



HighLoad++  
FOUNDATION

Яндекс



# Как мы считаем просмотры видео и зрителей трансляций для десятков миллионов пользователей в день

Иван Соколов

Группа рантайма видеохостинга

Дзен





# Сервисы

Дзен	Яндекс.ТВ
Я.Станция	Поиск по видео
Новости	Маркет

Суммарное  
DAU – десятки  
миллионов  
пользователей

# Виды контента

Блогеры	Спортивные матчи
Фильмы	Новостные сюжеты
Сериалы	Записи телепередач

Суммарно –  
десятки миллионов  
единиц контента



# Рантайм видеохостинга

>10k RPS

Батч-запросы

~100ms







The graphic features a red background with a subtle pattern. At the top, there are three circular elements: the Russian flag, the FIFA World Cup trophy, and the Croatian flag. Below these, the text "FIFA WORLD CUP RUSSIA 2018" is displayed in white. A dark blue horizontal bar contains the text "Russia vs. Croatia" in white. Below this bar, the text "Спайки – суперпопулярные ролики и трансляции" is written in white.

FIFA WORLD CUP  
RUSSIA 2018

<https://www.footballorg.in.com/russia-v-croatia-full-match-replay-world-cup-2018-quarter-final-itv1/>

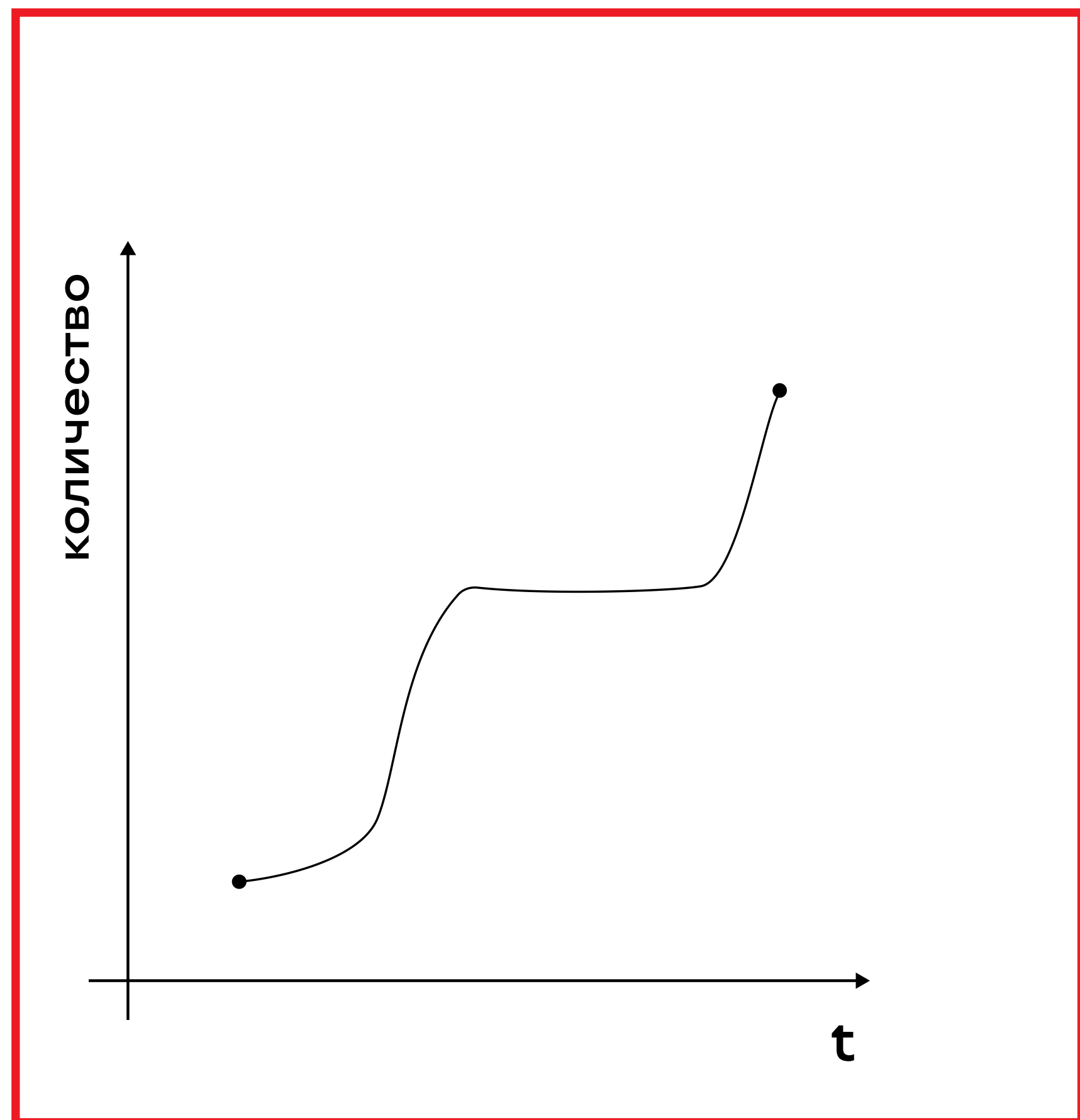
