



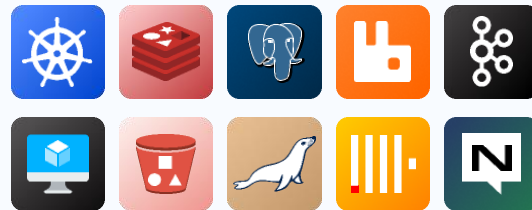
KATAMARINA

Управление GPU и AI/ML-портал

Katamarina — готовая инфраструктура для GPU

- Managed Kubernetes с автоматическим масштабированием
- GPU-ноды (виртуалки) продвигаются динамически по требованию
- Автоматический менеджмент и установка драйверов в VM
- Готовый мониторинг и дашборды на базе Victoria Metrics и Grafana
- Модуль подсчета использования ресурсов с разбивкой по тенантам

PaaS



IaaS



Множество сервисов доступны по клику



Managed Kubernetes



Managed Databases



Virtual Machines



Persistent Volumes



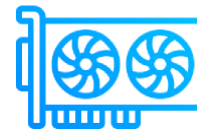
S3 Storage



Load balancers



Monitoring and Alerts



GPU devices

AI-Ready-инфраструктура: GPU для Kubernetes и VMs



Инфраструктурный слой GPU

Пулы ресурсов GPU, тайм-слийдинг, квоты по тенантам и проектам



Единый 스케дулинг

Единый планировщик и система управления для подов и VMs с метриками использования, на основе которых можно биллить клиентов и проводить глубокую аналитику



Данные и пайплайны

S3 storage и БД и AI/ML-моделей, пайплайны с автоматизацией на базе GitOps

Полный жизненный цикл для управления GPU



Automated GPU Provisioning



GPU Passthrough (VM and Kubernetes)



GPU Driver Management

- Полный автопилот для карт от NVIDIA
- GPU passthrough для видеокарт других производителей

Полный жизненный цикл для GPU:

- ▾ Автоматическая установка драйверов и управление GPU-ресурсами в Kubernetes
- ▾ Управление жизненным циклом GPU, resource-management
- ▾ Автоскейлинг и провиженинг ресурсов по требованию
- ▾ Безопасность и мульти-тенантность
- ▾ Вывод из эксплуатации и rolling upgrade



Marketplace

Marketplace

IaaS

Virtual Machines

Kubernetes

VM Instances

VM Disks

Buckets

PaaS

ClickHouse

FerretDB

Kafka

MySQL

NATS

PostgreSQL

RabbitMQ

Redis

NaaS

HTTP Cache

TCP Balancer

VPN

Administration

Info

Cluster: default

Namespace: tenant-kvaps

Available Products

Search

Marketplace

Edit Mode



Bucket

storage

S3 compatible storage



ClickHouse

Managed ClickHouse service



FerretDB

database

Managed FerretDB service



HTTP Cache

cache network

Layer7 load balancer and caching service



Kafka

messaging

Managed Kafka service



Kubernetes

compute

Managed Kubernetes service



MySQL

database

Managed MariaDB service



NATS

messaging

Managed NATS service



PostgreSQL

database

Managed PostgreSQL service



RabbitMQ

messaging

Managed RabbitMQ service



Redis

cache

Managed Redis service



TCP Balancer

network

Layer4 load balancer service



Virtual Machine

compute

Virtual Machine (simple)



VM Disk

storage

Virtual Machine disk



VPN

network

Managed VPN service



- Marketplace
 - Marketplace
- IaaS
 - Virtual Machines
 - Kubernetes
 - VM Instances
 - VM Disks
 - Buckets
- PaaS
 - ClickHouse
 - FerretDB
 - Kafka
 - MySQL
 - NATS
 - PostgreSQL
 - RabbitMQ
 - Redis
- NaaS
 - HTTP Cache
 - TCP Balancer
 - VPN
- Administration
 - Info

Cluster: Namespace:

apps.cozystack.io/v1alpha1/kuberneteses > Create

← Create apps.cozystack.io/v1alpha1/kuberneteses

OpenAPI Manual

metadata ^

name

spec ^

storageClass

version

host

nodeGroups ^

md0 ^

minReplicas*

maxReplicas*

instanceType*

ephemeralStorage*

roles

```

1  apiVersion: apps.cozystack.io/v1alpha1
2  kind: Kubernetes
3  metadata:
4    namespace: tenant-kvaps
5  spec:
6    addons:
7      certManager:
8        enabled: false
9      gatewayAPI:
10       enabled: false
11     velero:
12       enabled: false
13     fluxcd:
14       enabled: false
15     gpuOperator:
16       enabled: false
17     ingressNginx:
18       enabled: false
19       exposeMethod: Proxied
20     monitoringAgents:
21       enabled: false
22   controlPlane:
23     connectivity:
24       server:
25         resourcesPreset: micro
26     replicas: 2
27     scheduler:
28       resourcesPreset: micro
29     apiServer:
30       resourcesPreset: medium
31     controllerManager:
32       resourcesPreset: micro
33   host: ""
34   nodeGroups:
35     md0:
36       instanceType: u1.medium
37       maxReplicas: 10
38       minReplicas: 0
39       roles:
40         - ingress-nginx
41       ephemeralStorage: 20Gi

```

Что уже умеет платформа

- ✔ **Monitoring** — утилизация, нагрузка, неиспользуемые ресурсы.
- ✔ **Provisioning** — быстрое выделение GPU/vGPU для ML-проектов.
- ✔ **Multi-tenancy** — возможность создавать изолированные «личные пространства» для компаний/команд.
- ✔ **Billing & Quotas** — лимиты, тарификация, учет потребления.
- ✔ **Dynamic allocation** — перераспределение GPU при освобождении.
- ✔ **Inventory** — видеть все GPU, их локации и состояние.

