

Платформа облачных сервисов от Mail.ru Group



ФИНАНСОВЫЙ
УНИВЕРСИТЕТ

ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Платформа облачных сервисов от Mail.ru Group

Безопасная и масштабируемая IT-инфраструктура для разработки и использования облачных приложений

Сервисы:

Большие данные



Масштабируемый
сервис для анализа
больших данных
на базе Apache
Hadoop и Apache Spark

Базы данных в облаке



Управляемые
и масштабируемые
базы данных
PostgreSQL, MySQL,
MongoDB и ClickHouse

Машинное обучение



Сервис для быстрой
разработки
приложений на основе
машинного обучения



Большие данные в облаках

Масштабируемый PaaS-сервис для анализа больших данных (big data) на базе Apache Hadoop, Apache Spark, ClickHouse

[Попробовать бесплатно](#)[Рассчитать стоимость](#)

Замените дорогую и неэффективную локальную инфраструктуру обработки больших данных, которая нужна всего несколько часов в неделю, на мощную облачную инфраструктуру на основе Apache Hadoop, Apache Spark, ClickHouse.

Большие данные от Mail.ru Cloud Solutions: преимущества

Простота использования

Низкая стоимость

Интеграция

Надежность



Создание кластера за несколько минут. Сервис берет на себя создание машин и кластеров, настройку Hadoop, оптимизацию производительности



Можно добавлять модули, создавать нетиповые конфигурации кластера благодаря root-доступу к каждому узлу и к административному интерфейсу



Управление кластером через веб-интерфейс, командную строку, API



Гибкая масштабируемость вычислительных ресурсов и объема хранения — от пары до сотен серверов и обратно за 10 минут



Поддержка на русском языке

Простота использования

Низкая стоимость

Интеграция

Надежность



Оплата посекундно, за фактически потребленный дисковый объем



Не нужно платить за RAM и CPU остановленных виртуальных машин



Кластер Hadoop из 10 узлов — 39 ₽/час

Простота использования

Низкая стоимость

Интеграция

Надежность



С сервисами Mail.Ru Cloud Solutions: S3-совместимым объектным хранилищем, виртуальными машинами, кластерами Kubernetes, вычислениями на базе GPU



Приватная сеть с компонентами Hadoop и другими сервисами в вашей IT-системе для создания гибридного облака



Лёгкая миграция приложений для Hadoop с аналогичных облачных решений, таких как Amazon AWS, Microsoft Azure, Google Cloud

Простота использования

Низкая стоимость

Интеграция

Надежность



Сервис создан на основе глубокой экспертизы команды Mail.Ru в использовании Hadoop, Spark и ClickHouse



Хранение данных строго на территории РФ, в ЦОДах уровня TIER III, с тройной репликацией и непрерывным мониторингом безопасности

Области применения

Hadoop

Анализ агрегированных данных

Машинное обучение

Исследования

Финансы и банки

Spark

Агрегация данных, анализ в реальном времени

Анализ операционной деятельности компании

Транспорт

Диджитал-маркетинг

Денежные переводы

Фильтрация спама

Индустрия и IoT

Машинное обучение

Нейронные сети, глубокое обучение, распознавание образов, ИИ

Исследования

Поиск ископаемых, медицина, геномика, социология, астрономия, численные методы, классификация

Финансы и банки

Финансовый анализ, кредитный скоринг

Области применения

Hadoop

Анализ агрегированных данных

Машинное обучение

Исследования

Финансы и банки

Spark

Агрегация данных, анализ в реальном времени

Анализ операционной деятельности компании

Транспорт

Диджитал-маркетинг

Денежные переводы

Фильтрация спама

Индустрия и IoT

Анализ операционной деятельности компании

Агрегация data lake, ETL-обработка, оценка решений, анализ возможностей на рынке

Транспорт

Мониторинг, отчеты, перераспределение транспортной нагрузки

Диджитал-маркетинг

Сегментация пользователей по действиям на сайте, в приложении

Денежные переводы

Fraud detection в операциях с картами

Фильтрация спама

Выявление спама в письмах и комментариях, автоматическое удаление

Индустрия и IoT

Мониторинг и анализ состояния бизнес-процесса

Машинное обучение в облаке

Сервис для быстрой разработки приложений на основе машинного обучения



Машинное обучение от Mail.Ru Cloud Solutions помогает разработчикам и исследователям быстро создавать приложения на основе глубокого обучения без приобретения, конфигурирования и поддержки собственной инфраструктуры.

