Сервер для обучения LLM и тяжёлых AI-задач (4× NVIDIA H100)

📌 Назначение

Флагманское решение для:

- обучения и дообучения больших языковых моделей (LLaMA, Falcon, Baichuan)
- генеративного ИИ (видео, 3D, мультиязычные ассистенты)
- корпоративных ML-инфраструктур и дата-центров

Подходит R&D-командам, системным интеграторам, крупным ИТ-компаниям.

⚙️ Конфигурация оборудования

- **Форм-фактор:** 4U (стойка, активное охлаждение)
- **CPU:** 2× Intel Xeon Gold
- **ОЗУ:** 1024 GB ECC DDR4/DDR5
- **Хранение:** 4× NVMe 2TB + 2× SATA 4TB
- **GPU:** 4× NVIDIA H100 PCIe 80GB
- **Сеть:** 2× 25GbE (Mellanox ConnectX-6)
- **Дополнительно:** Software RAID, BMC/IPMI, HBA Broadcom 9500-16i

💻 Предустановленное ПО

- Rocky Linux 9 или Ubuntu 22.04 LTS
- NVIDIA CUDA Toolkit + драйверы для H100
- PyTorch с Apex, Megatron-LM, FSDP, DeepSpeed (по запросу)
- NVIDIA Triton Inference Server
- Slurm / Kubernetes (опционально)

💼 Сервис и сопровождение

- Предтестирование под нагрузкой и настройка BIOS/режимов питания
- Технический аудит и рекомендации по архитектуре
- Поддержка по драйверам и оптимизациям — 6 месяцев бесплатно

💰 Стоимость

- **Цена (DDP Москва + ПО):** 9 238 732 ₽
- Гарантия 3 года: +539 488 ₽
- Гарантия 5 лет: +944 104 ₽
- Настройка и сопровождение (если отдельно): 809 232 ₽

🧠 Почему выбирают нас?

- Работа напрямую с ODM (Inspur, Supermicro), оптимальные цены
- Тестирование под ваши модели до покупки
- Гибкость в BIOS/OS/driver настройках
- Работаем с НДС, документы, спецификации, техпаспорт

📦 Условия поставки

- Срок: 4–5 недель после оплаты
- Сервер поставляется готовым к запуску
- Транспортировка в промышленной упаковке (ударозащита, влагозащита)
