GPU-сервер 8×H100 (HGX SXM5)

Сервер обучения и инференса большого масштаба: 8 GPU H100 80 ГБ, NVLink, до 2 ТБ ОЗУ.

Описание

Флагманская платформа для обучения и дообучения LLM, работы с мультимодальными моделями и крупной инференс-нагрузки. Строится на базе HGX-платформы с 8 графическими ускорителями H100 80 ГБ и NVLink-соединением между ними.GPU: 8× H100 SXM5 80 ГБ (NVLink)CPU: 2× Xeon Scalable, 64+ ядра каждыйОЗУ: до 2 ТБ DDR5 ECCХранение: NVMe RAID, до 61 ТБСеть: 2×100G для кластерированияПоставка с пусконаладкой, настройкой SLURM/Kubernetes и гарантией на компоненты. Возможна аренда пилотного образца до закупки.

Технические характеристики

GPU	8× H100 SXM5 80 ГБ (NVLink)
CPU	2× Xeon Scalable, 64+ ядра каждый
ОЗУ	до 2 ТБ DDR5 ECC
Хранение	NVMe RAID, до 61 ТБ
Сеть	2×100G для кластерирования

Характеристики могут обновляться производителем. Финальная конфигурация фиксируется в спецификации счёта.