Мы поможем вам выбрать и внедрить нужные технологии машинного обучения, которые лучше всего подойдут для решения вашей задачи.



Рабочее место аналитика (data scientist)

Быстрое развёртывание сред обучения моделей позволяет разработчикам использовать популярные инструменты и фреймворки без трудоёмких настроек.

Среды развёртываются в виде образов с установленными и настроенными сервисами обучения нейронных сетей и драйверами.

Помимо самой среды, в каждый образ входят NVIDIA CUDA Deep Neural Network Library (cuDNN), пакетный менеджер Anaconda, системы контроля версий (git, svn, mercurial).



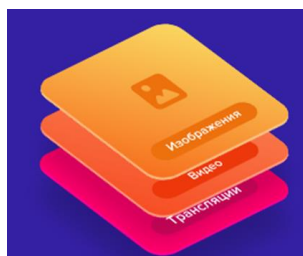
Готовые API для быстрого запуска сервисов

Компьютерное зрение Vision

Распознавание лиц,
объектов, номеров
автомобилей, улучшение
изображений

Аудиоаналитика Sounds

Перевод речи в текст,
синтез речи, голосовой
интерфейс приложений,
анализ тональности речи



> 25.000

объектов в базе определения
объектов

100 мс

среднее время на анализ
изображения

98%

Точность распознавания лиц

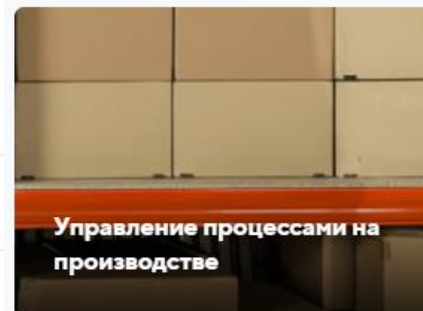
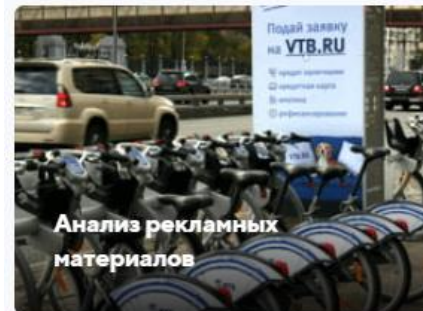
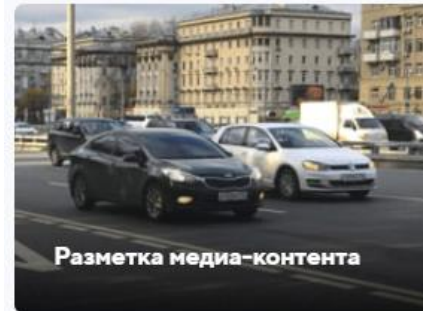


Примеры использования

Vision позволит автоматизировать любой человеческий труд, который основан на визуальной проверке, в основе которой лежит сверка оригинала (или заранее заданных условий) с тем, что происходит на фотографии, видео или в производственном процессе.

В Vision реализованы некоторые возможности, доступные через программные интерфейсы API. Возможности технологии не ограничиваются данным списком, мы создаем любые решения на базе компьютерного зрения.