# Russia vs. Croatia

Спайки – суперпопулярные  
ролики и трансляции

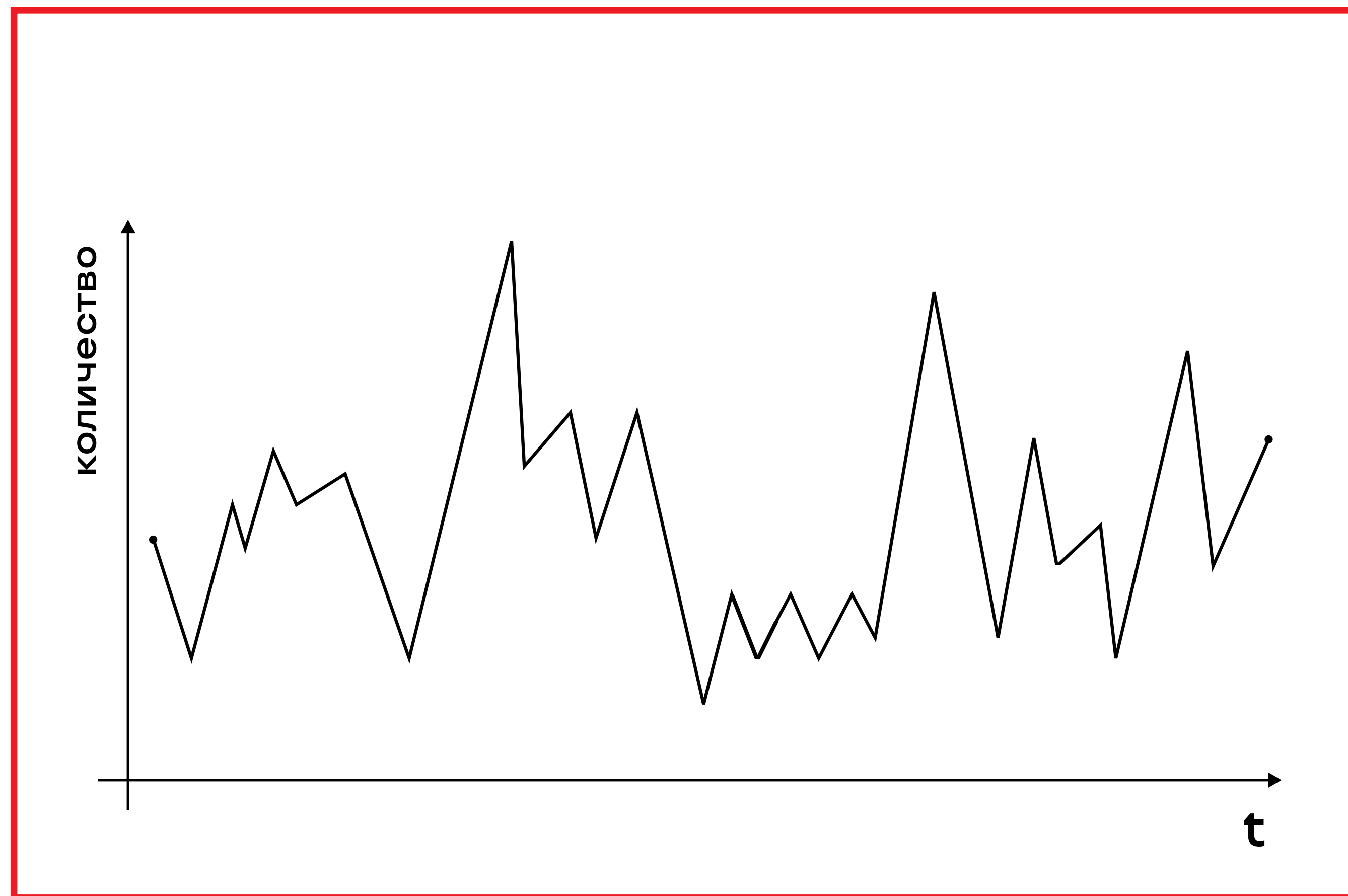




# Просмотры



# Зрители





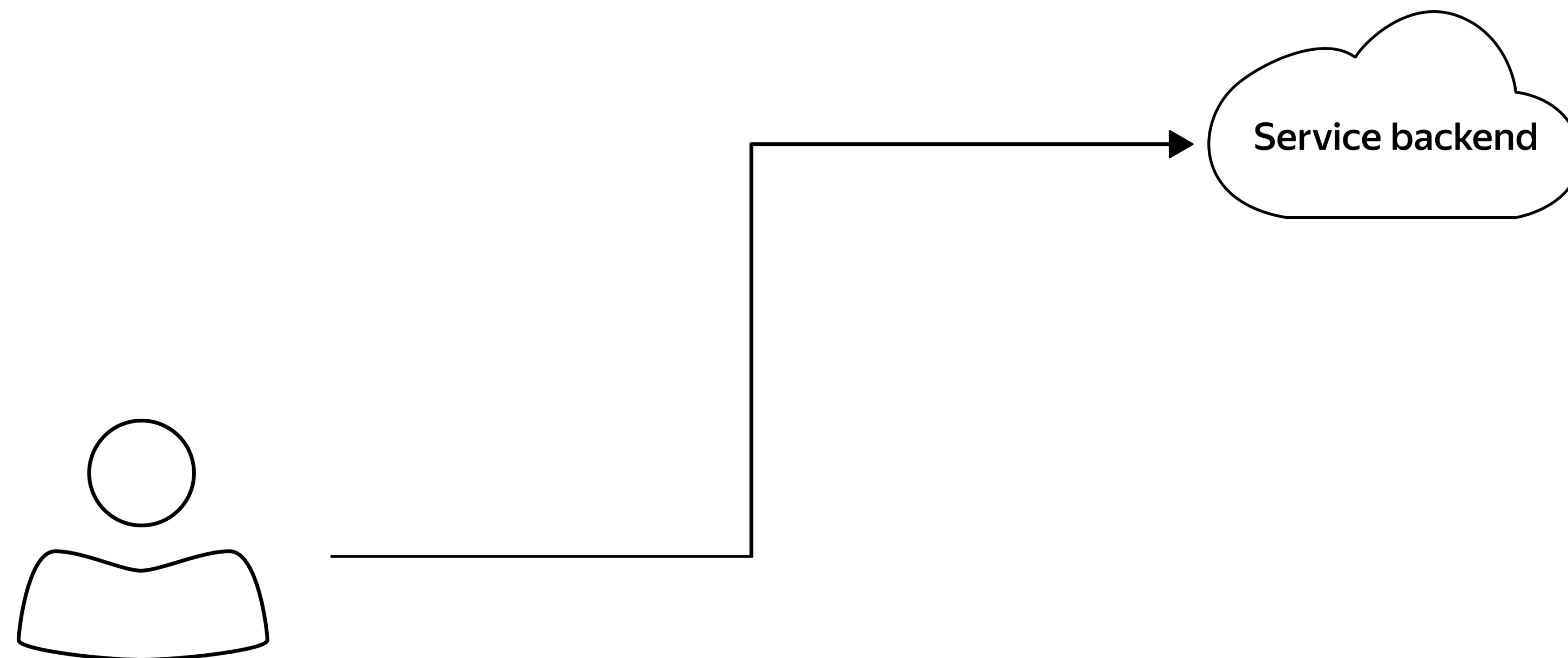
# 01

# Просмотры



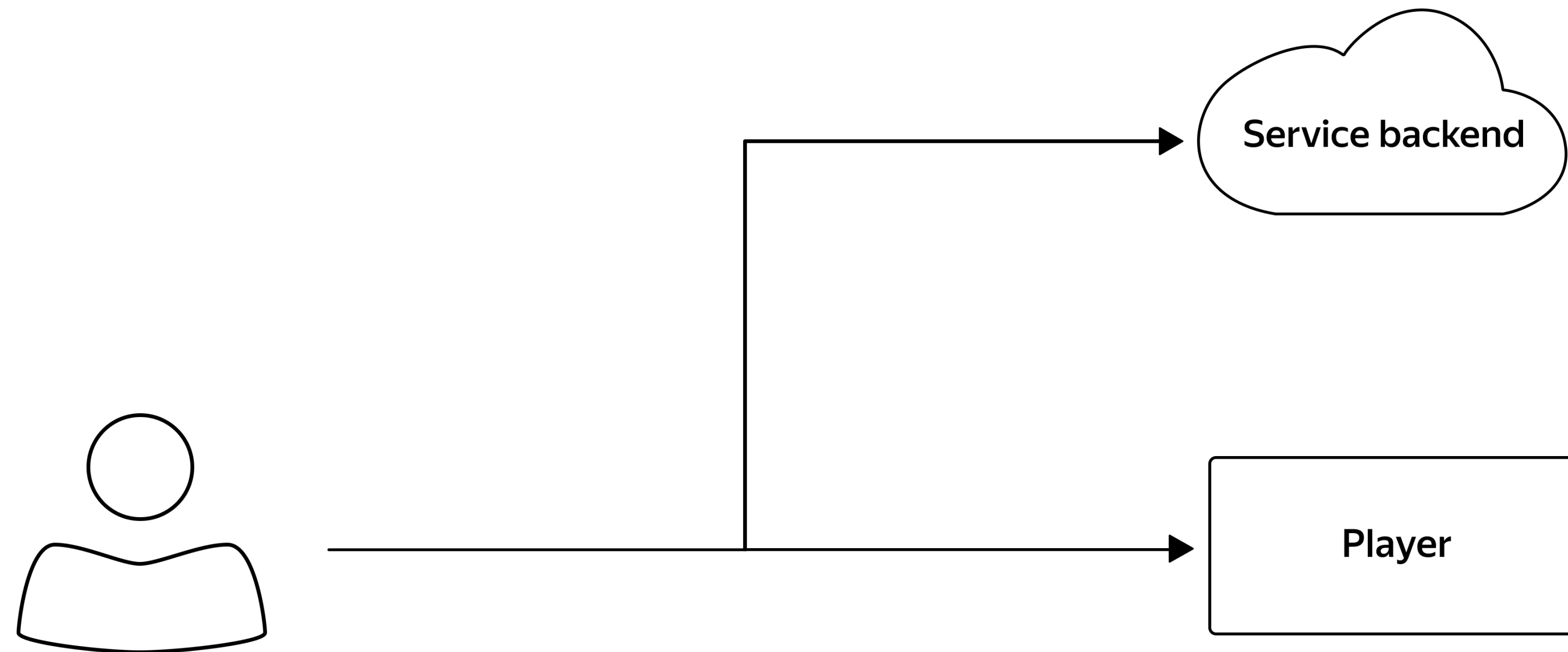


# Что такое просмотр



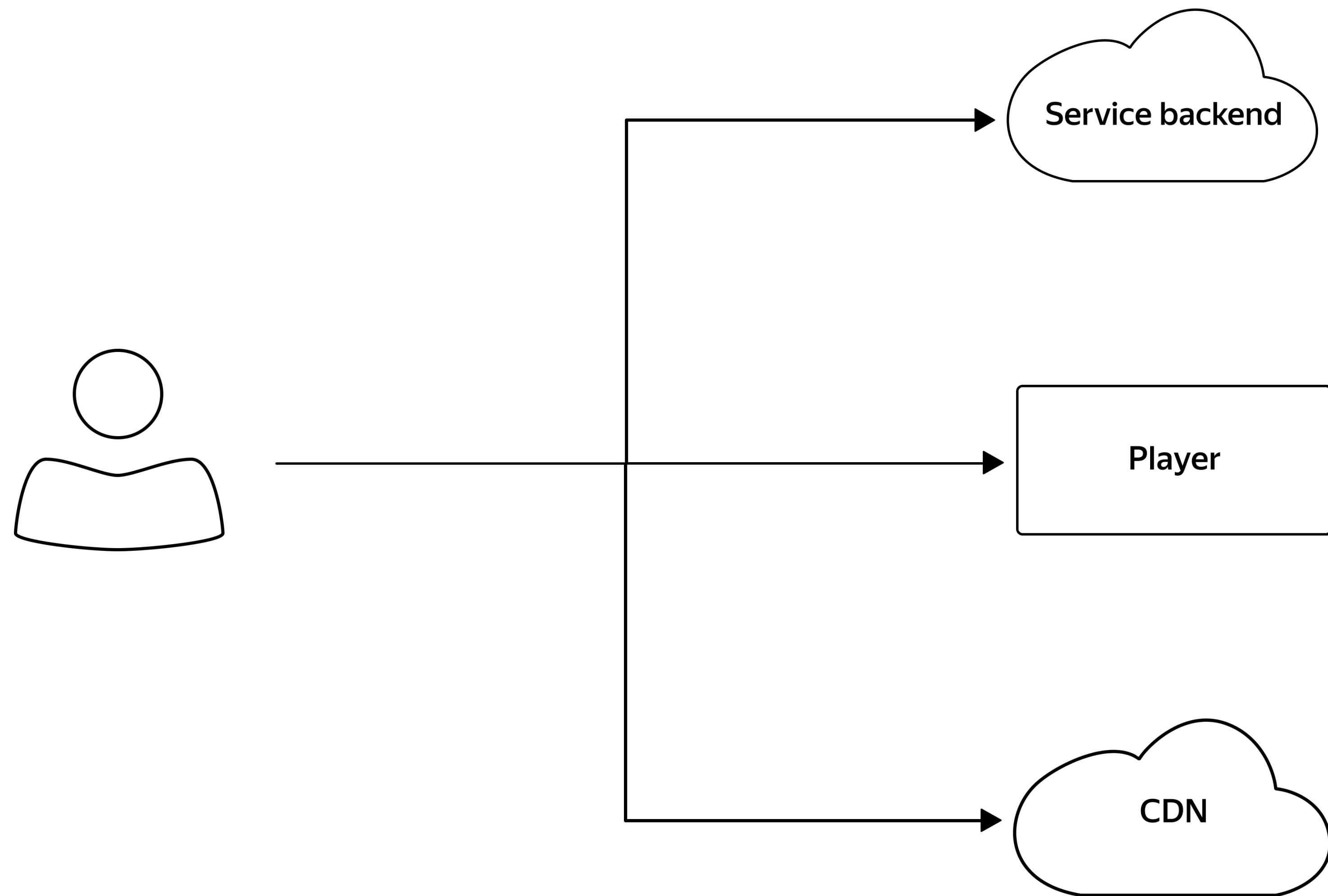


# Что такое просмотр



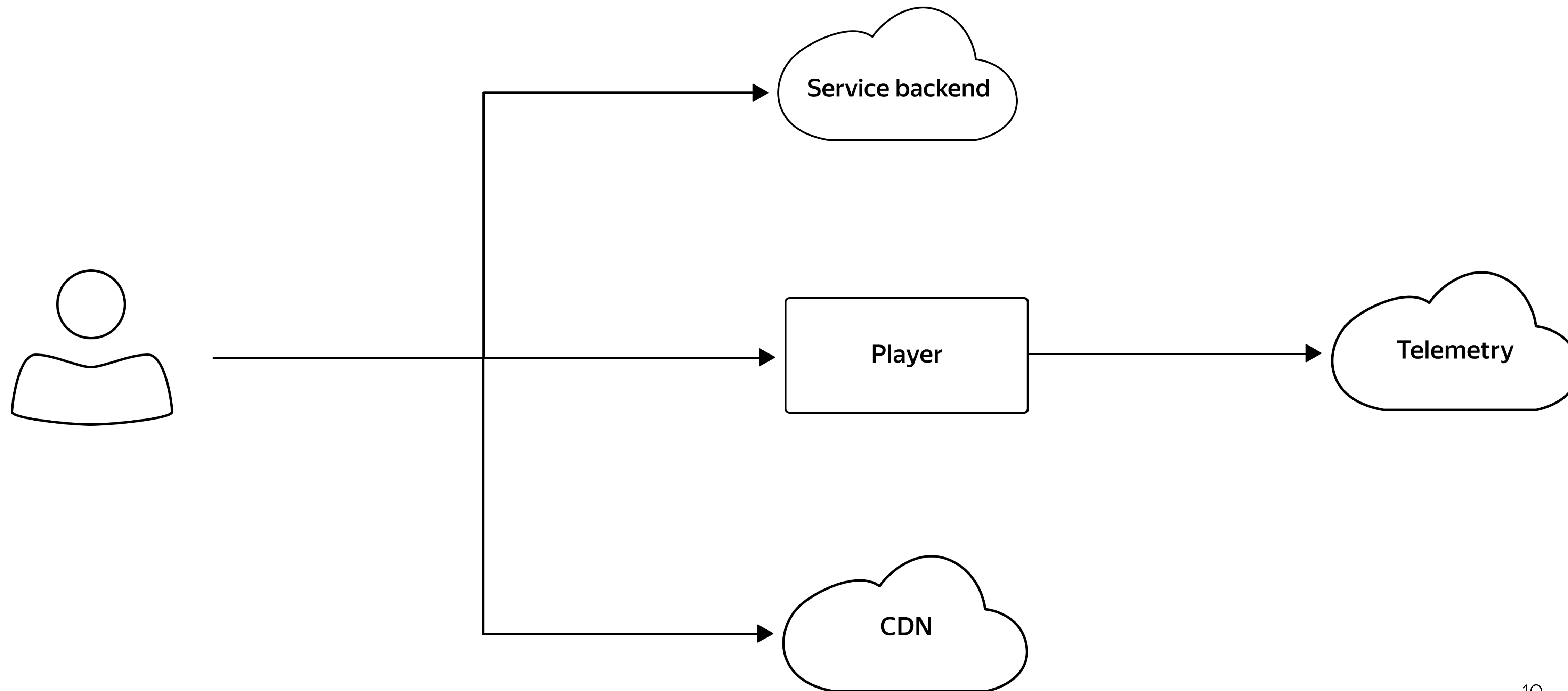


# Что такое просмотр



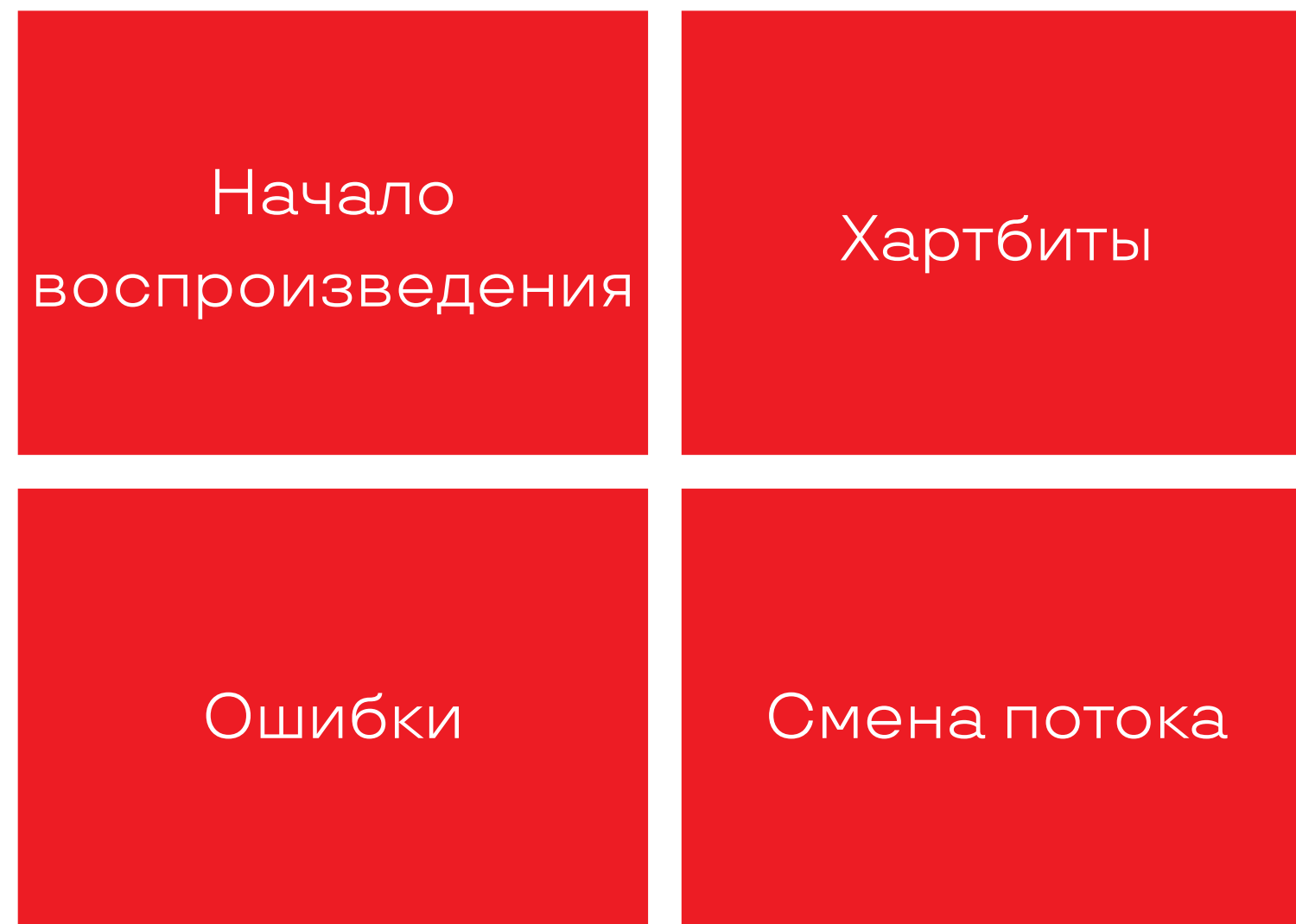


# Что такое просмотр

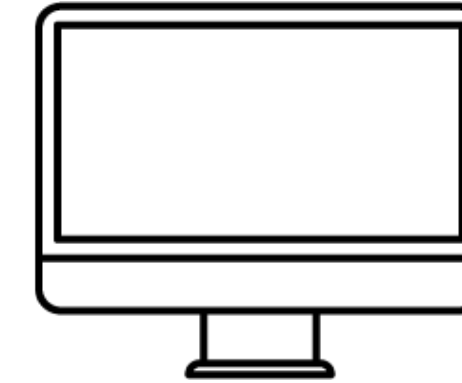




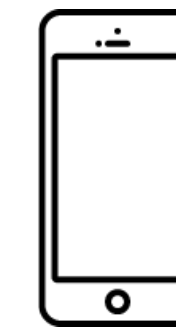
