Плата инференса Huawei Atlas 300I Duo (PCIe)

Atlas 300I Duo: PCIe-карта инференса на Ascend — видеоаналитика и LLM в существующий сервер. В РФ возят единицы, ниша свободна.

Описание

Atlas 300I Duo — двухпроцессорная PCIe-плата инференса Huawei на архитектуре Ascend: до ~280 TOPS INT8 (заявлено вендором) для видеоаналитики, OCR и LLM-инференса внутри уже стоящего у заказчика сервера. Вместо покупки нового AI-сервера за десятки миллионов — модернизация существующего за сотни тысяч рублей.Именно этот класс устройств отчёт по рынку называет свободной полкой: серверы Atlas возят все, платы и edge-платформы 200/300/500 — единицы. Спрос: вузы, НИИ, «инференс у станка», инференс 36% GPU-задач рынка.Мы поставляем с установкой, настройкой драйверов CANN, конвертацией ваших ONNX-моделей и гарантией. Расчёты ₽/¥, поставка из Китая. Цена по запросу.

Технические характеристики

Ускоритель	Ascend 910B-class, 2 ядра на плате
Производительность	до ~280 TOPS INT8 (заявлено вендором)
Интерфейс	PCIe 4.0 x16, двухслотовая, пассивное охлаждение
Память	LPDDR4X (по спецификации ревизии)
Стек	CANN, MindX, конвертеры ONNX/PyTorch
Установка и конвертация моделей — в комплекте, цена по запросу	

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