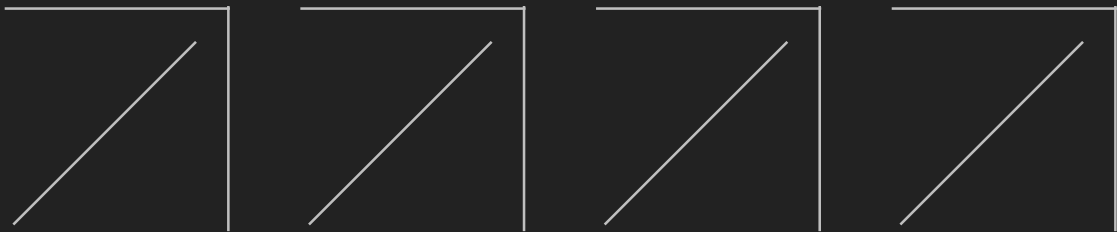


# Облачные сервисы для цифровой трансформации, ИИ и не только.



Используйте ИИ и масштабируйте возможности бизнеса в нашем облаке



**№1**  
**в сегменте IaaS**

на российском рынке

[iKS-Consulting, 2024](#)

**№1**  
**в сегменте PaaS**

на российском рынке

[iKS-Consulting, 2024](#)

**№1**  
**в AI-решениях**

на российском рынке

[CNews Analytics, 2024](#)

**В ТОП-10**

крупнейших IT-компаний в России

[CNews500, 2025](#)

**В ТОП-5**

самых дорогих компаний Рунета

[Forbes, 2025](#)

**В ТОП-30**

самых эффективных  
IT-компаний в России

[CNews Analytics, 2024](#)

**В ТОП-200**

мировых суперкомпьютеров  
Christofari и Christofari Neo

[TOP500, 2024](#)