# Телеметрия



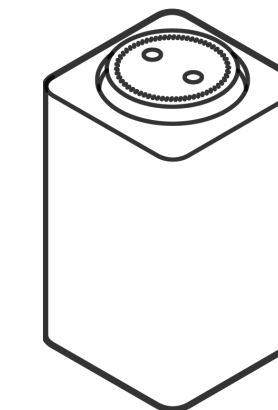
Десктопы



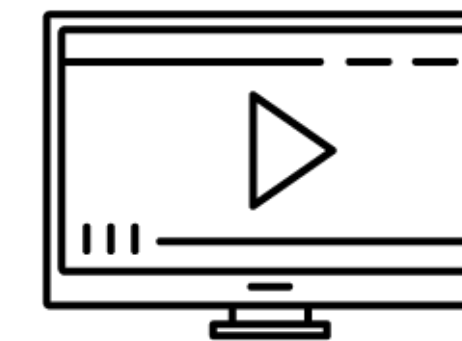
Мобильные



Умные устройства



ТВ





# Скорость доставки

+ Просмотры для выплат и аналитики считались по логам

+ Авторы ждали от часов до суток для получения статистики





© Netflix, Narcos, 2016





# Скорость доставки

+ Просмотры для выплат и аналитики считались по логам

+ Авторы ждали от часов до суток для получения статистики

+ Нужен счётчик, близкий к realtime





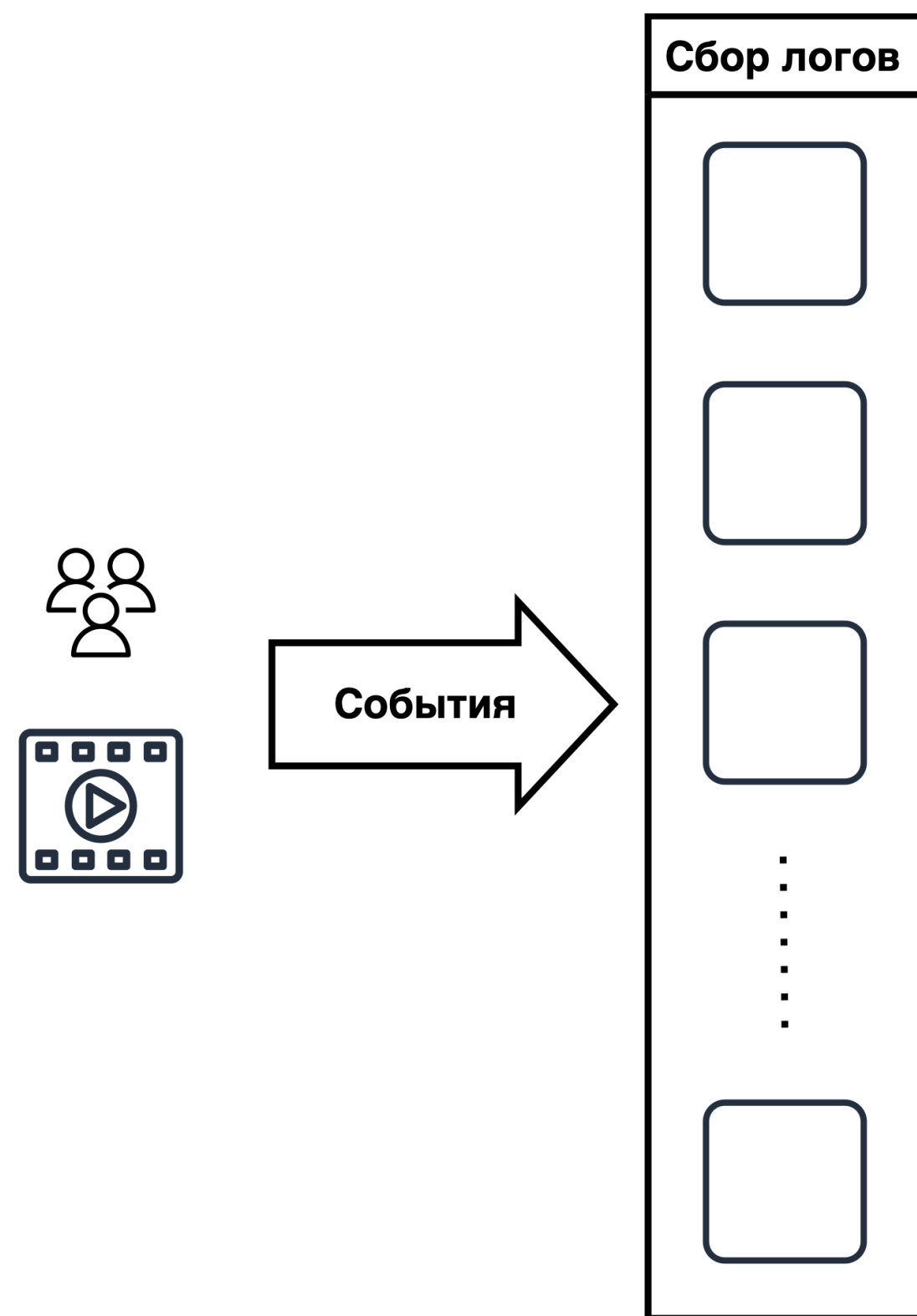
# 02

## Подсчёт и запись



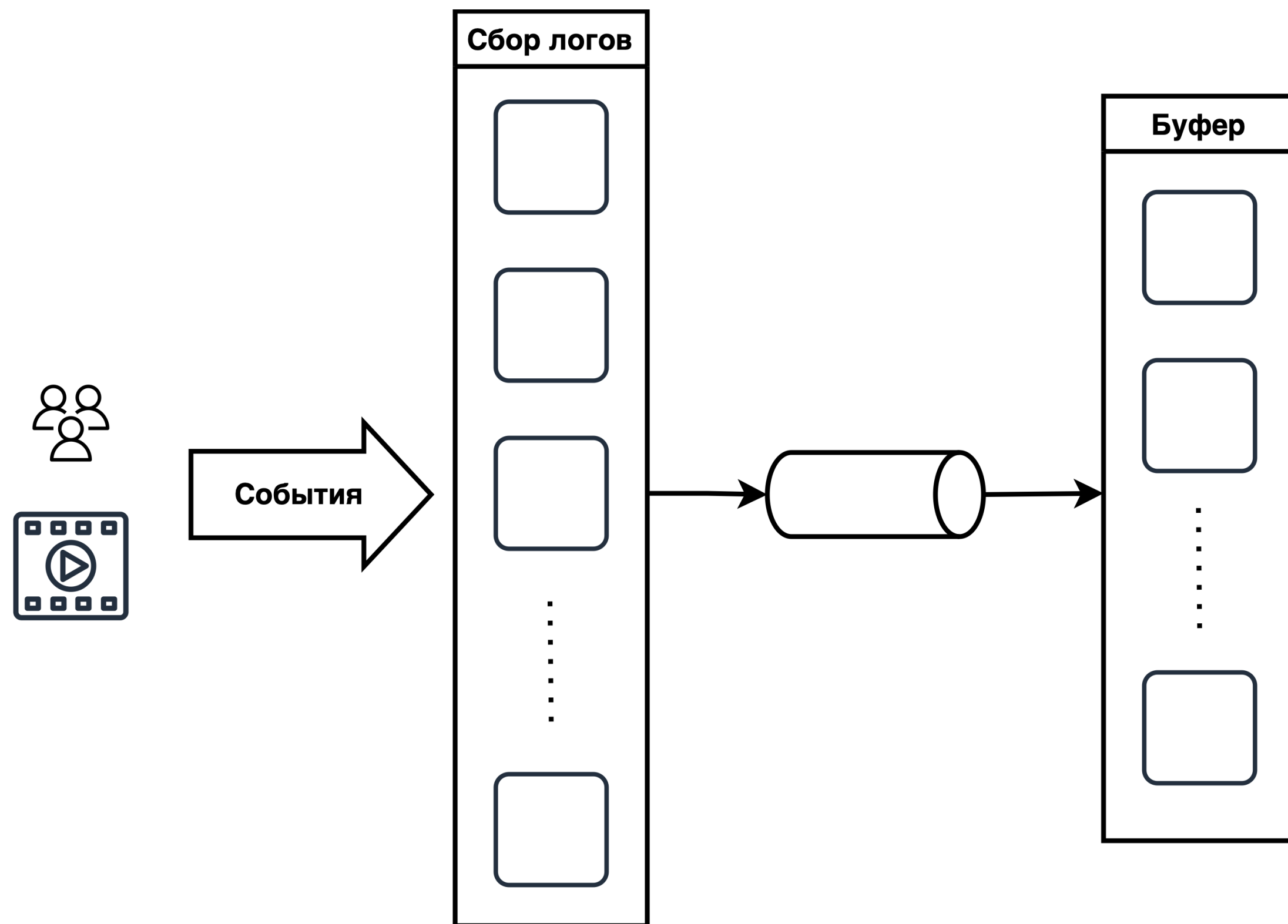


