

16. Стоимость владения системой в первый, второй и пятый годы с учётом роста данных и базы пользователей

Ниже приведена **примерная** оценка совокупной стоимости владения (ТСО) системой на 1-й, 2-й и 5-й годы с учётом условного роста данных и пользовательской базы. Поскольку каждая компания имеет свои скидки, региональные расценки и объёмы трафика, точные цифры будут отличаться. Однако данная модель даст **примерное представление**, из чего складывается итоговая сумма и как она эволюционирует.

1. Структура стоимости

Для простоты разобьём затраты на **четыре** основные группы:

1. Инфраструктура (облачные ресурсы)

- Контейнерные кластеры (Kubernetes) в различных облаках.
- Виртуальные машины/контейнеры (под каждую микросервисную группу).
- СУБД (SQL/NoSQL/Time Series) — либо управляемые сервисы, либо собственные кластеры.
- Data Lake, хранилища объектов (S3, Azure Blob, GCS).
- Сервис очередей/шины (Kafka, RabbitMQ) и т.д.

2. Хранение и передача данных

- Объём хранимых данных (в т.ч. резервные копии).
- Трафик между микросервисами, внешними устройствами (IoT), выгрузки в Data Lake и обратно.
- Рост данных во времени (журналы тренировок, соц-посты, аналитика).

3. Подписки и лицензии

- Специализированные инструменты DevOps (если используются коммерческие версии).
- Лицензии на аналитические/ML-платформы (в случае Enterprise-решений).
- Системы мониторинга, безопасности (в случае платных решений).

4. Человеческие ресурсы (люди)

- Команда DevOps/Cloud-инженеров, SRE (Site Reliability Engineers).
- Команда Data Engineers/ML-специалистов, сопровождающих Data Lake и стриминговую платформу.
- Разработчики микросервисов (Backend, Mobile/Web), QA, поддержка.
- Участие в поддержке 24/7 (если нужно гарантировать высокий SLA).

На практике «человеческие ресурсы» обычно оказываются одним из главных факторов затрат, особенно если речь идёт о сложной системе с мультиоблачной инфраструктурой.

2. Допущения и ориентиры роста

Чтобы прикинуть затраты, предположим **примерные** темпы увеличения пользователей и объёмов данных:

- **Год 1:**
 - Активная аудитория: 100k пользователей (из них одновременно активных, скажем, 5–10%).
 - Объём данных: до 1–2 ТБ (учитывая что соц. и тренировочные данные пока в зачаточном состоянии).
 - Нагрузка: ежедневные пики — тысячи одновременных запросов, небольшие IoT-потoki.
- **Год 2:**
 - Рост в 3–5 раз (до 300k–500k пользователей).

- Объём данных: 5–10 ТБ (учитывая накопление тренировок, соц-контента, логи).
- Нагрузка: новые челленджи, массовые мероприятия, IoT-данные растут.
- **Год 5:**
 - Кратное увеличение (1–2 млн пользователей, при удачном маркетинге и глобальной экспансии).
 - Объём данных: сотни ТБ (учитывая, что каждая тренировка и социальность несёт большие объёмы).
 - Высокая пиковая нагрузка — десятки тысяч одновременных запросов и постоянный поток IoT-событий.

Конечно, реальные цифры могут сильно отличаться в большую или меньшую сторону.

3. Примерный расклад годовых затрат (грубая модель)

Ниже — **очень упрощённый** ориентир (USD в год) при среднерыночных облачных расценках. В каждой категории даётся **вилка** (минимум-максимум), поскольку многое зависит от конкретных параметров (скидки провайдера, регион, трафик). Заложена базовая 24/7 поддержка и условная команда.

3.1. Год 1 (MVP, ограниченная аудитория)

1. Инфраструктура

- Kubernetes-кластер(ы), VMs, управляемая БД, Kafka.
- При 100k пользователей и небольшом объёме — можно уложиться в **~\$10k–\$15k в месяц** (порядка \$120k–\$180k в год).
- Это предполагает несколько десятков vCPU, несколько управляемых баз (SQL/NoSQL), умеренный объём S3/BLOB-хранилища.

2. Хранение и передача данных

- 1–2 ТБ данных в облачном объектном хранилище + NoSQL/SQL.

- Передача данных (egress) ещё не очень велика: ~\$2k–\$5k в месяц на трафик (зависит от географии и CDN), итого ~\$24k–\$60k в год.

