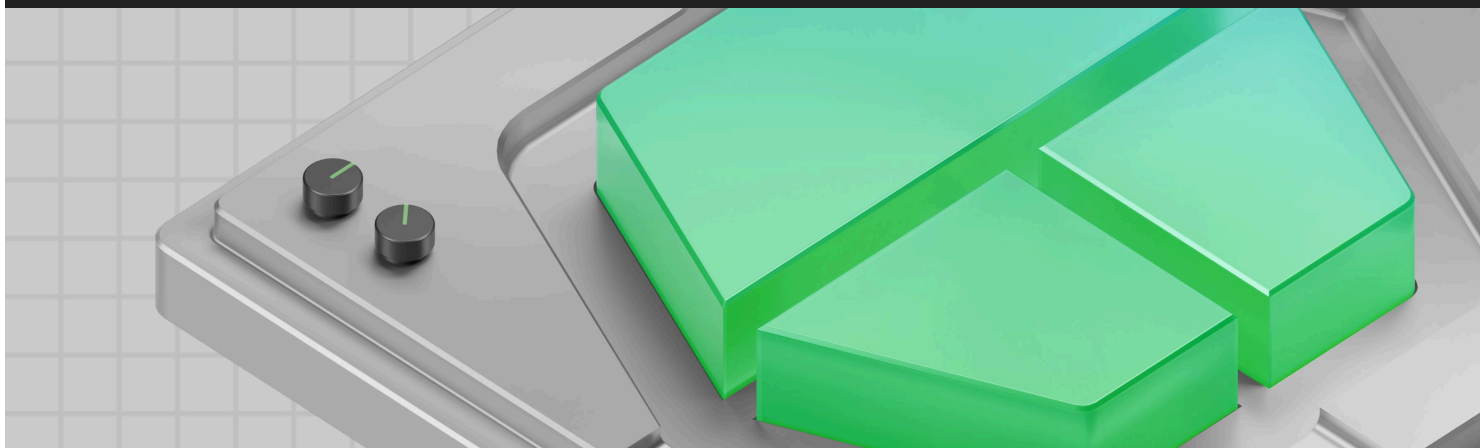


Cloud.ru Evolution Stack.ML

Устанавливается в ЦОДе заказчика

Гибридная платформа для ML-разработки и совместной работы Data Science-команд над созданием и тюнингом моделей машинного обучения



Проблемы ML-разработки

- Дефицит и низкая загрузка GPU
- Сложность развёртывания и подготовки окружения
- Прерывание задач, потеря прогресса и времени, рост затрат
- Ограничения ИБ по использованию облачных сервисов

Evolution Stack ML

- Развертывание решения полностью в ЦОДе заказчика «под ключ»
- Умный планировщик, резервирование и автоматическое распределение свободных GPU-ресурсов
- Автоматическое восстановление задач, процессов и сервисов при аппаратных и программных отказах

Ценность

- Запуск готовой on-premise ML-платформы в минимальные сроки
- Повышение утилизации GPU до 80–90%
- Существенное ускорение проверки гипотез и запуска решений

Опыт Cloud.ru

Подтвержденные кейсы до 95% утилизации по году

Кластеры 10 000+ GPU

Задачи на 1 000+ GPU

200+ команд на 1 кластер

Поставка
Cloud.ru

Cloud.ru Evolution Stack



Личный кабинет

OSS/BSS



Billing



IAM



LogAAS



MonAAS



AuditLog



Notifications



Managed Kubernetes



IaaS

ML services



Data Transfer



NFS



Jobs



Jupyter Notebooks



Model Monitoring



Experiments



Deploys



Allocations



Kubernetes

Workers



Worker Node



Worker Node



Worker Node



Worker Node

Host OS



CPU



NVME

Host OS



CPU



CPU



NFS

Инфраструктура
заказчика

Функции платформы

- Запуск задач распределенного обучения
- Запуск сред разработки Jupyter Notebook с готовыми окружениями
- Управление ресурсами (GPU, CPU, NFS). Контроль утилизации и автоматизированные средства управления утилизацией
- Интерфейсы: UI, API, Python SDK, CLI
- Self-healing задачи (контроль выполнения, авто-замена GPU-нод в случае аппаратных сбоев, контроль зависания, анализ ошибок)
- Управление планировщиком (очереди, спотовые инстансы, приоритеты, preemption режим)
- Командные пространства с shared и private ресурсами (GPU, CPU, NFS)
- Мониторинг обучения и управление экспериментами (MLflow, TensorBoard)
- Набор готовых образов для Jobs и Jupyter Notebooks с настроенными окружениями
- Инструменты observability: MonAAS, LogAAS
- OSS и BSS слой в составе облачной платформы Cloud.ru Evolution Stack

Внедрение

Комплекс услуг по проектированию, пилотированию, поставке, монтажу, настройке и вводу в коммерческую эксплуатацию.

Гарантия, поддержка, SLA

Расширенная техническая поддержка 24/7. Опция гарантированного SLA.

Получить
консультацию
эксперта

