

Внешний массив GPU/TPU-карт 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

RSC ScaleStream-C (JBOG) представляет собой **внешний массив с коммутатором** для подключения GPU/TPU-ускорителей, являясь ключевым компонентом вычислительных комплексов. Он предназначен для организации высокопроизводительных систем, состоящих из одного или нескольких серверов и внешних ускорителей. 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**.



Сделано в РФ 

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

Преимущества:

- ♦ **Форм-фактор:** универсальное шасси для 19-дюймовых серверных шкафов высотой 3U.
- ♦ **Высокая плотность:** возможность установки до 10 современных GPU/TPU-ускорителей PCIe x16 в корпус высотой всего 3U.
- ♦ **Эффективность:** имеется возможность использования гибридного охлаждения (воздушного и жидкостного) GPU/TPU ускорителей для эффективного теплоотвода.
- ♦ **Высокая надежность:** возможность подключения GPU-ресурсов к нескольким серверам обеспечивает отказоустойчивость.
- ♦ **Мониторинг и управление:** широкие возможности мониторинга и гибкие настройки позволяют создавать различные PCIe-конфигурации с GPU/TPU и подключенными серверами.
- ♦ **Сохранность инвестиций:** независимый жизненный цикл ключевых компонентов (серверов и GPU/TPU-ускорителей) позволяет обновлять их независимо.

Сферы применения:

- ♦ **Машинное обучение и ИИ:** обучение моделей нейросетей и инференс.
- ♦ **Высокопроизводительные вычисления (HPC):** научные исследования и симуляции.
- ♦ **Анализ больших данных:** обработка и анализ крупных наборов данных (Big Data).
- ♦ **Виртуализация:**
 - GPU/TPU-ресурсов.
 - графического рабочего пространства (VDI).
- ♦ **Криптография и блокчейн:** майнинг криптовалют и задачи распределенных реестров.



Технические характеристики

Форм-фактор	Высота 3U, установка в стандартный 19-дюймовый шкаф 42U (RSC ScaleStream)
Количество графических ускорителей	До 10-ти графических процессоров Nvidia H100 или H200 NVL с интерфейсом PCI Express, до 10-ти карт PCIe (до 600 Вт)
Связь между графическими ускорителями (GPU Link)	Высокоскоростные соединения NVLink от Nvidia для объединения видеокарт в вычислительный кластер
Пиковая производительность для GPU- ускорителей	300 Тфлопс (FP64, non-Tensor) на 1 массив с 10-ю картами Nvidia H200
Пиковая производительность для TPU- ускорителей	до 960 TOPS (int8) на 1 массив с 10-ю картами LinQ HPQ
Типы поддерживаемых устройств	GPU, TPU, FPGA, NIC и другие стандартные PCIe-устройства (карты расширения)
Размеры карт расширения	Любые карты PCIe с форм-фактором «двойная ширина»
Максимальное энергопотребление на один GPU, Вт, не более	600
Количество подключаемых серверов к одному массиву	До 4-х
Интерфейс хоста-сервера	В комплект входят 4 карты HBA (host bus adapter) для серверов
Сетевой интерфейс управления (Ethernet)	Контроллер BMC с портом Ethernet (RJ45) для управления JBOG и питанием GPU/TPU
Система управления и мониторинга	Redfish, RESTful API, GUI разработки РСК
Функции управления	<ul style="list-style-type: none">◆ Компонуемость TPU/GPU на шине PCIe◆ Поддержка режима «горячей замены» GPU◆ Поддержка технологии GPU-Direct, позволяющей двум или более графическим процессорам напрямую обмениваться данными◆ Конфигурирование PCIe-порта к портам серверов и устройств◆ Конфигурируемое виртуальное дерево PCIe-сервера◆ Просмотр с сервера назначенных TPU/GPU-карт◆ Управление аутентификацией пользователей◆ Управление электропитанием PCIe-слотов

Технические характеристики

Функции мониторинга	<ul style="list-style-type: none">◆ Информация о состоянии графического процессора (температура, загрузка, скорость работы и использование памяти)◆ Мониторинг производительности и ошибок работы портов серверов и устройств◆ Защита устройства от перегрева◆ Мониторинг связи и статус устройств
Тип охлаждения	Воздушное (для TPU-карт) или гибридное (для GPU-карт)
Модули охлаждения	5 высокопроизводительных двухроторных вентиляторов с режимом «горячей замены»
Блоки питания (PSU)	4 блока питания мощностью 2,2 кВт, с выходным напряжением 12 В
Электропитание	230 В переменного тока, 8,8 кВт
Размеры сервера (ШхВхГ), мм	483x132x927
Вес массива (без GPU), не более, кг	25