3. Подписки и лицензии

- Часто на старте можно пользоваться open-source DevOps-инструментами, не платя больших лицензионных сборов.
- Бюджет ~\$10k–\$50k в год на случай использования дополнительных платных сервисов (например, Splunk Cloud для логов, некоторые ML-сервисы и т.д.).

4. Человеческие ресурсы

- Минимальная команда: 1–2 DevOps, 5–7 разработчиков, 1–2 QA, 1–2 аналитика/ML-специалиста (при старте).
- Всё зависит от региона, но для упрощения считаем, что расходы на такую команду могут начинаться **от \$1M в год** (если это квалифицированная команда в западных странах/мегаполисах).
- Если часть команды в более доступных регионах, можно снижать. Но в любом случае это часто **самая весомая статья**.

Итог по Году 1

- Инфраструктура + хранение/трафик + подписки: **\$200k–\$300k** (примерно).
- Команда (з/п + налоги + overhead): **\$1M–\$1.5M**.
- **Общая вилка:** от **\$1.2M** до **\$1.8M** в год (может быть и выше, если офисы в дорогих локациях и нет скидков).

3.2. Год 2 (увеличение трафика и данных в 3–5 раз)

1. Инфраструктура

- Нужно масштабировать Kubernetes-кластеры, базы, Kafka кластеры.
- Затраты могут возрасти до **\$20k–\$40k в месяц** (т.к. растут объёмы запросов), т.е. до ~\$240k–\$480k в год.

2. Хранение и передача данных

- Уже 5–10 ТБ на хранении, активная соцсеть, рост IoT-потока.

- Трафик может «съедать» ~\$5k–\$15k в месяц, итого \$60k–\$180k в год.
- Плюс возможно расширение Data Lake: ещё ~\$50k–\$100k в год на хранение, если начинаем собирать большой объём «сырых» данных.

3. Подписки и лицензии

- Возможно, компания переходит на платные решения для мониторинга, APM, SIEM, enterprise-поддержку Kafka, и т.д.
- Затраты могут подняться до ~\$50k–\$150k в год.

4. Человеческие ресурсы

- Количество пользователей растёт, появляются дополнительные задачи (поддержка, локализация, интеграции).
- Команда может вырасти до 15–20 человек (или более). Это увеличивает фонд оплаты труда до ~\$2M–\$3M в год (при условии сохранения уровня квалификации).

Итог по Году 2

- Инфраструктура + хранение/трафик + подписки: ~\$400k–\$800k
- Человеческие ресурсы: ~\$2M–\$3M
- **Общая вилка: \$2.4M–\$3.8M в год.**

(При резком успехе и большем росте аудитории затраты могут вырасти ещё сильнее.)

3.3. Год 5 (масштаб глобального уровня)

Предположим, к 5-му году у нас 1–2 млн реальных пользователей (получилось раскрутить бренд, отвоевать рынок) и накоплено **сотни ТБ** данных, а также продвинутая аналитика (ML, real-time рекомендации, глобальные челленджи).

1. Инфраструктура

- В нескольких регионах/облаках кластеры Kubernetes, продвинутый CI/CD, большой Kafka/streaming, обслуживающий сотни миллионов сообщений в день.

- Месячные расходы на облако могут достичь **\$100k–\$300k** (или выше, если очень большие объёмы), то есть **\$1.2M–\$3.6M** в год.

2. Хранение и передача данных

- Сотни ТБ в Data Lake, активная NoSQL-репликация, архивация старых данных, частые batch-загрузки.
- Трафик между регионами может стоить **десятки тысяч** долларов в месяц.
- Совокупно (storage + egress + DB) это может вырасти до **~\$500k–\$1M+** в год.

3. Подписки и лицензии

- При большом масштабе часто приобретаются Enterprise-подписки (Confluent Kafka, Datadog/Splunk, большие ML-платформы), в совокупности **\$200k–\$500k** в год — а то и выше.

4. Человеческие ресурсы

- Полноценная распределённая команда (30–50+ человек): разработчики, мобильщики, DevOps, SRE, Data Engineers, ML-инженеры, аналитики, support, менеджеры продукта.
- Общий ФОТ (фонды оплаты труда) может составлять **\$5M–\$10M** в год (если в основном команда расположена в регионах с «западными» зарплатами).

Итог по Году 5

- Инфраструктура + хранение/трафик + подписки: **\$2M–\$5M** (или выше, в зависимости от агрессивного роста).
- Команда: **\$5M–\$10M**.
- **Общая вилка: \$7M–\$15M+** в год.