- Распознавание лиц
- Определение сцен и объектов
- Определение качества изображения
- Отслеживание людей
- Распознавание знаменитостей
- Распознавание текста на изображениях
- Обнаружение автомобилей
- Увеличение разрешения
- Поиск достопримечательностей
- Выявление дефекта на производстве
- Определение номеров автомобилей
- Модерация изображений

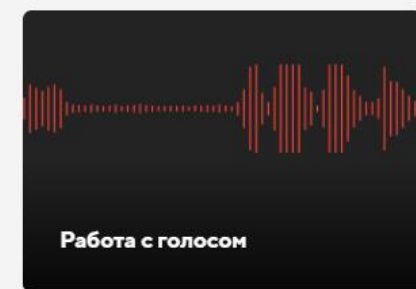
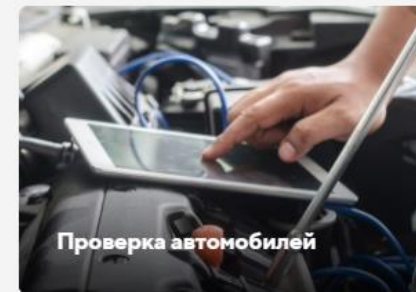
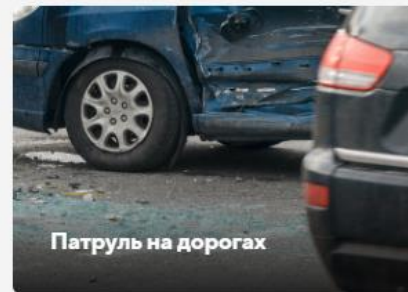
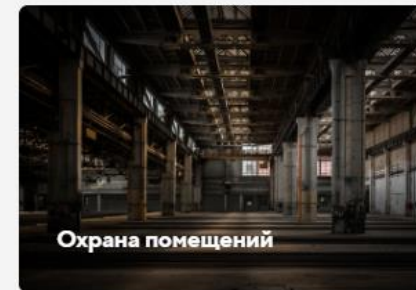
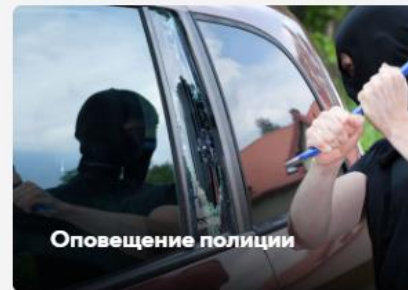


Сервис аудиоаналитики Sounds

Технология распознавания звуков и речи на базе машинного обучения от Mail.Ru Group

Sounds — облачная технология обработки и синтеза звуков от Mail.Ru Cloud Solutions.

- Обнаруживает и анализирует любые звуки или их сочетание в аудиопотоке.
- Распознает отдельные звуки и их комбинации в аудиопотоке
- Предугадывает ситуацию на основе сочетания звуков
- Различает громкость, тональность, интенсивность звучания
- Уведомляет о том, что обнаружен определенный звук
- Узнает и разделяет звуки даже в большом шуме
- Работает с речью: распознает голоса и преобразует в текст



Аналитика больших данных Яндекс

Аналитика больших данных Яндекс

PREVIEW

Yandex Data Proc

Сервис помогает разворачивать кластеры Apache Hadoop® и Apache Spark™ в инфраструктуре Яндекс.Облака. Вы сами определяете размер кластера, мощность узлов и набор сервисов Apache® (Spark, HDFS, YARN, Hive, HBase, Oozie, Sqoop, Flume, Tez, Zeppelin).

Сервис Yandex Data Proc находится на стадии Preview и не тарифицируется. Инфраструктура Яндекс.Облака защищена в соответствии с ФЗ-152.



Быстрые операции с кластерами

Создание кластера займёт всего несколько минут. Вам не нужно думать о создании хостов, настройке, установке пакетов и привязке хостов в единый кластер — все это сервис сделает автоматически. В дальнейшем вы сможете легко изменить количество или вычислительные мощности хостов кластера.

Гибкая настройка каждого кластера

У вас будет полный контроль над кластером за счёт доступа ко всем хостам с правами пользователя «root». Вы можете устанавливать только необходимые сервисы Hadoop, загружать свои приложения и, при необходимости, настраивать кластер в любой момент.



Аналитика больших данных Яндекс

PREVIEW

Yandex Data Proc



Быстрые операции с кластерами

Создание кластера займёт всего несколько минут. Не нужно думать о создании хостов, настройке, установке пакетов и провязке хостов в единый кластер — все это сервис сделает автоматически. В дальнейшем можно легко изменить количество или вычислительные мощности хостов кластера.

