ИИ-СЕРВЕРЫ И ASCEND

Atlas 800I A2, инференс

4U-сервер инференса с 8 модулями Ascend AI. Рассчитан на работу генеративных больших моделей: умная поддержка клиентов, генерация текстов, накопление знаний.



Ключевые характеристики

Форм-фактор	4U, сервер ИИ
NPU	8 × модуль Ascend AI
Процессор	4 × Kunpeng 920
Память	32 слота DDR4, до 3200 MT/s; модули 16 / 32 / 64 GB
Локальные диски	варианты: 8 × 2,5″ SATA + 2 × 2,5″ NVMe; 4 × 2,5″ SATA + 6 × 2,5″ NVMe
RAID	0 / 1 / 10 / 5 / 50 / 6 / 60
Сеть	8 × 200GE QSFP напрямую, протокол RoCE
Межсоединение NPU	8 NPU по схеме HCCS Full mesh, двусторонняя пропускная способность 392 GB/s
Энергоэффективность	0,4 PFLOPS на kW
PCIe	до 3 слотов PCIe 4.0
Питание	4 блока 2,6 kW с горячей заменой, 2+2; 200 – 240 V AC, 240 V DC
Охлаждение	воздушное, 8 вентиляторных модулей с горячей заменой, N+1
Рабочая температура	5 – 35 °C
Габариты	175 мм (В) × 447 мм (Ш) × 790 мм (Г)

Для каких задач

генерация контента · вопрос-ответ · умная поддержка клиентов · рекомендации в интернет-сервисах · модерация контента · поиск и обработка текста (NLP)

Запросить КП: +7 (495) 718-20-10, info@huaweirussia.ru

Срок подготовки КП: 3 рабочих дня. Цена — в КП, с учётом объёма и условий оплаты.
Характеристики по данным Huawei. Лист подготовлен Renomme Group