4. Вывод

1. **На старте** (Год 1) система обходится в среднем **\$1–2M** в год (при условии полноценной команды и базовых облачных расходах).
2. **Ко 2-му году** при трёх-пятикратном росте аудитории и данных, ТСО может достичь **\$2.5–4M** в год.

3. К 5-му году, если проект развивается глобально (миллионы пользователей и большие объёмы данных), совокупные затраты (инфраструктура + команда) могут перевалить **\$7–15M+** в год.

Компромиссы:

- Можно экономить на некоторых аспектах (например, использовать больше open-source и self-managed решений вместо дорогих Enterprise-подписок), но тогда растут риски по надёжности и время, затрачиваемое командой на поддержку.
- Можно размещать часть сервисов on-premise (для экономии), но придётся платить за «железо», поддерживать его и всё равно планировать затраты на масштабирование и сетевую инфраструктуру.

Главный фактор:

- **Зарплаты и команда** — обычно крупнейшая статья расходов. Высококвалифицированные DevOps, Data Engineers, ML-специалисты, Fullstack-разработчики стоят дорого, особенно в развёрнутых проектах уровня «миллионы пользователей» и «сотни ТБ данных».

Таким образом, **стоимость владения** (ТСО) в долгосрочной перспективе растёт в несколько раз по сравнению с начальными этапами — что характерно для систем, ориентированных на **большой охват аудитории, анализ больших данных и высокие стандарты надёжности и безопасности**.

Детализированные расчёты затрат на первый год владения системой

Ниже приведены **примерные** (но уже более **детализированные**) расчёты затрат на **первый год** владения системой, объясняющие, откуда появляются суммы в районе **\$1–2 млн**. Нужно учесть, что каждая компания имеет свои тарифы, скидки от облачных провайдеров и зарплатные вилки, поэтому реальные цифры в конкретном проекте могут отличаться.

1. Общая структура затрат в первый год

Чтобы понять, «откуда взялось» итоговое число, разложим его на четыре составляющие:

1. **Инфраструктура (облачные ресурсы)**
2. **Хранение и передача данных**
3. **Подписки и лицензии**
4. **Человеческие ресурсы (зарплаты, налоги, overhead)**

В конце подытожим, какой ориентир по каждому блоку складывается в течение **12 месяцев**.

2. Детализация инфраструктурных расходов

Допустим, у нас ~100k пользователей (из них 5–10% активных ежедневно), базовые микросервисы (10–12 сервисов), и мы решили использовать **облачную инфраструктуру** с Kubernetes.

2.1. Kubernetes-кластер (Production + Staging)

- **Compute (виртуальные машины / узлы кластера):**
 - Допустим, у нас 6 узлов (по 8 vCPU, 32 ГБ RAM) на продакшене и 3 узла (поменьше) на staging.
 - Приблизительная стоимость в публичном облаке может составлять ~\$0.3–\$0.4/час за узел.
 - Итого, на 6 узлов прод + 3 узла staging = $9 \text{ узлов} * 0.3\$ * 24 * 30 \approx \$1944/\text{месяц}$ только за «голые» узлы (в реальности чуть выше, если брать запас по ресурсам, аварийные ноды и т.д.).
 - Плюс **дополнительные сервисы**: плата за master (если не managed), load balancers, NAT-gateways, IP-адреса. Это может добавить ещё \$200–\$500/месяц.
- **Резерв для автоскейлинга:**
 - Во время рекламных акций или челленджей мы можем автоматически поднимать +2–3 узла, добавляя ~\$500–\$1000/месяц.

- Итого, суммарно на Compute + overhead K8s получаем **\$2500–\$3500/месяц**.

2.2. Базы данных

- **SQL (реляционная база):**
 - Допустим, для User Management + транзакций (покупки) нам нужна управляемая СУБД (например, RDS, CloudSQL).
 - Средний класс (db.m5.large, db.m5.xlarge) ~\$0.3–\$1/час, плюс плата за storage (100–200 ГБ).
 - Примерно **\$1000–\$1500/месяц**.
- **NoSQL** (например, MongoDB Atlas, DynamoDB, Cassandra-кластер):
 - На старте, при не очень больших объёмах, ~\$500–\$800/месяц.
 - Это покрывает соц. ленту и базовые тренировочные данные.
- **Time Series DB** (InfluxDB, Timescale Cloud, etc.):
 - Для подробных тренировочных метрик (пульс, шаги), на старте может быть ~\$300–\$500/месяц (управляемые планы).