OpenAI API Pricing · TokenCost.app · Oplexa AI Budget Report

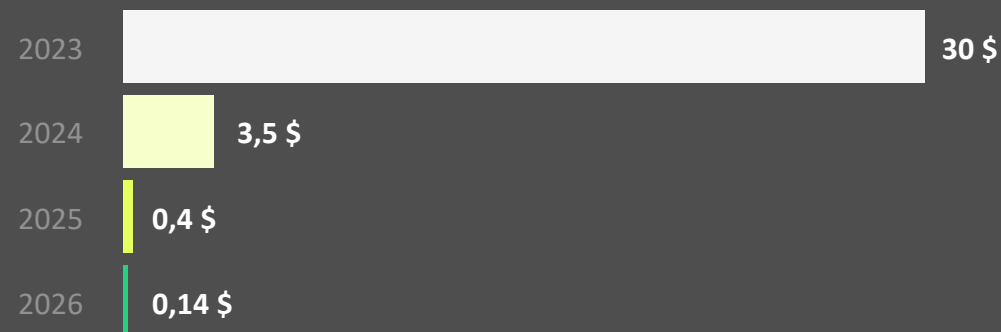
## ПАРАДОКС ДЖЕВОНСА

**ТОКЕНЫ ПОДЕШЕВЕЛИ  
В 200+ РАЗ. РАСХОДЫ  
ВЫРОСЛИ ВТРОЕ**

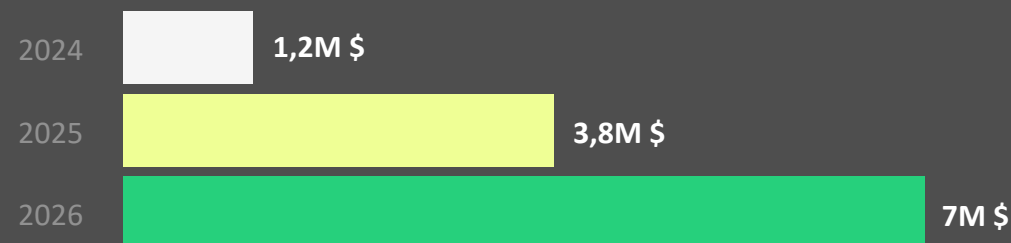
**БАРЬЕР ВХОДА В ИИ  
СНИЗИЛСЯ, НО ЧТОБЫ ПОЛУЧИТЬ ROI,  
НУЖНА ЦЕЛЬ И ОБЛАКО**



## СТОИМОСТЬ 1М ТОКЕНОВ (GPT-4 УРОВЕНЬ)

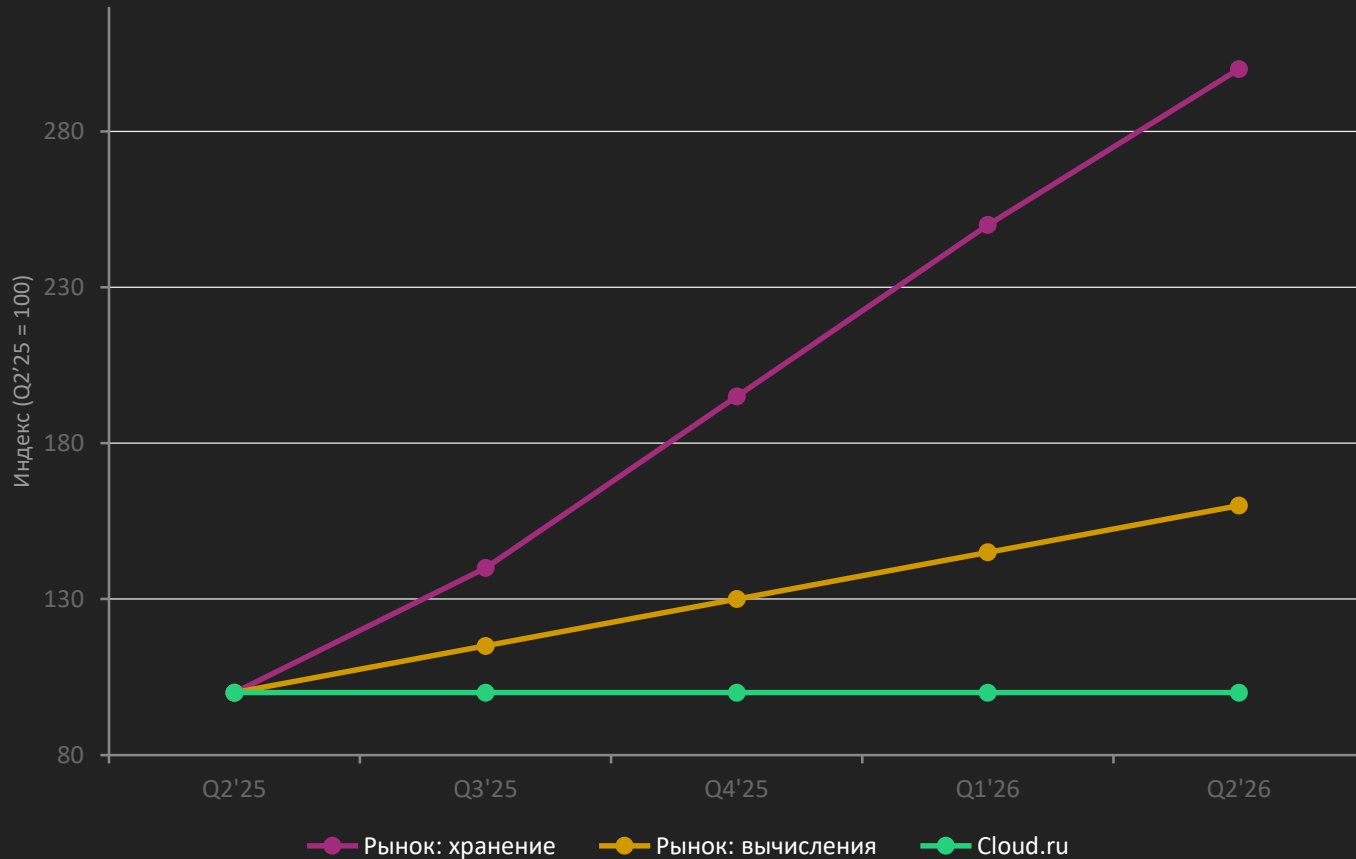


## СРЕДНИЙ БЮДЖЕТ КОМПАНИИ НА ИИ



# РАЗВИТИЕ ИИ НЕВОЗМОЖНО БЕЗ ОБЛАКОВ

Индекс закупочных цен: Q2'25 = 100; до Q1'26 — факт, далее — прогноз



Рынок: хранение, за год

**×3**

**+200%**

Рынок: вычисления, за год

**×1,6**

**+60%**

Цены Cloud.ru, за год

**0%**

**без роста**

## Что это значит для AI-пилотов

- Бюджеты требуют сценарного планирования
- CAPEX существенно растёт
- Облако повышает предсказуемость, даёт гибкость, за счёт OPEX

