QazCloud

VK Cloud

Безопасность хранения данных и кейсы использования Cloud ML Platform

Александр Перепелов,
технический директор QazCloud

Александр Волынский,
технический менеджер VK Cloud

Распределение ответственности в облаке

QazCloud

Модель построения системы работы с данными

Заказчик
ответственность за собственные данные

- Данные пользователей
- Клиентское ПО, приложения, система идентификации
- Серверные операционные системы
  Базы данных

Провайдер
ответственность за инфраструктуру

- Виртуализация, системное ПО, бэкапирование, репликация
- Вычислительные ресурсы, хранение, каналы связи
- ЦОД провайдера, энергетика, ИБП, охлаждение, физическая охрана

Модель построения системы работ с данными

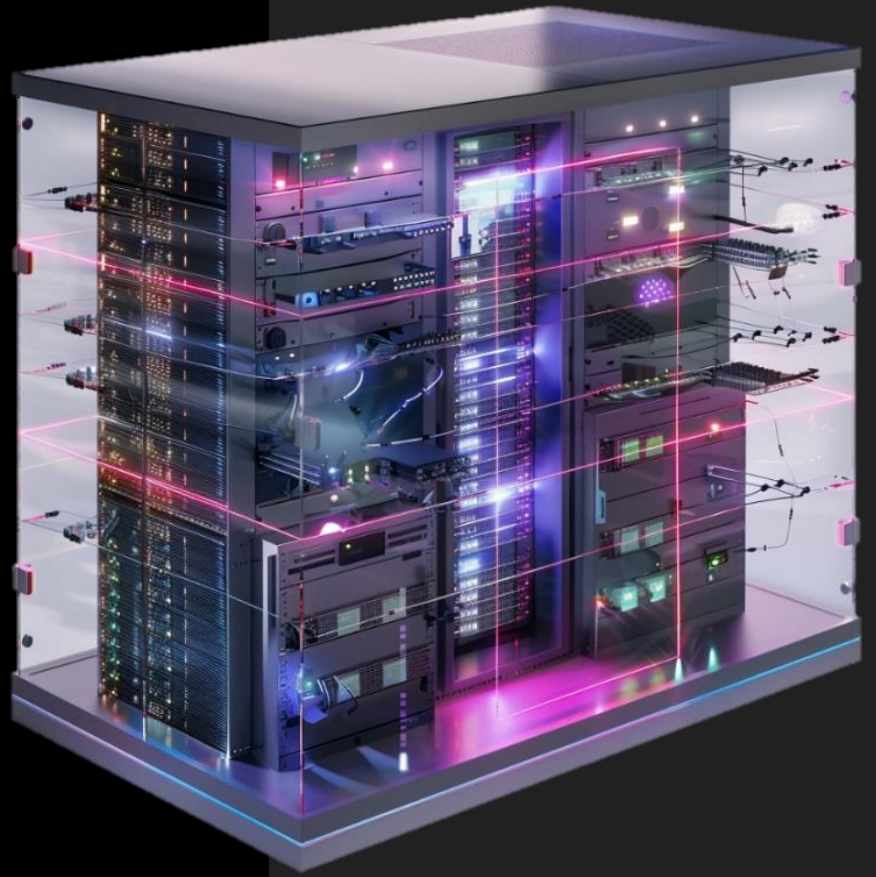
QazCloud

Выбор провайдера, с учетом архитектуры хранения данных

1. Предлагает ли Ваш провайдер два ЦОД-а?
2. Готов ли Ваш провайдер предоставлять гибридное облако?
3. Поддерживает ли Ваша Компания лучшие практики по хранению данных?

Если ответ на эти вопросы – нет,
то как вы будете поступать в случаях:

- пожара;
- землетрясения;
- физической поломки оборудования;
- хакерской атаки, шифрования.



Распределенное хранение

Варианты хранения

1. **Гибридное облако**

   Собственная инфраструктура
   + данные в облаке

2. **Бэкапирование в облаке**

   Ваши данные уже в облаке, но провайдер предоставляет услугу хранения
   в удаленном ЦОД

3. **Два и более ЦОД для размещения критически важных ИТ систем**

   Ваши данные в облаке, при этом наиболее критичные ИТ системы имеют полную копию (active-active / active-passive)
   в географически удаленном ЦОД

QazCloud

Можно ли получать выгоду от распределенного хранения?

QazCloud

Используйте резервные данные
для работы Вашей нейросети, не создавая
нагрузку на продуктивные системы.

QazCloud

Благодарю за внимание!