Дорожная карта построения AI-портала

Architecture, pilot, migration & custom development

Фаза 1 (готово)

Поддержка GPU

- Автопилот с GPU от NVIDIA
- Passthrough для других GPU
- Мониторинг и учет ресурсов

Фаза 2 — 2 месяца

Первое MVP с AI-сервисами

- Добавление в дашборд платформы для AI-сервисов
- Поддержка GPU других производителей
- Добавление на дашборд популярных self-hosted моделей

Фаза 3 — 3 месяца

Расширение сервисов

- Улучшенная оркестрация AI-сервисов
- Добавление полноценного набора инструментов enterprise-уровня для работы с GPU
- Поддержка MIG

Фаза 4

Следующие шаги

- Тонкая подстройка AI-платформы
- Добавление новых компонентов
- Добавление автоматизации для GPU других производителей

 Infrastructure

 AI Platform MVP

 Enterprise Level AI Platform

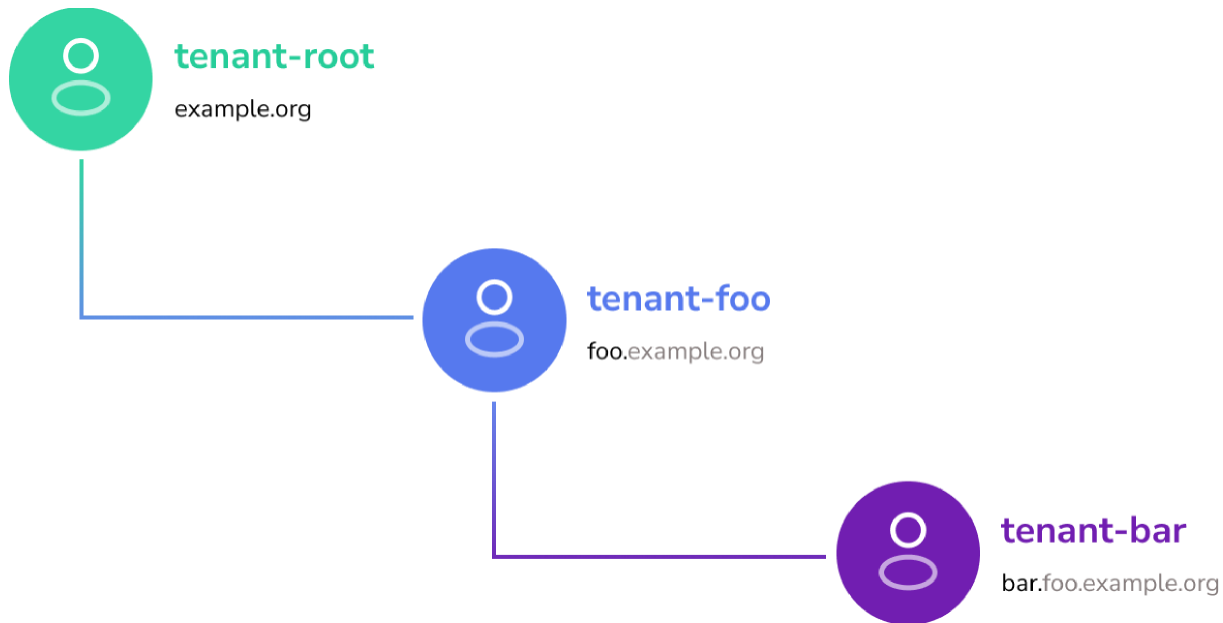
 Expansion

Преимущества выбора решения от Katamarina

- Готовая на 80% инфраструктура для AI
- Быстрый старт в продакшен, быстрое поэтапное добавление новых компонентов
- Возможность получать обновления платформы и компонентов
- Поддержка от вендора с сильной экспертизой и сообществом
- Легкая интеграция с приватным и публичным облаком холдинга

Эффективность и безопасность из коробки

Наша уникальная система тенантов позволяет эффективно использовать облачные ресурсы для control plane, оптимизировать расходы и усиливает безопасность



Катмарина основывается на технологиях Cozystack - являющимся проектом CNCF (Linux Foundation)

Это признание международного сообщества, гарантия отсутствия вендор-лока, снижение рисков для адоптеров и ускорения развития продукта за счет комьюнити.



CLOUD NATIVE COMPUTING FOUNDATION



Преимущества для энтепрайза

- Поддержка 24/7 с гарантированным SLA
- Персональный роадмап
- Выделенный канал поддержки
- Система биллинга,
- White-labeling / Air-gapped
- Курсы и обучение



Компании, которые уже используют решения на основе Cozystack (публичная информация от вендора Aenix)



Развитие продукта с учетом требований в РФ, СНГ, БРИКС

- Включено в реестр российского ПО (<https://reestr.digital.gov.ru/reestr/8093831/>)
- Использование аудированного ядра Linux от ИСП РАН
- Сертификация ФСТЭК (в процессе)
- Сертификация по требованиям стран СНГ (в планах)
- Развитие русскоязычного комьюнити и локального форка <https://cozystack.ru/>
- Локализованные курсы и обучение





« Контакты

Сайт: <https://katamarina.ru/>

Почта: info@katamarina.ru

Канал: https://t.me/katamarina_ru

Чат: https://t.me/cozystack_ru