Источники: [TrendForce 2026](#), [Аналитика Cloud.ru](#)

Тестирование агентов — дорогой, но необходимый процесс

**100% ГИПОТЕЗ (POC)**

Поиск бизнес-кейсов

**40% ТЕСТИРОВАНИЕ**

Проверка на данных

**12%  
ПРИМЕНЕНИЕ И  
ROI**

**88% GenAI проектов** не доходят до реального использования. Покупка собственного оборудования для них ведет к прямым убыткам



## Взрывной рост внедрения

К концу 2026 года **40%** корпоративных приложений будут включать AI-агентов (рост с 5% в 2025 году)



## Цена ошибки

- Мощные GPU, необходимые для тестирования AI-агентов, составляют **30–80%** стоимости инфраструктуры.
- Проверка гипотезы занимает 1–4 месяца; закупка оборудования под разовый PoC может привести к потерям до 80% вложений.
- Срок поставки и настройки собственной инфраструктуры — от 3 недель до двух кварталов



## Оптимальная стратегия

1. Используйте **облачные ресурсы** для R&D
2. Масштабируйте только те кейсы, которые приносят прибыль

# Cloud.ru Evolution

Публичное облако  
собственной разработки  
для быстрого старта



## Экономичность

Прозрачное ценообразование  
и контроль затрат в личном  
кабинете в режиме онлайн



## Импортонезависимость

Собираем платформу силами  
своей команды на свободно  
распространяемых компонентах



## Понятный интерфейс

Запускать и управлять  
конфигурацией сервисов можно  
в консоли без запроса  
в поддержку



## Безопасность

Размещаем инфраструктуру  
в ЦОД Tier III, соблюдаем  
требования 152-ФЗ, прошли  
аттестацию УЗ-1

# ДЕЛАЕМ ДОСТУП К ОБЛАКАМ ПРОСТЫМ И РАБОТУ С НИИ УДОБНОЙ

## Персональный AI-помощник в облаке

Гига помощник — новый член вашей команды на базе GenAI, который упростит работу в облаке

[подробнее](#)

## 100+ сервисов

Для любых задач в облаке: от размещения сайта и тестирования до запуска и адаптации ML-моделей

[подробнее](#)

## Импортонезависимые решения

Облачная платформа Cloud.ru Evolution на базе собственных разработок команды Cloud.ru

[подробнее](#)



## Цифровая среда для работы с GenAI

Облачная Evolution AI Factory, которая объединяет готовые инструменты для быстрого внедрения LLM и создания мультиагентных систем

[подробнее](#)



## Защищенные ЦОД

Девять дата-центров в России с уровнем надежности Tier III, современные серверы и последние версии ПО

[подробнее](#)

## Безопасность данных

Сертификаты для работы с персональными данными, ГИС, КИИ и финансовыми операциями

[подробнее](#)

# AI FACTORY

Foundation  
Models



ML  
Inference



Jupyter  
Notebooks



ML  
Finetuning



Managed  
RAG



AI Agents



## Foundation Models



Используйте готовые топовые модели



QWEN



Llama



DeepSeek



Оплата за токены по мере потребления



ML Inference



Jupyter Notebooks



ML Finetuning






















Managed RAG



















AI Agents









|  |   |  |  |  |  |
|--|---|--|--|--|--|
| <p>Foundation Models</p>  | <h2>ML Inference</h2>  <p>Разворачивайте свои модели в готовом окружении</p>  <p>Платите только за потраченные GPU-минуты</p>  <p>General Availability</p>  | <p>Jupyter Notebooks</p>  | <p>ML Finetuning</p>  | <p>Managed RAG</p>  | <p>AI Agents</p>  |
|--|---|--|--|--|--|