Александр Перепелов

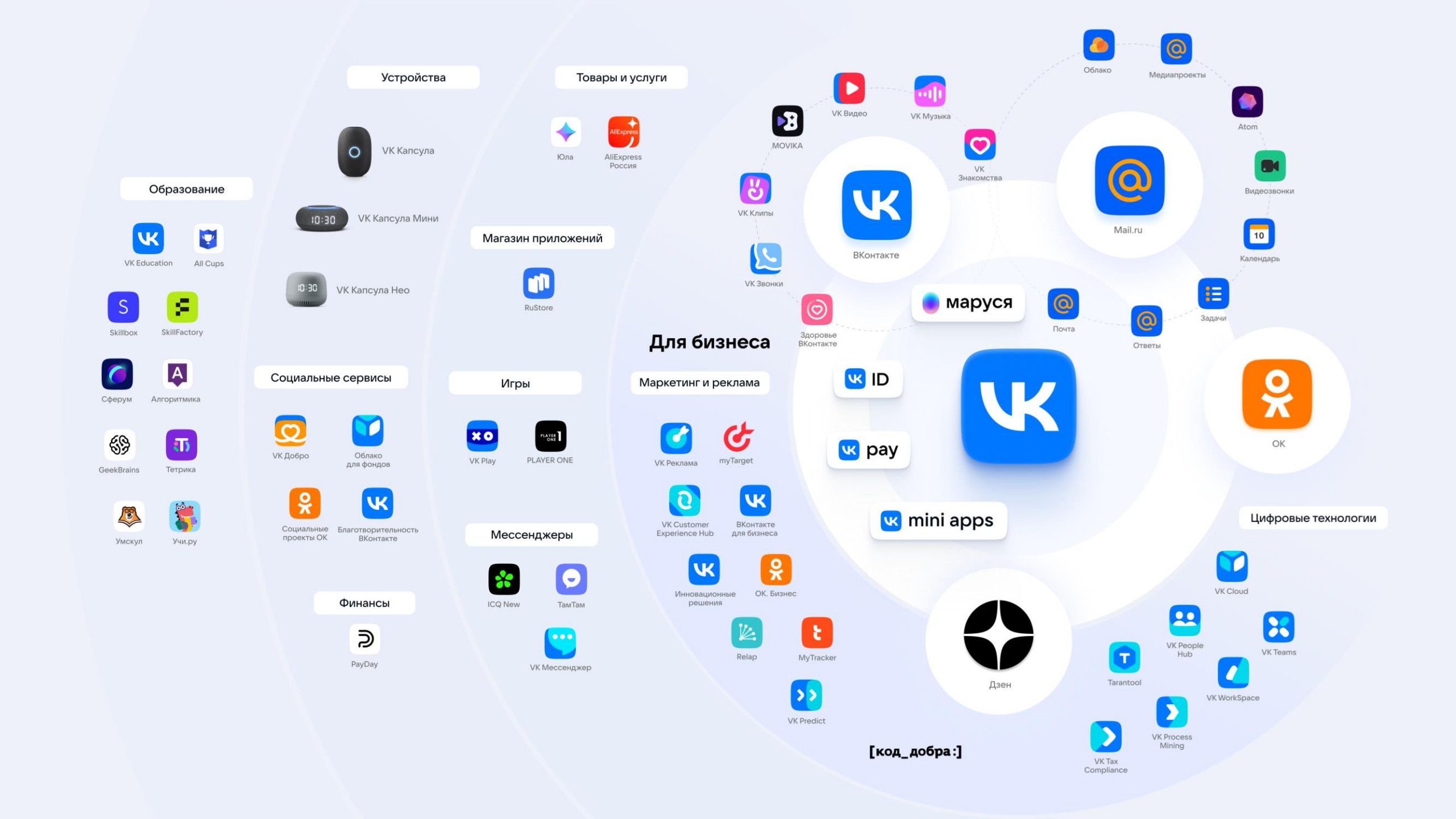
Технический директор QazCloud

VK Cloud

Cloud ML Platform: кейсы и опыт использования

Платформа для полного цикла ML-разработки
и совместной работы дата-команд

бизнес


Устройства
Товары и услуги
VK Капсула
Юла
AliExpress Россия
VK Капсула Мини
Образование
VK Education
All Cups
Магазин приложений
VK Капсула Нео
RuStore
Skillbox
SkillFactory
Сферум
Алгоритмика
Социальные сервисы
Игры
GeekBrains
Тетрика
VK Добро
Облако для фондов
VK Play
PLAYER ONE
Умскул
Учи.ру
Социальные проекты ОК
Благотворительность ВКонтакте
Мессенджеры
Финансы
ICQ New
ТамТам
PayDay
VK Мессенджер
Для бизнеса
Маркетинг и реклама
VK Реклама
myTarget
VK Customer Experience Hub
ВКонтакте для бизнеса
Инновационные решения
OK. Бизнес
Relap
MyTracker
VK Predict
Облако
Медиапроекты
Atom
VK Видео
VK Музыка
MOVIKA
VK Знакомства
Видеозвонки
VK Клипы
ВКонтакте
Mail.ru
Календарь
VK Звонки
маруся
Задачи
Здоровье ВКонтакте
Почта
Ответы
ID
OK
pay
mini apps
Цифровые технологии
VK Cloud
VK People Hub
VK Teams
Дзен
Tarantool
VK WorkSpace
VK Process Mining
VK Tax Compliance
[код_добра:]

VK Cloud

Облачная платформа
для цифровизации бизнеса
в Казахстане

7 лет

с момента создания
продукта VK Cloud

>8000

компаний строят бизнес
на решениях VK Cloud

Что мы предлагаем в Казахстане

- ЦОД в Астане, партнер QazCloud
- Соответствие закону 94-V о персональных данных
- Сертификация PCI DSS
- Юрлицо — «ВК Тех Казахстан»
- Все расчеты в тенге
- Локальная команда в Алматы и Астане

VK Cloud — фундамент для построения платформы данных

Виртуальные серверы

Конфигурации необходимой мощности
с внешними IP и трафиком 1 Гбит/с

Виртуальные сети

Объединение серверов
в локальных сетях,
публичный и приватный DNS,
VPN, балансировка
и фильтрация трафика

Объектное хранилище

Хранение и передача данных
в исходном виде
по S3 API и через UI.
От мегабайт до петабайт

Cloud Backup

Резервное копирование
облачных серверов
и баз данных

Кластеры Kubernetes

Автомасштабирование
ресурсов и ускорение
доставки приложений
с Kubernetes как сервисом

Личный кабинет

- Self-Service
- Биллинг по ресурсам, командам, проектам