2.3. Очередь/шина (Kafka или RabbitMQ)

- Kafka (Confluent Cloud, MSK и т.д.):
 - Начальный кластер (несколько broker'ов) может стоить **\$1000–\$2000/месяц**.
 - Если берём open-source на тех же узлах Kubernetes, часть затрат включается в compute. Но, как правило, всё же бюджет на поддержку Kafka вылезает (доп. узлы, диски).

2.4. Прочие сервисы и сетевой трафик

- Load Balancers, NAT, egress-трафик:
 - В облаке за каждый балансировщик (LB) и IP могут брать \$20–\$50/месяц, плюс плата за гигабайты исходящего трафика.

- Если приложение активно используется (особенно при CDN), можно заложить ~\$500–\$1000/месяц на эти сетевые сервисы.
- Резервное копирование (Backups), мониторинг (Prometheus/Grafana в облаке или платный Datadog), логи:
 - Могут добавить ещё **\$300–\$1000/месяц**.

Итого по инфраструктуре (в месяц)

Складываем грубо:

1. **K8s compute + overhead:** \$2500–\$3500
2. **SQL:** \$1000–\$1500
3. **NoSQL + TSDB:** \$800–\$1300 (суммируя)
4. **Kafka:** \$1000–\$2000
5. **LB + NAT + monitoring + backups:** \$800–\$2000

Итого в месяц: **\$6100–\$10,600** (очень примерная вилка).

Умножаем на 12 месяцев: получаем **~\$73k–\$127k в год** только за compute + базы + очереди + базовый monitoring.

Но в реальности:

- При пиках может понадобиться автоскейлинг, +\$1000–\$2000/месяц.
- В некоторых облаках дороже egress-трафик.
- Возможно, нужно больше узлов для staging, dev, QA.

Поэтому **\$120k–\$180k/год** за инфраструктуру — реалистичная вилка, которая раньше упоминалась.

3. Хранение и передача данных (подробнее)

3.1. Хранилище объектов (S3, Azure Blob, GCS)

- На 1-й год предполагаем до **1–2 ТБ** рабочих данных (включая файлы тренировок, соц-медиа, фото).
- Ставка хранения ~\$20–\$25/ТБ в месяц (в зависимости от региона и класса хранения).
- Итого: ~\$20–\$50/месяц => \$240–\$600/год, что довольно скромно.

Но куда дороже может стать:

3.2. Трафик (egress, CDN)

- Если пользователи со всего мира активно загружают/скачивают медиа, а CDN не всё кэширует, может выйти **\$1k–\$3k/месяц**.
- При 100k пользователей и не слишком активном стриминге, можно заложить ~\$2k/месяц => **\$24k/год**.

3.3. Data Lake (при включении Big Data/ML)

- На 1-м году MVP Data Lake часто небольшой. Но если мы хотим собирать все события (логи, телеметрию), легко набирается пару терабайт.
- Хранение сырых данных ~\$25/ТБ/месяц => +\$600/год за 2 ТБ.
- Если используется частая **batch-аналитика** (Spark), придётся платить за кластеры EMR/Dataproc (\$1k–\$2k в месяц), но на MVP можно обойтись редкими запусками.

Итого по блоку «хранение и трафик» (в год)

- **\$24k–\$60k** (учитывая, что egress-трафик может сильно колебаться в зависимости от географии и CDN).
- Если Data Lake активно используется, может добавиться ещё ~\$10k–\$30k.

4. Подписки и лицензии

4.1. DevOps-инструменты

- Часто для начала используют open-source: Jenkins, GitLab CE, Grafana, Loki, Prometheus.
- Но если нужен Datadog/Splunk/ELK Cloud, Atlassian (Bitbucket, Jira) Enterprise, то можно платить **\$1k–\$5k/месяц**.

4.2. Мониторинг, APM, SIEM

- При использовании SaaS Datadog/Logs/SIEM и объёмных логах может легко выйти ~\$2k–\$4k/месяц => \$24k–\$48k/год.
- Если берём только open-source, может быть ниже (но придётся тратить время DevOps).

4.3. ML-сервисы

- На 1-м году MVP часто обходятся минимальным ML, если оно не ключевое. Но, допустим, есть подписка на «enterprise ML-платформу», тогда +\$10k–\$30k/год.