|  |   |   |  |  |  |
|--|---|---|--|--|--|
| <p>Foundation Models</p>  | <p>ML Inference</p>  | <h2>Jupyter Notebooks</h2>  <ul style="list-style-type: none"><li> Работать на GPU в привычной среде Jupyter Notebooks</li><li> Платите только за потраченные GPU-минуты</li><li> General Availability</li></ul>  | <p>ML Finetuning</p>  | <p>Managed RAG</p>  | <p>AI Agents</p>  |
|--|---|---|--|--|--|

|  |   |  |   |  |  |
|--|---|--|---|--|--|
| <p>Foundation Models</p>  | <p>ML Inference</p>  | <p>Jupyter Notebooks</p>  | <h2>ML Finetuning</h2>  <ul style="list-style-type: none"><li> Дообучайте большие модели под свои задачи</li><li> Платите только за потраченные GPU-минуты</li><li> General Availability</li></ul>  | <p>Managed RAG</p>  | <p>AI Agents</p>  |
|--|---|--|---|--|--|

|   |   |   |   |  |   |
|---|---|---|---|--|---|
| Foundation Models   | ML Inference  | Jupyter Notebooks   | ML Finetuning   | <b>Managed RAG</b><br><br>Интегрируйте свою базу знаний в LLM<br><br>Платите только за токены используемых моделей<br><br>General Availability | AI Agents   |
|  |  |  |  |   |  |

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| Foundation Models   | ML Inference  | Jupyter Notebooks   | ML Finetuning   | Managed RAG   | <h2>AI Agents</h2> <ul style="list-style-type: none"><li>✦ Создавайте своих ИИ-агентов</li><li>📁 Платите только за токены используемых моделей</li><li>📈 General Availability</li></ul> |
|  |  |  |  |  |    |



Гига

ИИ-помощник



Гига

ИИ-помощник

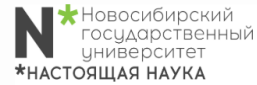




Разработчик «Жива Технологии» развернул приложение с AI-ассистентом в облаке Cloud.ru

От облака нам были нужны не только виртуальные машины с GPU, но и полный спектр сервисов для резервного копирования, мониторинга и логирования работы всей нашей инфраструктуры, чтобы промышленная среда была всегда доступна и приложение работало корректно.

Скорость работы LLM-модели для нас приоритет. Пока модель распознает нутриенты по фото за 8-10 секунд, мы стремимся сократить время до 2-3 секунд. И нам важно



НГУ разместил умного чат-бота в Telegram с помощью облака Cloud.ru

Студенты IT-лаборатории НГУ используют ресурсы облака Cloud.ru Evolution для проекта – размещения чат-бота на базе большой языковой модели, который выдает ответы на вопросы абитуриентов и студентов быстрее, чем за 10 секунд.

Нам было важно, чтобы работа в консоли была простой и удобной – мы занимаемся проектом чисто научной командой и не могли выделить отдельного технического специалиста, который взял бы на себя настройку сложной инфраструктуры.



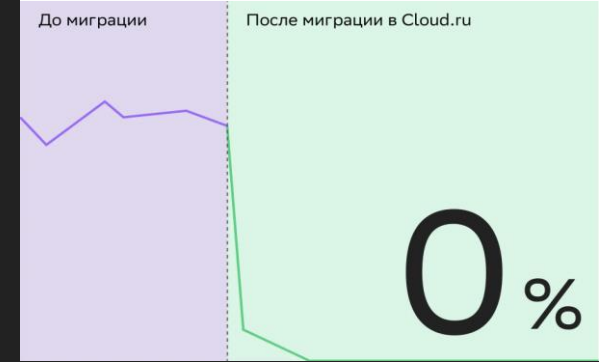
«Литрес» ускорила аналитику в 30 раз с Cloud.ru

Литрес построила облачное аналитическое хранилище на базе КХД Cloud.ru и сократила время обработки и анализа данных.

Как компания-лидер на рынке лицензионных цифровых книг в России и странах СНГ создала собственное аналитическое хранилище в Cloud.ru и теперь обрабатывает запросы при обращении к данным за 5 секунд



UVENCO: телеметрия без простоев в облаке Cloud.ru



Железо не дешевеет, а наш парк автоматов разрастается, добавляются новые разработки.