- Квоты
- Права доступа
- Мониторинг
- Алертинг
- API

Базы данных

Полностью управляемые
базы данных MySQL,
PostgreSQL, Redis, ClickHouse,
Arenadata DB (аналитическая БД
на основе Greenplum),
Tarantool

Big Data

Обработка и анализ больших данных с помощью Cloud Big Data (облачная платформа на базе Arenadata DB, Arenadata Hadoop, Apache Spark, Cloud Streaming)

Машинное обучение

Полный цикл ML-разработки и совместной работы data-команд
с помощью Cloud ML Platform.
Модели компьютерного зрения и сервис для распознавания и синтеза речи доступны по API

Управление инфраструктурой как кодом

Свобода автоматизации
с RestAPI, Terraform

Скоро

Платформенные сервисы для работы с данными в VK Cloud

Cloud ML Platform

ML-разработка полного цикла
без необходимости настраивать инструменты
и администрировать инфраструктуру.

Быстрый запуск экспериментов
с данными и обучение ML-моделей,
легкое масштабирование ресурсов.

Подключение
источников, обработка
данных, подготовка фич

Эксперименты с данными
и ML-моделями. Трекинг
и версионирование экспериментов,
данных и артефактов.

Деплой ML-моделей, интеграция
в продукт, мониторинг
и обновление моделей
в production

Компоненты Cloud ML Platform

JupyterHub

Среда для проведения экспериментов с данными.

Содержит набор популярных библиотек
и преднастроенное окружение для GPU.

Поддерживает одновременную работу множества пользователей.

MLflow

Сервис для MLOps задач.

Позволяет организовать централизованный трекинг и хранение моделей, параметров экспериментов, артефактов работы дата-специалистов.

MLflow Deploy

Позволяет решить задачи деплоя ML моделей.
Предоставляет возможность сделать доступной модель по REST API в несколько строк кода.

Также позволяет управлять выделенными ресурсами для деплоя с помощью Python client.

Компоненты Cloud ML Platform

Spark в Kubernetes

Managed Spark
с использованием
возможностей K8s
для масштабирования
на лету.

Сервис для тех, кому нужен
Spark без необходимости
устанавливать
и поддерживать Hadoop.

JupyterHub + MLflow в Kubernetes

Готовое рабочее пространство
с библиотеками
для обучения ML-моделей
и экспериментов
с большими данными.