# Пайплайн



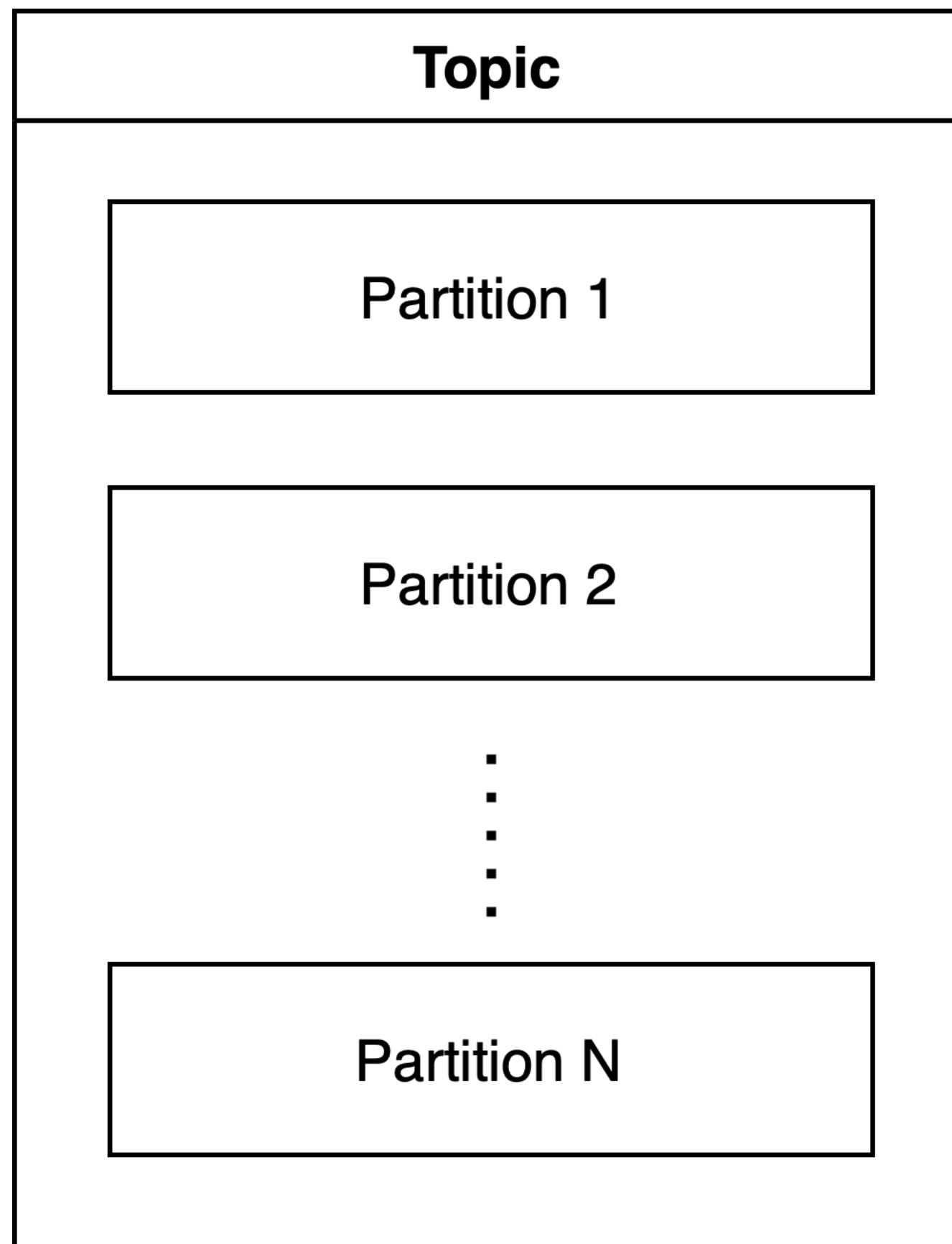


# Пайплайн



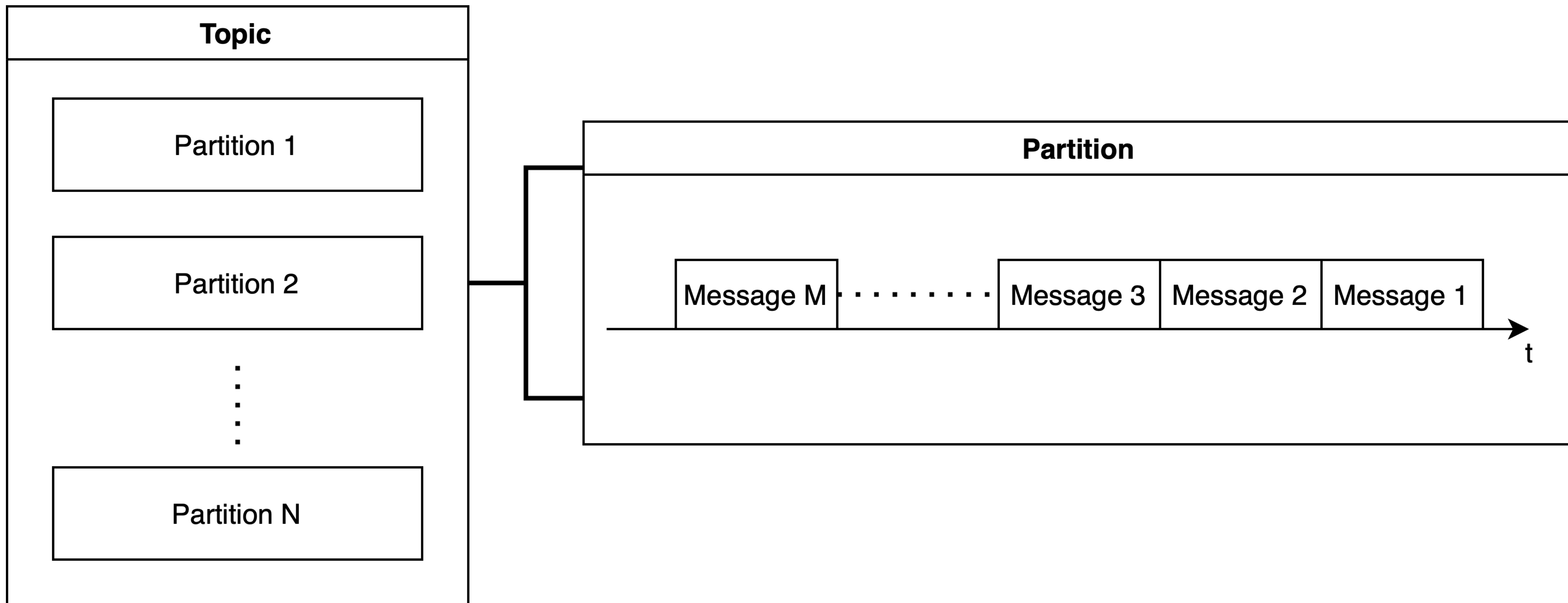


# Logbroker



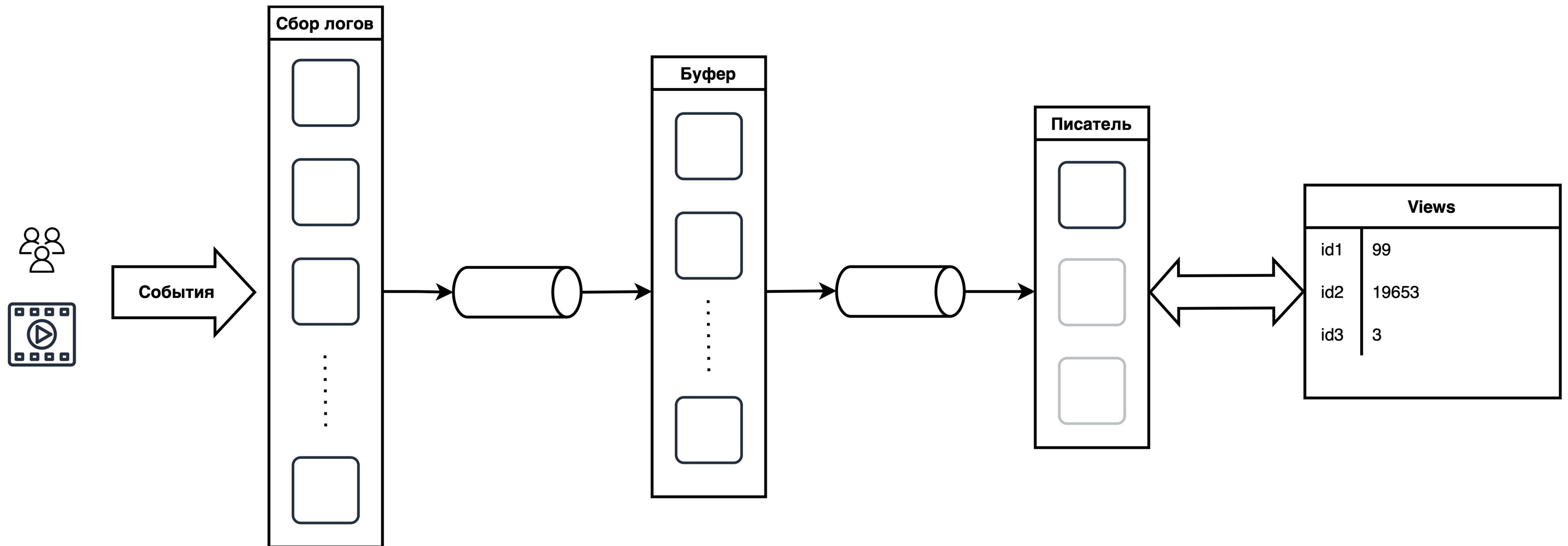


# Logbroker





# Пайплайн





# 03

## Чтение в рантайме





# Технологии

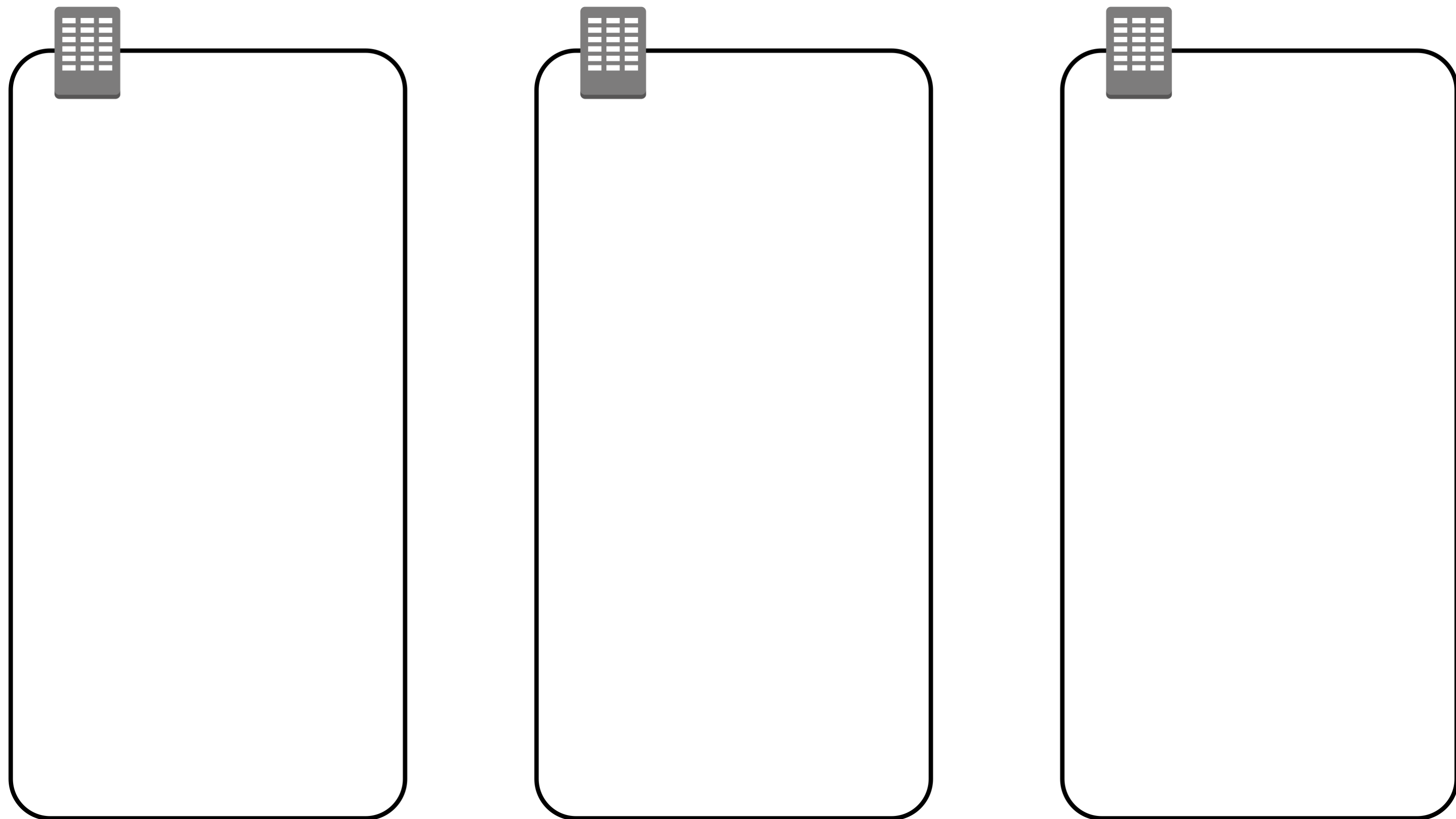
+ YDB – распределённая SQL СУБД

+ Apphost – API Gateway



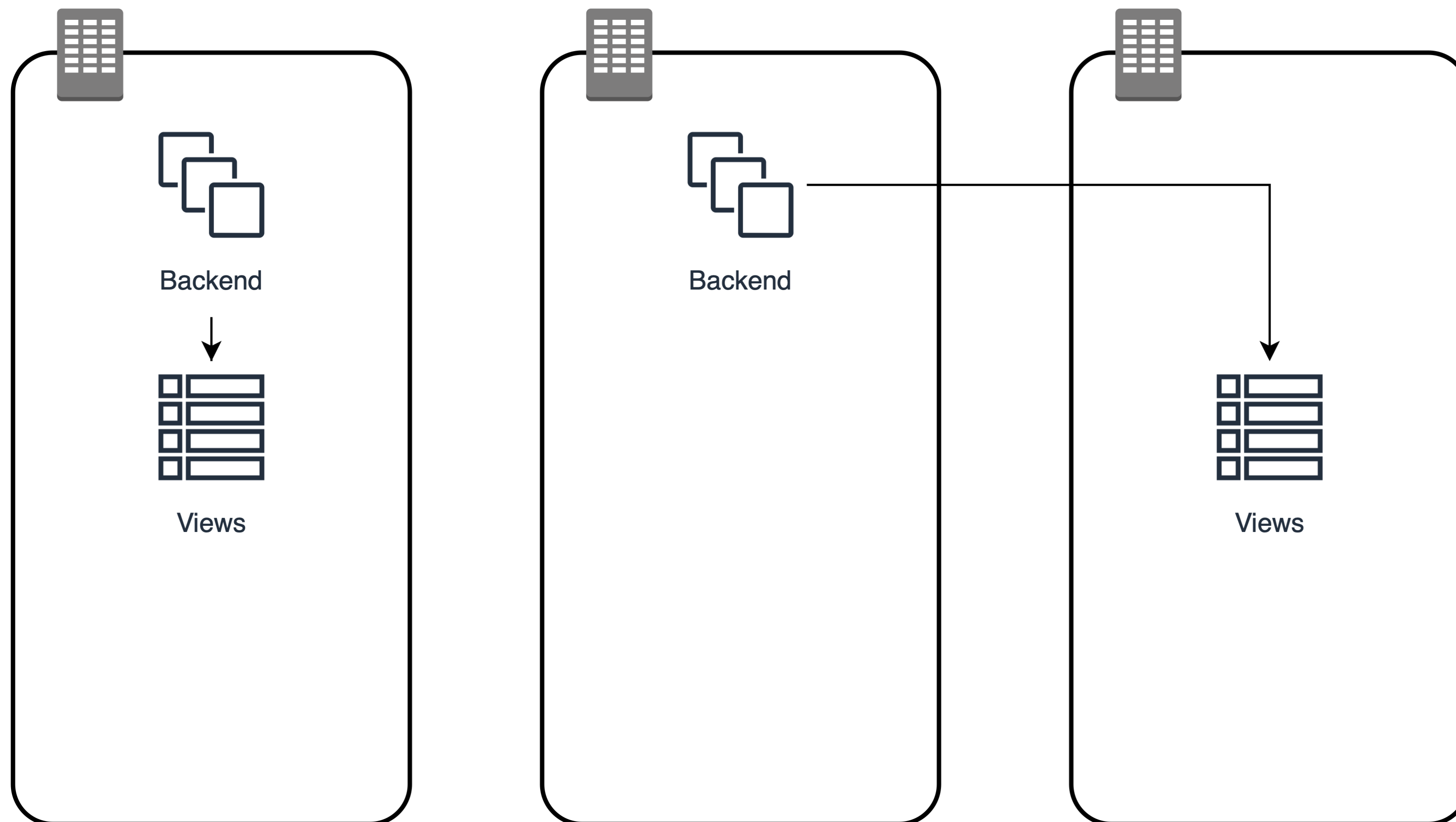


# Локальность данных



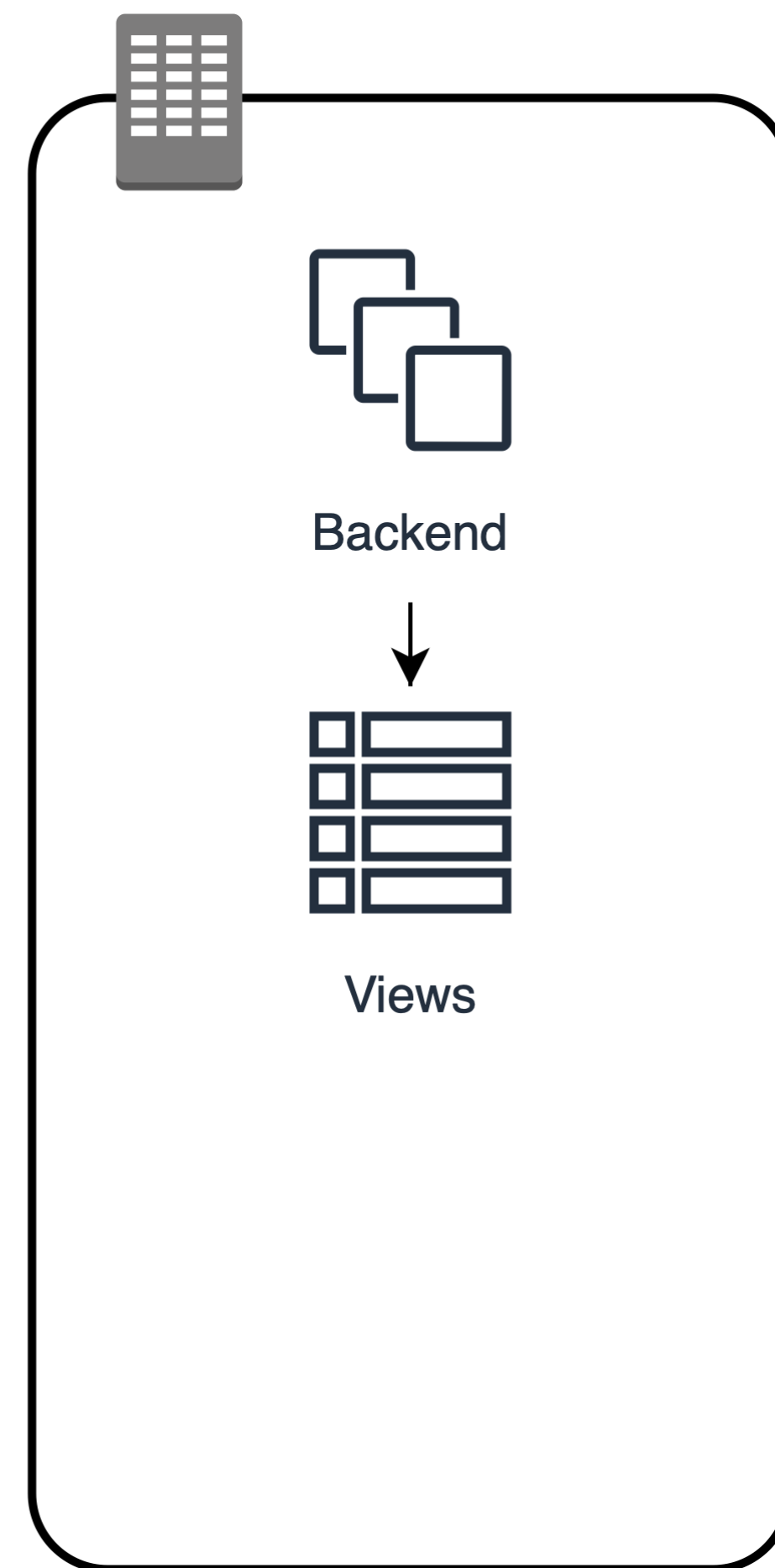
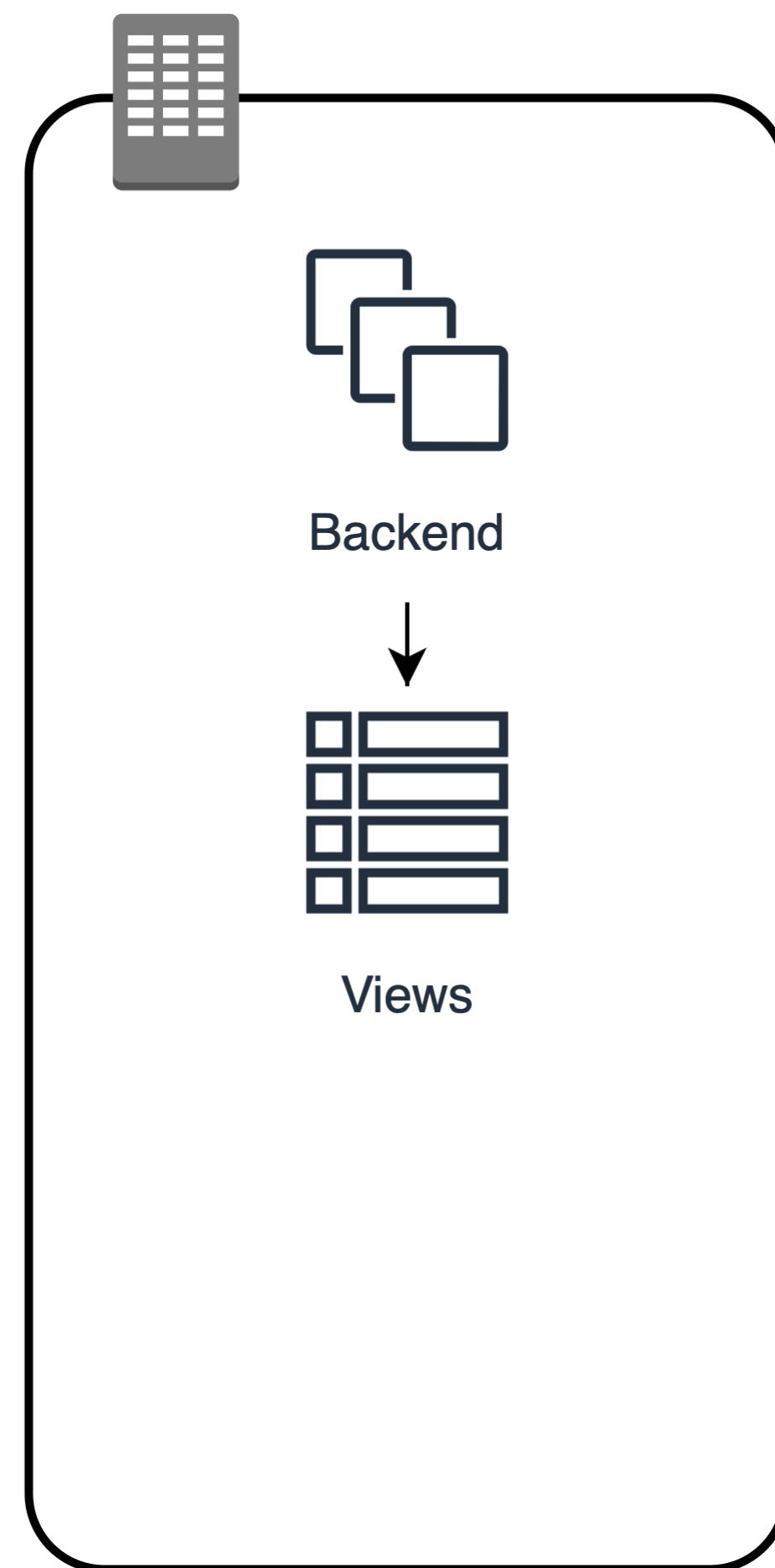
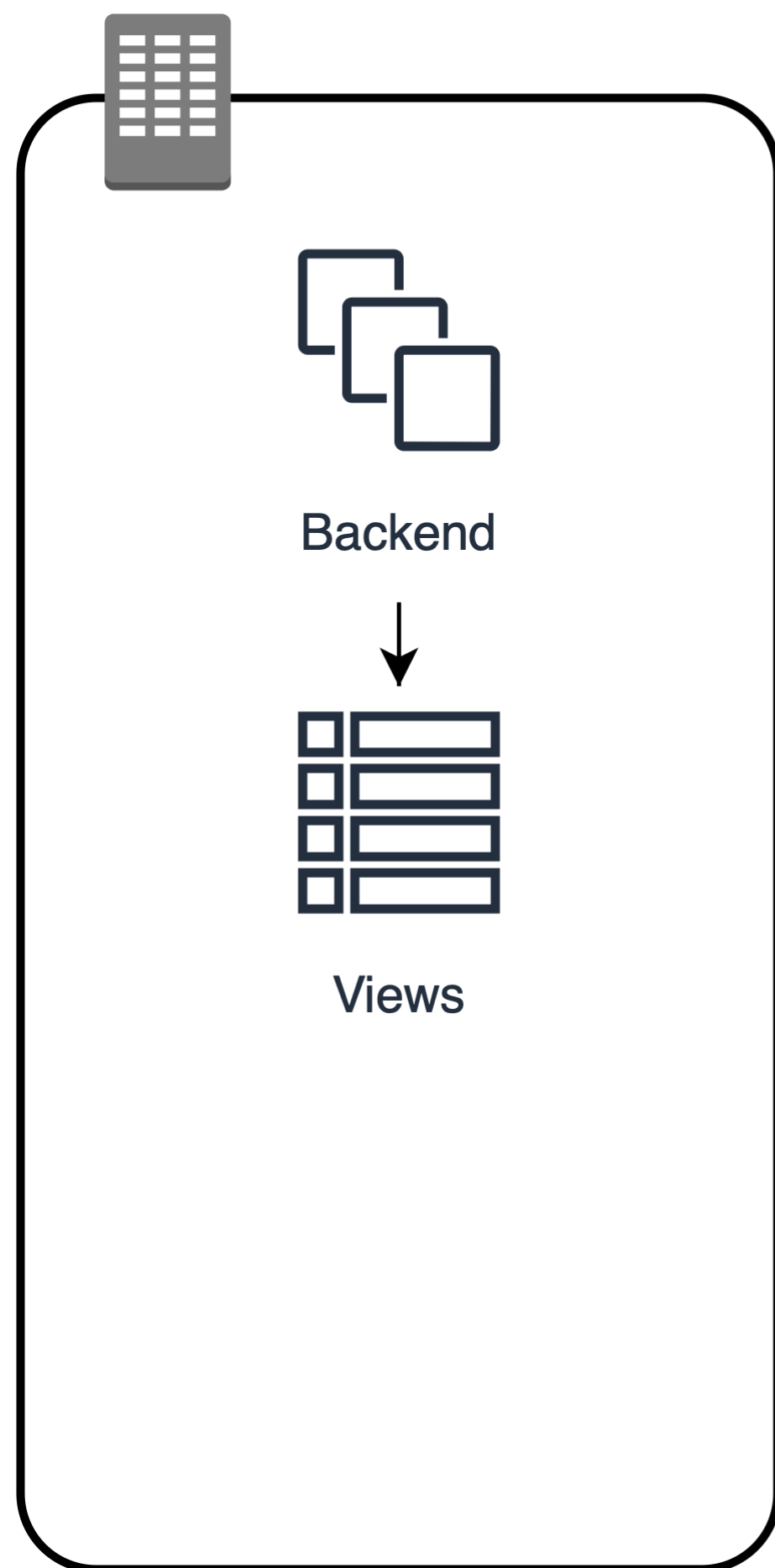