Мы решили, что перенесем в облако одну из самых критичных для нас инфраструктур – онлайн-телеметрию. Так мы избавимся от закупки серверов на миллионы

Понятно и  
относительно  
легко



С осторожностью

- Ввод и обработка данных (сканы документов, CRM)
- Поддержка клиентов по рутинным вопросам (ссылки на документацию, передача запросов людям)
- Административные задачи в офисе (напоминания оформление новых сотрудников)
- Обработка и анализ документов (извлечение ключевой информации, проверка корректности заполнения)
- Анализ покупателей (адресные предложения)
- Суммаризация созвонов, встреч, переписки
- Творческие задачи (маркетинговые кампании, дизайн, написание статей)
- Высокорискованные финансовые операции и решения
- Задачи связанные с безопасностью и жизнью людей
- Процессы без четкого регламента, неструктурированные данные
- Юридические и этические вопросы

## AI-агент или Co-Pilot? 5 уровней автономности AI

- 1. Режим оператора.** Люди сохраняют полный контроль, AI лишь помогает. Агент генерирует предложения и черновики ответов, человек принимает решения и выполняет действия
- 2. Режим сотрудничества.** AI дает рекомендации, а человек одобряет или отклоняет
- 3. Режим консультанта.** AI автономно выполняет предварительно утвержденные действия низкого риска. Если попадается решение высокого риска, агент запрашивает помощь человека
- 4. Режим утверждающего.** AI-агенты сами справляются с большинством операций и оценивают их уровни риска. Если риск превышает определенное пороговое значение, агент запрашивает одобрение у человека.
- 5. Режим наблюдателя.** Агенты работают автономно в пределах определенных границ. Люди только мониторят производительность, при этом у них остается возможность вмешаться.



## Цифровизация бизнеса

- Ускорение time to market
- Доступ к новым технологиям
- Быстрое масштабирование
- Гибкость инфраструктуры
- Лучший контроль и управление ресурсами



## Безопасность и регулирование

- Облачный провайдер обеспечивает лучшую защиту от кибератак
- Соответствие сервисов требованиям ФЗ и регуляторов
- Своевременные обновления облачных сервисов
- Высокий показатель SLA и система компенсаций



## Модернизация IT-ландшафта

- Рост производительности и надежности IT-систем
- Возможность замены устаревшего оборудования и лицензий ПО
- Переход от IaaS- к PaaS- и SaaS-моделям
- Ускоренное развертывание стандартных IT-решений
- Возможность остановить проект, потерявший свою актуальность



## Экспертиза сотрудников

- Поддержка инфраструктуры на стороне провайдера
- Снижение нагрузки на IT-департамент клиента



## Затраты на IT-инфраструктуру

- Снижение CAPEX и затрат на ПО, отсутствие амортизации, переход на OPEX
- Минимизация простоев инфраструктуры
- Сокращение затрат на ФОТ IT-специалистов

## IaaS

Инфраструктура  
любого уровня сложности

## PaaS

Десятки готовых сервисов,  
партнерские решения в облаке

## ML&AI

Платформа машинного обучения  
на базе мощнейших  
суперкомпьютеров России

## Сценарии использования облачных сервисов

Миграция традиционных приложений в облако

Аналитика данных в облаке

Разработка и тестирование

Платформа для мобильных и веб-приложений

Разработка и эксплуатация ML-моделей

Резервное копирование

Аварийное восстановление

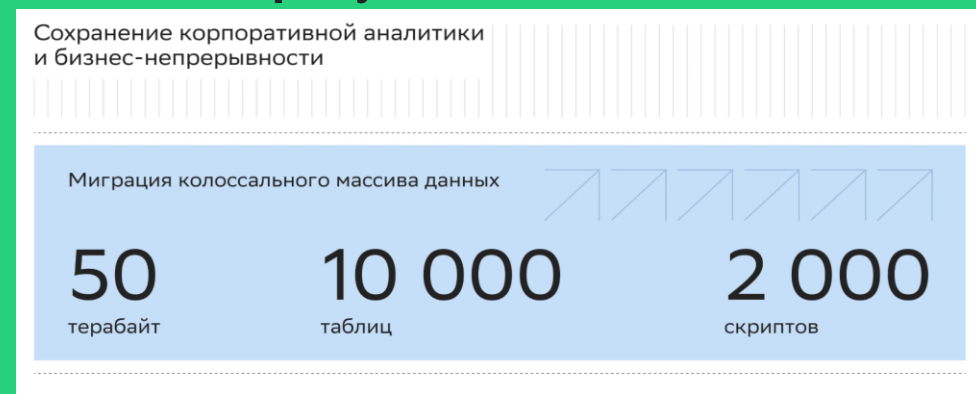
Удаленный рабочий стол

Рекомендательные системы

# 1 000 сервисов и 50 ТБ данных: как METRO перенесла весь IT-ландшафт и BI-систему в облако Cloud.ru.

Ритейлер не только выполнил плановую локализацию IT-ландшафта, но и в экстренном порядке за 3 месяца перенес в облако корпоративную BI-систему без даунтайма.