Гибкая настройка каждого кластера

Полный контроль над кластером за счёт доступа ко всем хостам с правами пользователя «root». Можно устанавливать только необходимые сервисы Hadoop, загружать свои приложения и, при необходимости, настраивать кластер в любой момент.

Эластичность

Добавляйте динамические новые хосты в кластер для увеличения его мощности и платите только за время их использования. Храните данные в объектном хранилище и удаляйте неиспользуемые хосты, чтобы сэкономить на вычислительных ресурсах.

Выбор типа хранилища

Для каждого кластера можно выбрать свой способ хранения данных. Предоставляются два варианта: стандартное сетевое и быстрое сетевое хранилище. Первый из них более дешёвый, а второй обеспечивает более высокую скорость.

Изоляция и шифрование

В Yandex Data Proc данные разных клиентов Яндекс.Облака полностью изолированы друг от друга. Благодаря отсутствию общих компонентов никто не сможет получить доступ к загруженным данным. Возможно настроить шифрование для кластера, если это необходимо.



Yandex Data Science Virtual Machine

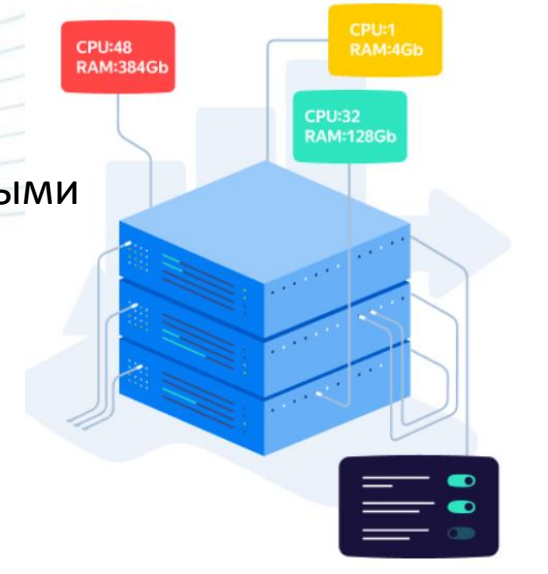
Yandex Data Science Virtual Machine (DSVM) — это виртуальная машина с предустановленными популярными библиотеками для анализа данных и машинного обучения. DSVM можно использовать как среду для обучения моделей и экспериментов с данными.

Предустановленное программное обеспечение

Операционная система: Ubuntu 18.04

Установленные пакеты:

- Менеджер пакетов conda с Python 2.7 и Python 3.6.
- Инструменты для интерактивных и воспроизводимых вычислений Jupyter Notebook и JupyterLab.
- Библиотеки Machine Learning:
 - CatBoost;
 - LightGBM;
 - XGBoost;
 - TensorFlow;
 - PyTorch.
- Система управления контейнерами Docker.
- Консольные клиенты систем контроля версий: SVN, Git, Mercurial.
- Библиотеки NumPy, scikit-learn, SciPy, оптимизированные с помощью Intel Math Kernel Library и Data Analytics Acceleration Library.
- Оптимизированные библиотеки для работы с изображениями: libjpeg-turbo, Pillow-SIMD.



Yandex SpeechKit

Сервис распознавания и синтеза речи. Инфраструктура сервиса спроектирована с учётом высоких нагрузок, чтобы обеспечивать доступность и безотказную работу системы даже при большом количестве одновременных обращений. Именно на основе SpeechKit работает **Алиса** — голосовой помощник Яндекса.

Поддержка трёх языков

Сервис работает с аудио и текстами на трёх языках: русском, английском и турецком. Поддержку каждого из этих языков легко добавить в любой момент — при выходе на новую аудиторию не потребуется подключать сервисы сторонних разработчиков.

Естественное звучание

Если формировать речь из слов, произнесённых актёром, звучание получается неестественным. Yandex SpeechKit складывает речь более чем из миллиона отдельных фонем, а интонация задаётся нейросетью, обученной на множестве примеров из реальной жизни. Поэтому когда система воспроизводит текст, результат не режет слух.