# Локальность данных



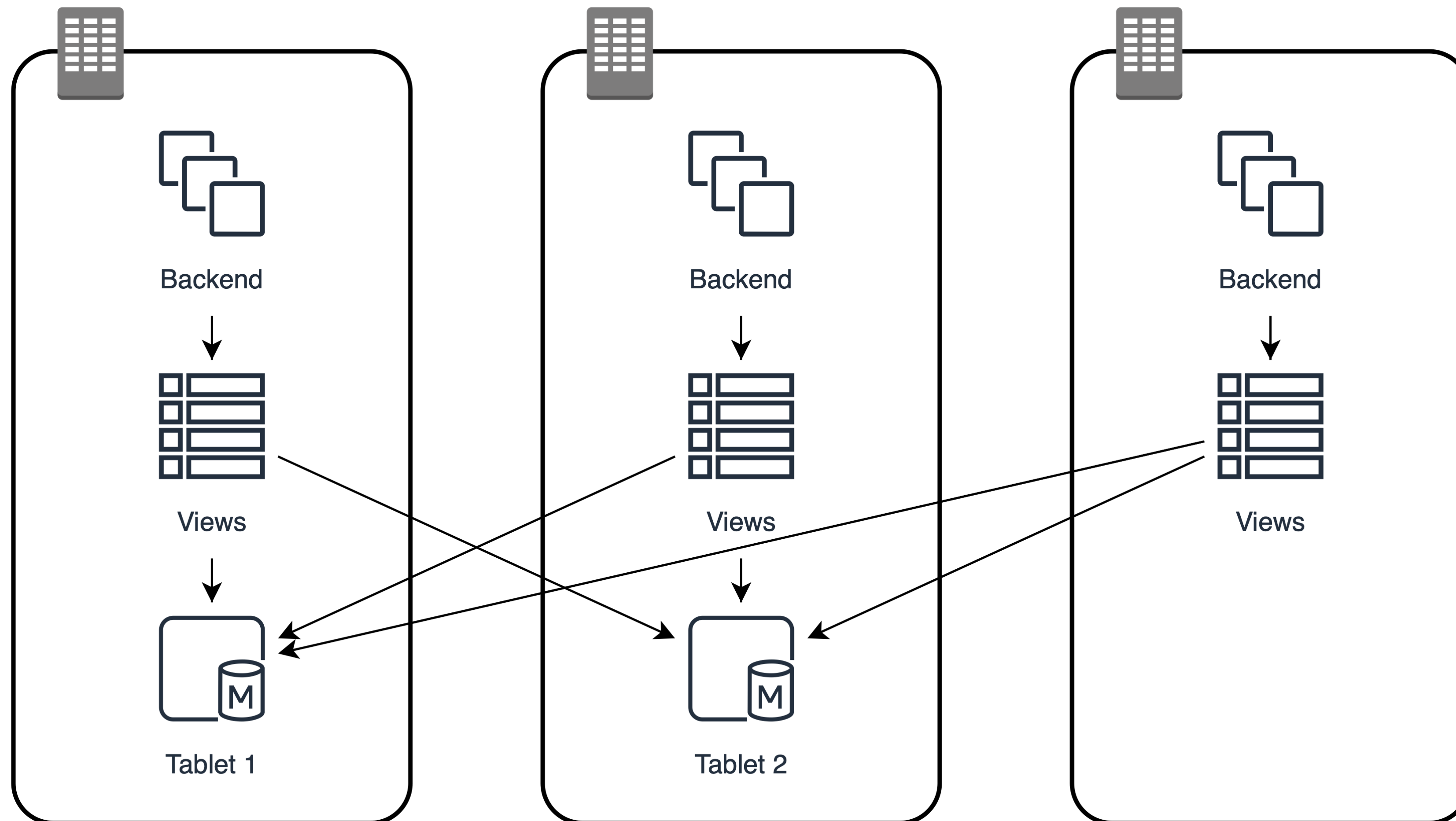


# Локальность данных



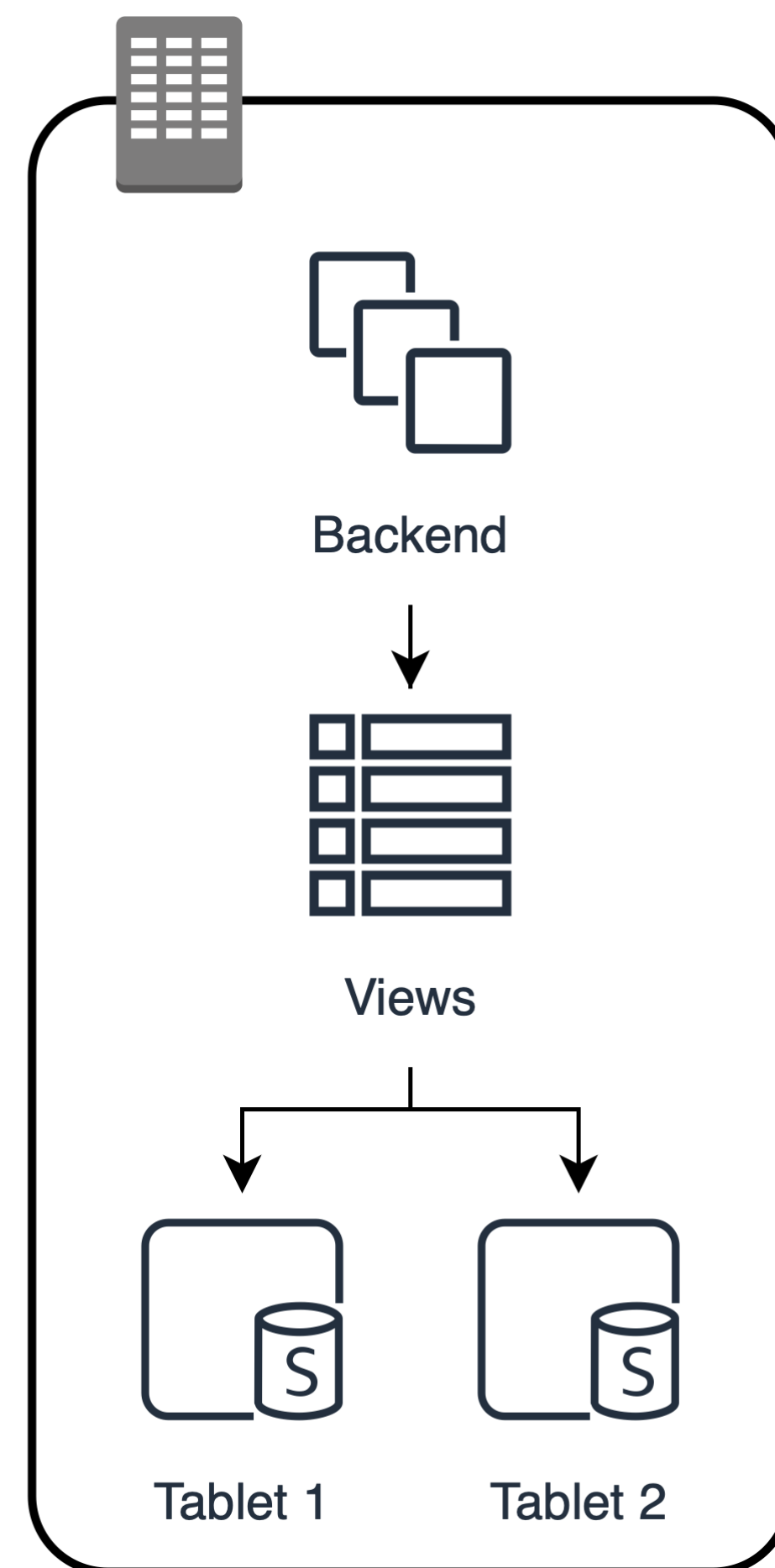
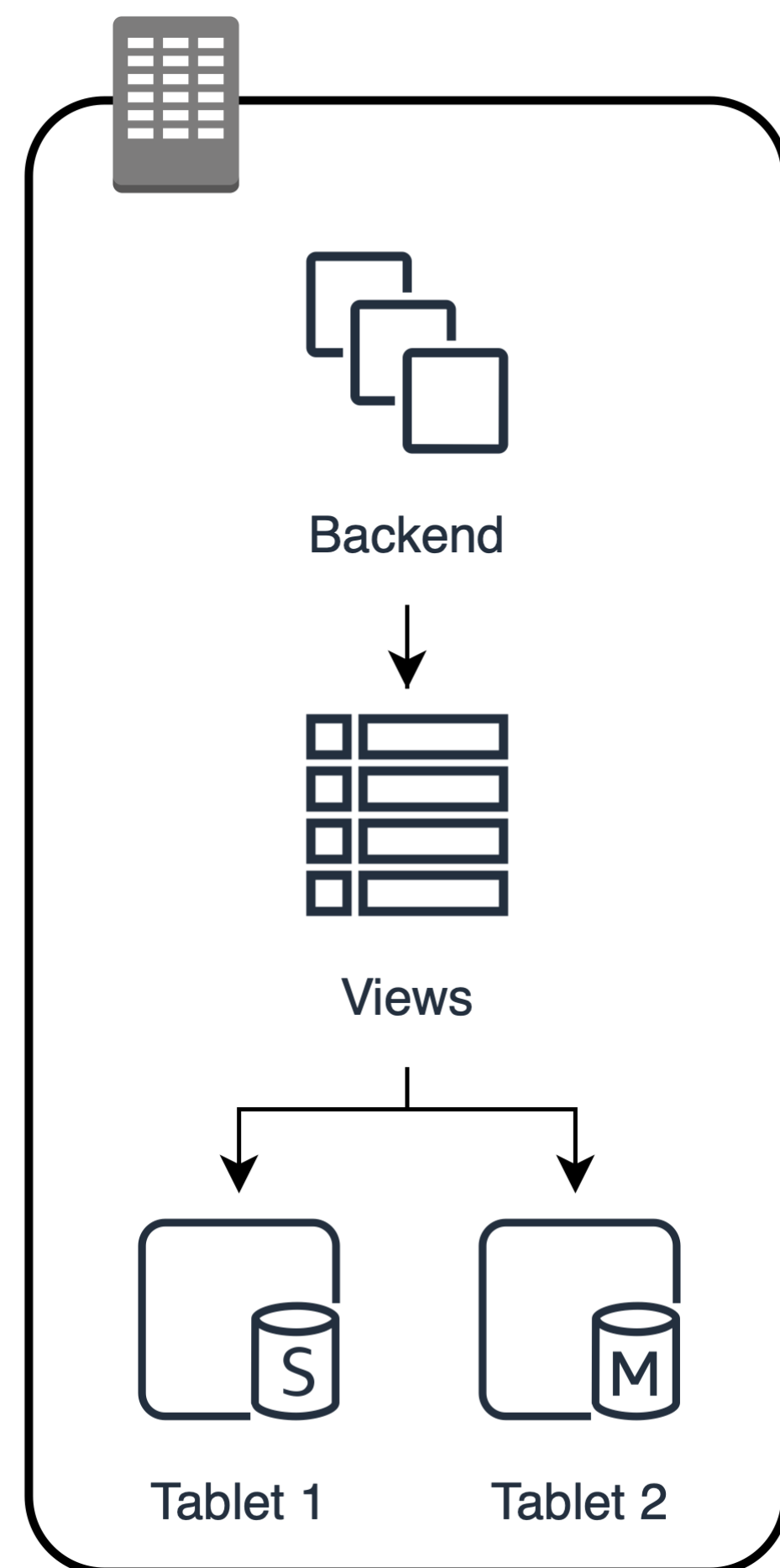
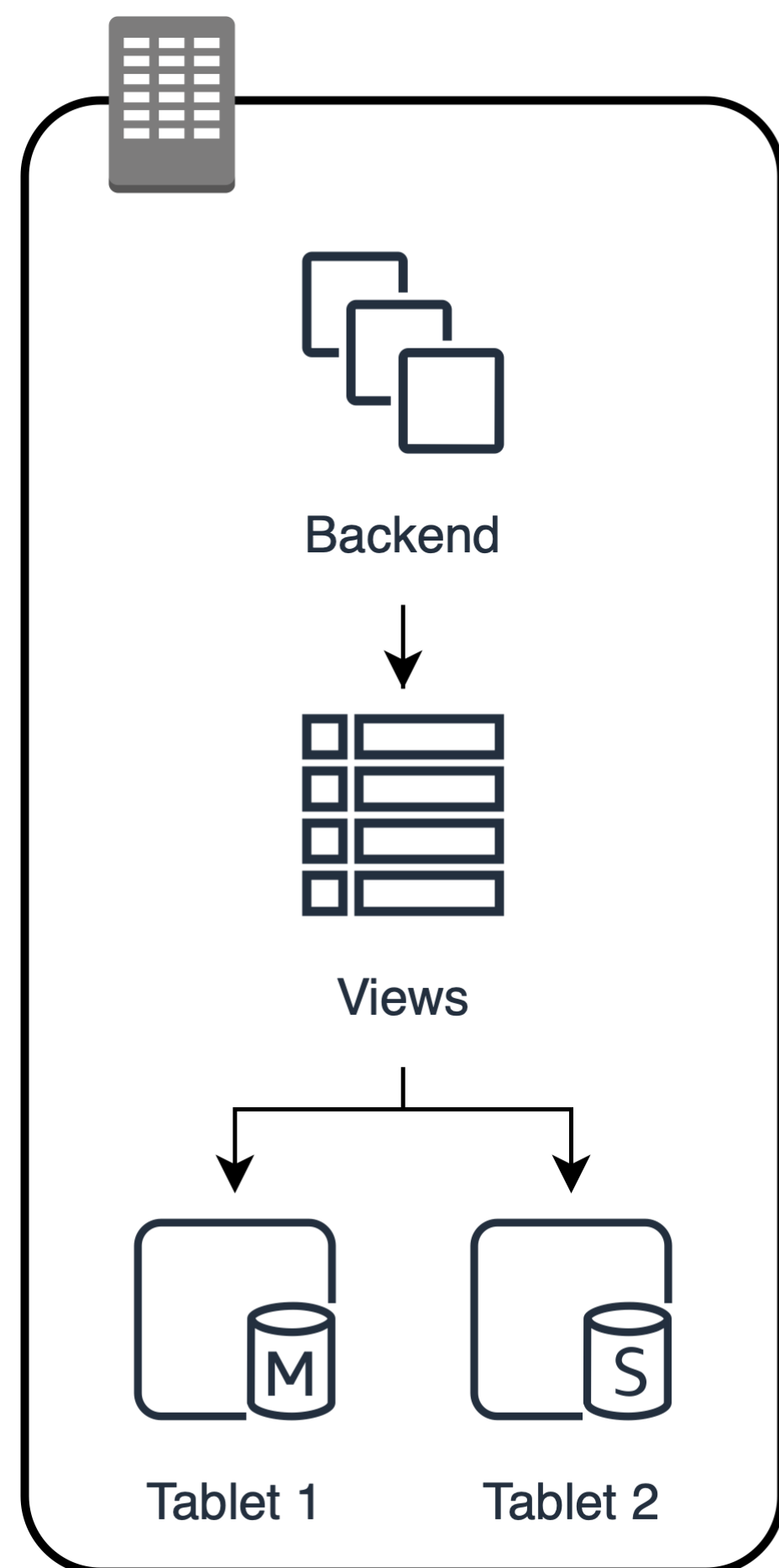