## Ключевой результат



## Задача

Чтобы повысить эффективность решения своих технологических задач, в 2019 году METRO начала использовать облачную инфраструктуру по модели IaaS: уже на тот момент руководство компании понимало, что за облаками — будущее.

С 2019 по 2023 METRO все больше разворачивалась в сторону облачных решений — активно велась разработка cloud native продуктов. А в конце 2023 года компания, находясь под санкционным давлением, приняла решение найти отечественного провайдера со зрелым SaaS-облаком, на мощности которого в будущем можно будет перенести весь централизованный IT-ландшафт компании.

Нам было нужно перенести весь IT-ландшафт компании для управления B2B-каналом: B2B-витрину, а также системы сборки, управления мастер-данными, ценообразования и не только. Всего порядка 20 продуктов и 200 микросервисов. **Антон Антоничев** IT-директор METRO

# КУПЕР ускорил подготовку аналитической отчетности благодаря переходу в облако Cloud.ru

## Задача

Перенести из облака в облако 40 ТБ сырых бизнес-критичных данных вместе с дев-контуром, с которым работают аналитики и разработчики Купера. Перенос данных должен произойти без остановки аналитических процессов.

## Решение

Специалисты облачного провайдера развернули СУБД в сервисе [Evolution Managed ArenadataDB](#) на платформе [ArenadataDB](#). В этой версии клиент может конфигурировать и деликатно настраивать кластер, мониторить текущие запросы с анализом скорости их исполнения, работать с удобными инструментами бэкапирования и восстановления, а также пользоваться базой знаний по сервису и получать круглосуточную вендорскую поддержку.

## Результат

Купер без остановки процессов перенес в облако Cloud.ru 40 ТБ данных и тестовый контур, с которым работают аналитики и разработчики. Инфраструктура, развернутая с сервисом Evolution Managed ArenadataDB, работает в режиме T-1, когда данные за вчера доступны на следующий день – это позволяет обеспечивать управленческую отчетность и ad-hoc аналитику быстрее.

Миграция в облако Cloud.ru ускорила выполнение запросов, увеличила прозрачность работы с данными и позволила гибко масштабироваться под рост запросов.

## Почему Cloud.ru

Провайдер предложил аналитическую базу данных на базе решения с открытым исходным кодом Greenplum. Также Купер оценил техническую поддержку и скорость реагирования на запросы со стороны команды Cloud.ru во время нагрузочного тестирования.

Миграция нашей инфраструктуры с одного из ведущих российских провайдеров в Cloud.ru позволила нам

## ИНФОРМАЦИЯ К РАЗМЫШЛЕНИЮ

1. В компании есть или готовится план цифровой трансформации
2. Компания разрабатывает ИИ Агентов/Ботов или планирует использовать ИИ в своей деятельности
3. Нелинейный рост затрат на ИТ, CAPEX vs. OPEX
4. Явно выраженная сезонность бизнеса – неравномерная загрузка ИТ инфраструктуры
5. Проблемы с ИТ которые приводили к замедлению работы или временной приостановке бизнеса
6. Были инциденты с кибербезопасностью, готовится план резервного копирования и аварийного восстановления данных
7. Длительные сроки и сложные процессы закупки оборудования для новых проектов
8. Сложности с поиском квалифицированных ИТ специалистов
9. Задержки при разработке и запуске новых приложений

**Ответ «да» на один или нескольких вопросов  
– есть предпосылки для использования облачных сервисов!**

# СПАСИБО ЗА ВНИМАНИЕ!

**КОНТАКТЫ:**

**ГРИГОРИЙ ЖУКОВ**  
[GVZHUKOV@CLOUD.RU](mailto:GVZHUKOV@CLOUD.RU)



**Кейсы клиентов**