Синтез в реальном времени

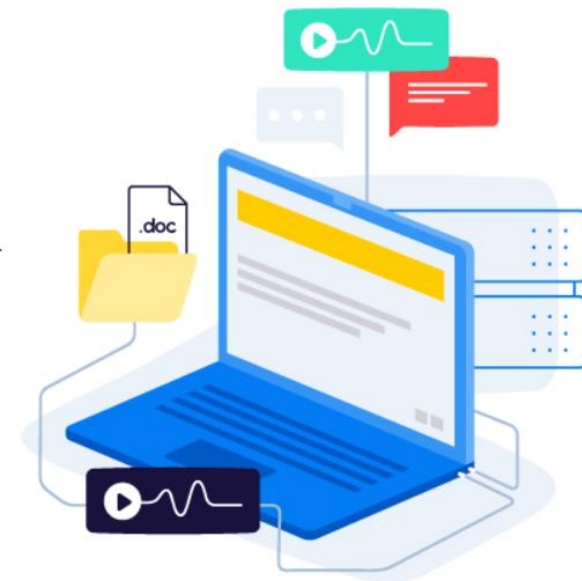
Отправив текст для синтеза, ваш сервис или приложение сразу же получит в ответ аудиозапись: задержка настолько мала, что позволяет создавать программное обеспечение с потоковым воспроизведением.

Прозрачное ценообразование

Когда вы присылаете аудиозапись, стоимость формируется по её длительности, а когда на вход подаётся текст — по количеству символов. Это позволяет строить точные прогнозы расходов.

Удобный API

Для обмена данными с сервисом предусмотрен HTTP API. Поэтому вам не придется разворачивать и поддерживать свою инфраструктуру, чтобы в сжатые сроки реализовывать нужные функции.



Yandex Translate

Сервис для интеграции алгоритмов Яндекс.Переводчика в приложения или веб-проекты для конечных пользователей. Yandex Translate поддерживает более 90 языков и умеет переводить отдельные слова и целые тексты.

Более 90 языков с автоопределением

Yandex Translate даёт возможность переводить между множеством языков — их уже больше 90. Запрос с исходным текстом от вашего сервиса может и не содержать код, указывающий, на каком языке текст написан: API определит это автоматически.

Гибридная система

Существует два актуальных подхода к автопереводу. Статистический подход отличается от нейросетевого тем, что текст разбивается на короткие фрагменты. За счёт этого отдельные слова переводятся точнее. Но чтобы анализировать предложения целиком и понимать фразы в контексте, необходим машинный интеллект на основе нейронных сетей. В API задействованы обе модели: алгоритм каждый раз выбирает, какую из них применить.

Алгоритм продолжает обучаться

Машинное обучение не просто помогло запустить Yandex Translate, но и способствует улучшению переводов. В обработку постоянно попадают новые тексты, относящиеся к множеству языковых пар. При этом ваши данные, переданные в сервис, не используются для таких целей.

Какие задачи можно решать с Yandex Translate:

- Упрощение коммуникации
- Расширение аудитории сайта
- Построение платформы создания контента
- Предварительный перевод



Yandex Vision

Анализ изображений с помощью моделей машинного обучения. Yandex Vision подходит, например, для создания электронного архива документов с возможностью поиска или для автоматизации бизнес-процессов. Для работы с сервисом можно использовать простые REST API или gRPC API.

Распознавание текста на изображениях

Yandex Vision распознает текст на отсканированных страницах документов и возвращает найденные блоки текста, строки и слова с указанием их местоположения на изображении. Вы также можете использовать сервис для анализа фотографий, например находить дорожные знаки или номера машин.

Более 40 языков

Yandex Vision даёт возможность находить и распознавать текст на изображениях для более 40 языков — среди них русский, английский, немецкий, японский и многие другие.

PREVIEW. Модерация изображений

Технология классификации изображений поможет проанализировать контент, который загружают ваши пользователи. Применяйте признаки модерации, чтобы определить нарушение авторских прав по наличию водяных знаков или отфильтровать другой нежелательный контент.