Скоро

Дополнительные компоненты

Data Version:
версионирование данныхи
пайплайнов.

Feature Store: каталог фич
для тренировки ML-моделей.

Data Catalog: организация
работы с Data Lake, DWH.

Скоро

Проведение хакатонов

Участники трех федеральных хакатонов использовали ML Platform.

Быстрое предоставление ресурсов для обучения ML-моделей десяткам команд.

Отзыв одного из участников:

«В целом сложилось очень приятное впечатление!

Мы ограничились запуском jupyter notebook, и его пробросом, чтобы его локально можно было бы открыть и работать в нем. Парень кто настраивал сказал, что это самый приятный опыт использования и все очень понятно (читаешь название кнопки, кликаешь, и она делает именно то что надо)»

Кейс. Предсказание оттока клиентов

Нам требовалось разработать модель оттока на основе машинного обучения для автоматического предсказания склонности клиента к оттоку в течение прогнозируемого периода.

В качестве периода прогнозирования было решено выбрать 60 дней — если в течение этого времени человек не возвращается в магазин за новой покупкой, мы считаем, что он ушел в «отток».

Отзыв команды:

«В качестве среды для упаковки ML-моделей и их автоматического развертывания в облаке мы выбрали MLflow Deploy.
Под саму модель подняли конфигурацию на 2 CPU 2 GB.

По результатам наших тестов выделенных мощностей достаточно для обработки 10 000 RPM (запросов в минуту) при 50 переменных.»

Кейс – Банк из топ-10 по России

Активно использует используют JupyterHub + GPU для задач команды Data Science.

Также команда пользуется S3 для хранения датасетов.

Решает задачи:

- проведения экспериментов с данными;
- обучение ML моделей для видеоаналитики.

LLM и PEFT на практике

С применением PEFT
и подхода LoRa нам удалось
дообучить модель Pythia-12B
на собственном датасете.

Ресуры: одна GPU A100 с 40 GB.
Во время тренировки потребляет от 19
до 38 GB в зависимости от параметров.

Мы тренировали нашу модель около 28 часов,
после чего она стала понимать вопросы,
заданные на русском языке и отвечать на них.

Пример работы

ВОПРОС:

Что такое etcd в Kubernetes?

ОТВЕТ:

Etcd - это распределенное хранилище данных, которое используется в Kubernetes для хранения конфигурационных данных и состояния кластера. Он является частью инфраструктуры Kubernetes, которая обеспечивает надежность, отказоустойчивость и масштабируемость кластера.

ВОПРОС:

Как получить список подов в Kubernetes?

ОТВЕТ:

Для получения списка подов в Kubernetes нужно использовать команду kubectl get pods. Например, kubectl get pods -o jsonpath='{range.items[*]}{.metadata.name}{“\n”}{end}’.
В результате получим список подов в формате JSON.

Пример работы

ВОПРОС:

Когда родился Пушкин?

Ответь на этот вопрос, используя данный текст: Алекса́ндр Серге́евич Пу́шкин — русский поэт, драматург и прозаик, заложивший основы русского реалистического направления, литературный критик и теоретик литературы, историк, публицист, журналист.
Один из самых авторитетных литературных деятелей первой трети XIX века.

Википедия: Дата и место рождения: 6 июня 1799 г., Москва, Россия

ОТВЕТ:

Пушкин родился 6 июня 1799 года в Москве, Россия.

Пример работы

ВОПРОС:

Объясни мне что такое сервис?

Ответь на этот вопрос, используя данный текст:
A Kubernetes Service is an abstraction which defines a logical set of Pods running somewhere in your cluster, that all provide the same functionality. When created, each Service is assigned a unique IP address (also called clusterIP). This address is tied to the lifespan of the Service, and will not change while the Service is alive. Pods can be configured to talk to the Service, and know that communication to the Service will be automatically load-balanced out to some pod that is a member of the Service.

ОТВЕТ:

Kubernetes Service - это абстракция, которая определяет логическое сочетание подов, работающих в вашем кластере, которые обеспечивают одинаковую функциональность. Когда создан, каждый Service имеет уникальный IP-адрес (также называемый clusterIP), который связан с жизненным циклом Service и не меняется, пока Service будет жив. Pods могут быть настроены, чтобы общаться с Service, и знать, что общение с Service будет автоматически масштабировано на Pod, который является членом Service.

VK Cloud

Спасибо за внимание!

Александр Волынский
Технический менеджер продукта Cloud ML Platform VK Cloud

tg: @volinski

бизнес