Итого по подпискам (в год)

- **\$10k–\$50k** (если используем в основном open-source и недорогие планы).
- Или **\$50k–\$100k** (если берём enterprise-поддержку Kafka, Confluent Cloud, Datadog, Splunk).

5. Человеческие ресурсы (команда)

Это самая «гибкая» статья, зависящая от:

- Местоположения (США/Европа vs. Восточная Европа/Азия).
- Уровня специалистов (Junior, Middle, Senior).
- Требуемых компетенций (DevOps, Data Engineer, ML, Mobile, QA).

5.1. Минимальная команда для MVP

- **1 DevOps** (поддержка Kubernetes, CI/CD, IaC)
- **5–6 разработчиков** (Backend, Frontend/Mobile)
- **1 QA** (мануальный + автотесты)
- **1–2 Data/ML** (базовая аналитика, работа с NoSQL/TSDB)
- **1 продуктовый менеджер** (или менеджер проекта).

Итого **9–11 человек**.

Пример: если это Западная Европа/США

- Средний fully-loaded cost (зарплата + налоги + бенефиты) для среднего уровня (Middle) ~\$100k–\$150k/год на человека. Старшие инженеры (Senior) бывают дороже.
- Предположим, **средняя** ставка \$110k, а у 2–3 человек (ведущих) \$140k+.

Тогда:

- 9 человек * 110k \$/год = \$990k
- Пара ведущих инженеров (добавим \$30k на каждого) => +\$60k, вместе ~\$1.05M.
- Плюс иногда есть overhead (затраты на офис, оборудование, ПО, курсы) +5–10%. Получаем **\$1.1–\$1.15M**.

Если часть команды (или вся) в других регионах

- Затраты могут снизиться в 1.5–2 раза, но при этом учесть разницу часовых поясов, риски качества.
- Всё равно остаётся вилка **\$0.7–\$1M/год** как **суммарный** фонд оплаты труда+налоги+overhead.

5.2. Дополнительные роли

- Техподдержка (Support), работа с пользователями (если нужно 24/7) — отдельная статья. На первом году MVP иногда справляются разработчики + 1–2 человека первой линии.

6. Суммирование для Первого года

Сопоставим всё вместе:

1. **Инфраструктура:** ~\$120k–\$180k/год
2. **Хранение, трафик:** ~\$24k–\$60k (может быть чуть выше, если активная соцсеть).
3. **Подписки и лицензии:** ~\$10k–\$50k (условно).
4. **Человеческие ресурсы:** ~\$700k (если команда в более дешёвом регионе) до \$1.2–\$1.5M (если high-level команда на Западе).

Возьмём **некий усреднённый** сценарий:

- Инфраструктура: \$150k
- Хранение, трафик: \$40k
- Лицензии: \$30k
- Команда (10 человек, смешанная география): \$900k

Итого: \$1.12M.

Если же все работают в дорогом регионе, да ещё и с Enterprise-подписками, легко достигаем **\$1.5M–\$2M**.

7. Резюме и важные замечания

1. **Реальная стоимость** всегда **зависит** от:

- Нагрузки (пиковые/средние), числа окружений (Dev, QA, Stage, Prod).
- Типов и размеров баз данных, интенсивности IoT-трафика.
- Выбора технологий (open-source vs. Enterprise SaaS).

- Локации и структуры команды.
2. **Самая крупная статья** в большинстве случаев — **люди**. Инфраструктура по сравнению с ФОТ (фондом оплаты труда) может оказаться лишь 10–20% всей суммы.
 3. На первом году, когда продукт только растёт, ещё не накоплен огромный объём данных и нет постоянных челленджей со 100k одновременных пользователей, поэтому облачные расходы держатся относительно скромными (\$10–\$15k/месяц). Но **стоит** добавить 2–3 разработчика или нанять Senior ML/Architect — и общая сумма зарплат может вырасти на сотни тысяч долларов за год.

Таким образом, **\$1–\$2M** на первый год — это действительно **реалистичный** диапазон при масштабах в **100k пользователей** и серьёзной (но ещё не колоссальной) инфраструктуре. Нижняя планка (\$1M) характерна для комбинированной команды (частично в дешёвых регионах), активного использования open-source, аккуратных облачных расходов. Верхняя планка (\$2M) — если всё делается на высоком уровне, с Enterprise-подписками, полной командой в дорогом регионе, с резервом на масштабирование.