# Локальность данных



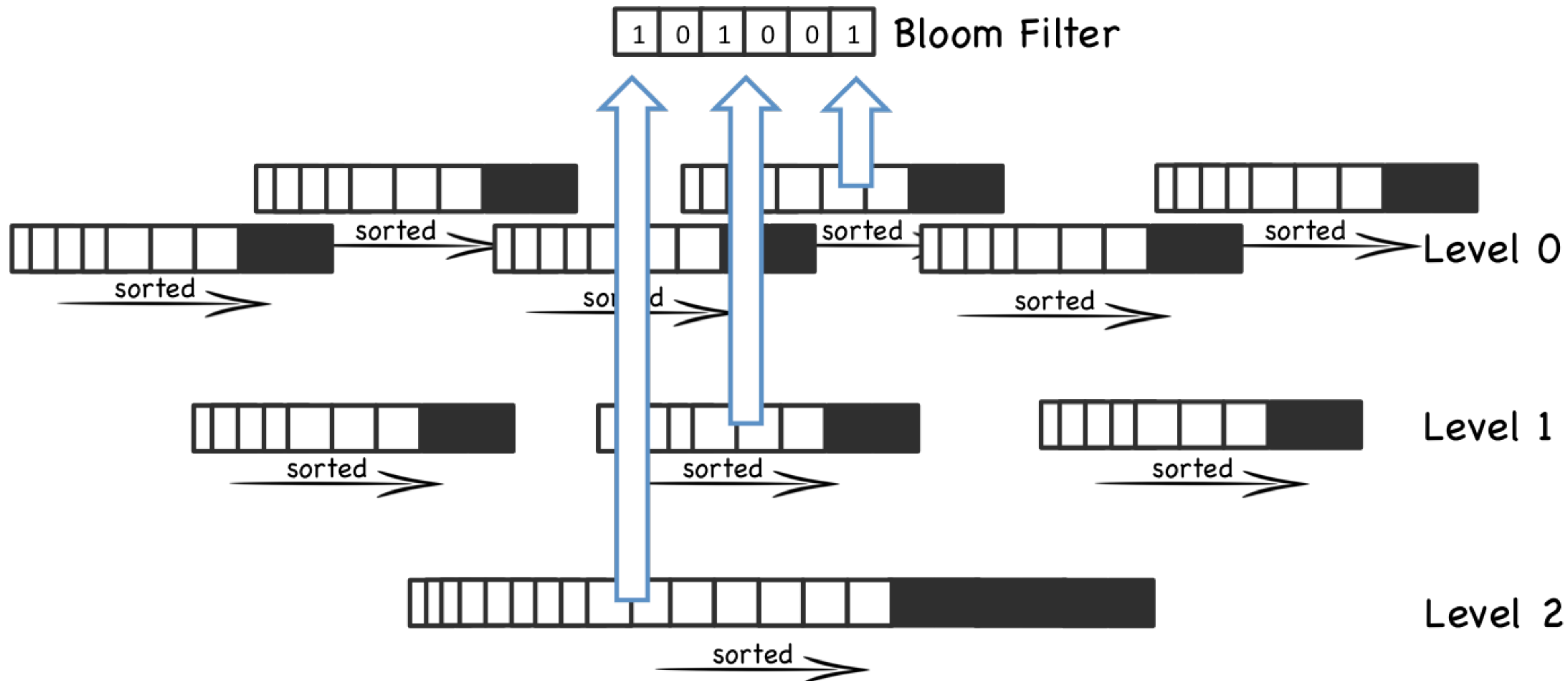


# Локальность данных





# Несуществующие ID



<http://www.benstopford.com/2015/02/14/log-structured-merge-trees/>





# Особенности реализации

+ Горизонтальное масштабирование на запись и чтение

+ Суммарное время на обновление счётчика – ~ 10 секунд

+ Время на чтение батч-запросов из сотен ключей – 10-12мс по p99





# 04

# Зрители



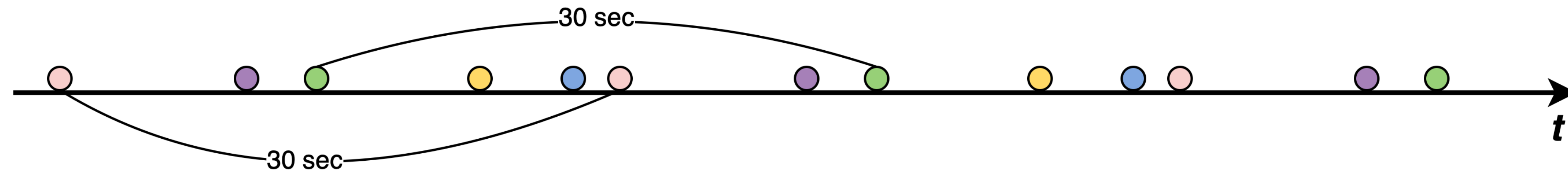


# Храним счётчик?

ID	Зрители
id1	653
id2	60
id3	2667
id4	14
id5	65



# 30-секундные хартбиты

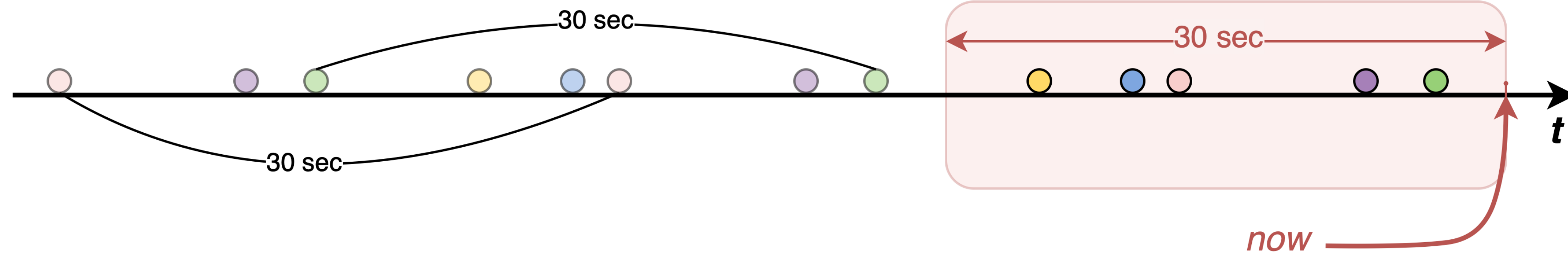


## Heartbeats

- user 3
- user 5
- user 7
- user 10
- user 18



# Скользящее окно



## Heartbeats

- user 3
- user 5
- user 7
- user 10
- user 18



# Храним события?

ID	Timestamp	Пользователь
id1	1645293584	● user 3
id2	1645293586	● user 10
id2	1645293592	● user 5
id1	1645293614	● user 3
id1	1645293584	● user 7





# Храним события?

ID	Timestamp	Пользователь
id1	1645293584	● user 3
id2	1645293586	● user 10
id2	1645293592	● user 5
id1	1645293614	● user 3
id1	1645293584	○ user 7



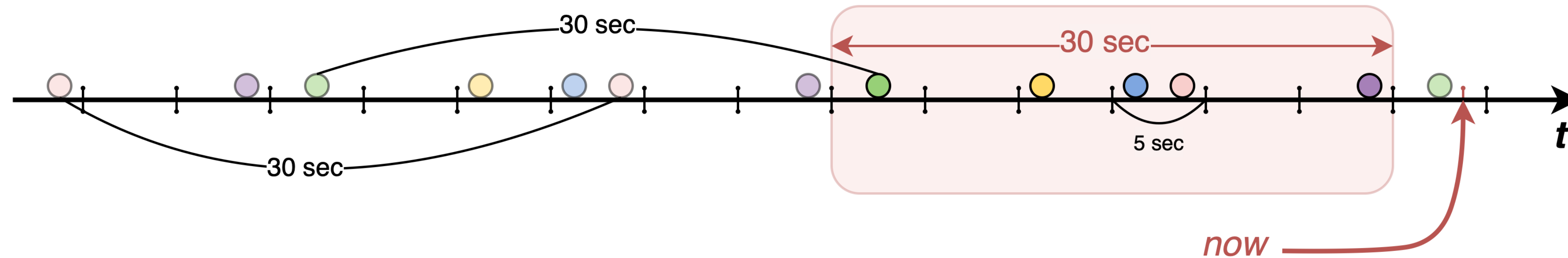


# Храним данные по 5-секундным бакетам

ID	Timestamp ↓ 5	Зрители за 5 сек
id1	1645293580	543
id2	1645293585	34
id2	1645293590	55
id1	1645293610	601
id1	1645293575	478



# Делаем скользящее окно прыгающим

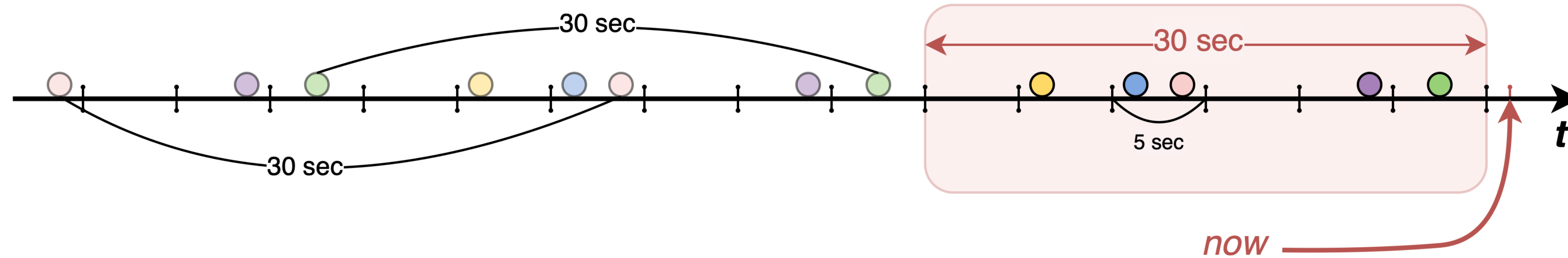


## Heartbeats

- user 3
- user 5
- user 7
- user 10
- user 18



# Делаем скользящее окно прыгающим



## Heartbeats

- user 3
- user 5
- user 7
