, как:

- Объединить разнородные источники данных и уровни управления в единую модель. (Это особенно важно для деагрегированных сред)
- Повысить прозрачность и понимание сложностей НРС-инфраструктуры.
- Как запросы к графу отражают смысловые связи, а не структуру API.
- Использовать ИИ (LLM) как проводник и интерпретатор в сложной системе
- Использовать ИИ (LLM) для анализа текущего состояния инфраструктуры

Спасибо!



rscgroup.ru
hq@rsc-tech.ru