- user 10
- user 18



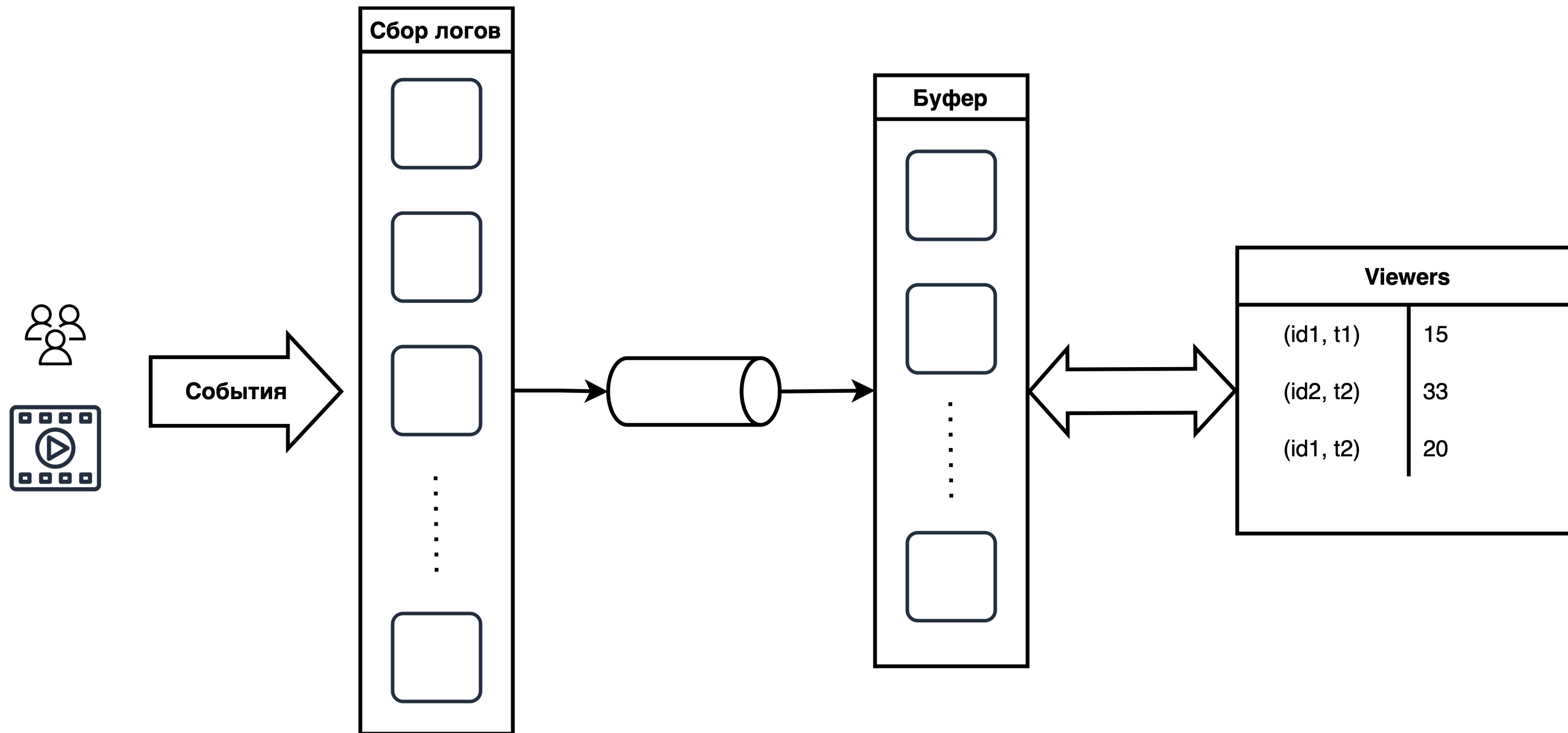
# Счётчик зрителей

+ По смыслу realtime-величина, но считаем её значение в дискретные моменты времени

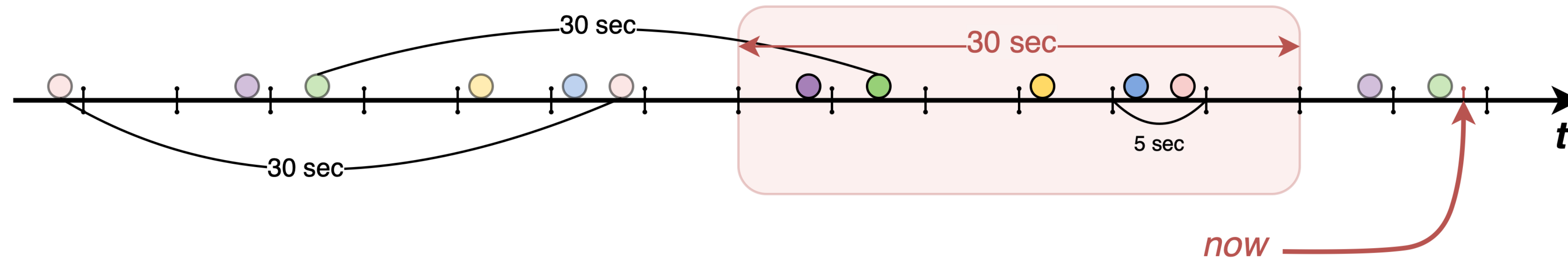
+ Телеметрия посылает события о начале воспроизведения и хартбит раз в 30 секунд



# Пайплайн



# Учитываем время буферизации



## Heartbeats

- user 3
- user 5
- user 7
- user 10
- user 18



# Чтение

$$viewers_{video\_id} = \sum_{i=0}^5 bucket_{video\_id}(t_{now} - t_{now} \bmod 5 - 5(i + 2))$$

$t_{now}$  – текущий unix timestamp



# Что-то забыли...

- Время чтения стало плавно деградировать с каждым днём
- Таблица быстро росла, несмотря на автоудаление данных по TTL



# Что-то забыли...

- Время чтения стало плавно деградировать с каждым днём
- Таблица быстро росла, несмотря на автоудаление данных по TTL
- Для прошедших трансляций тоже приходят хартбиты :)



# Итого

- 1** Если пользователи не приходят и не уходят с трансляции – счётчик стабилен
- 2** Новые пользователи учитываются в течение 5-10 секунд
- 3** Ушедшие с трансляции – вычитаются через 35-40 секунд





# Underflow 😊



Разбитые мечты !  
**BMW 1** Серия за 750  
ТЫСЯЧ

 Яндекс.... 12 февраля

18446744074 млрд просм...



ГОД НА **BMW 118i**\_  
МИНУСЫ, ПЛЮСЫ,  
САМЫЙ ЧЕСТНЫЙ ОТЗЫВ

 Яндекс.... 15 февраля

18446744074 млрд просм...





HighLoad++  
FOUNDATION

Яндекс

# Спасибо!

Иван Соколов

Группа рантайма видеохостинга

Дзен

[i-sokolov@yandex-team.ru](mailto:i-sokolov@yandex-team.ru)

Telegram: @ivsokolov

