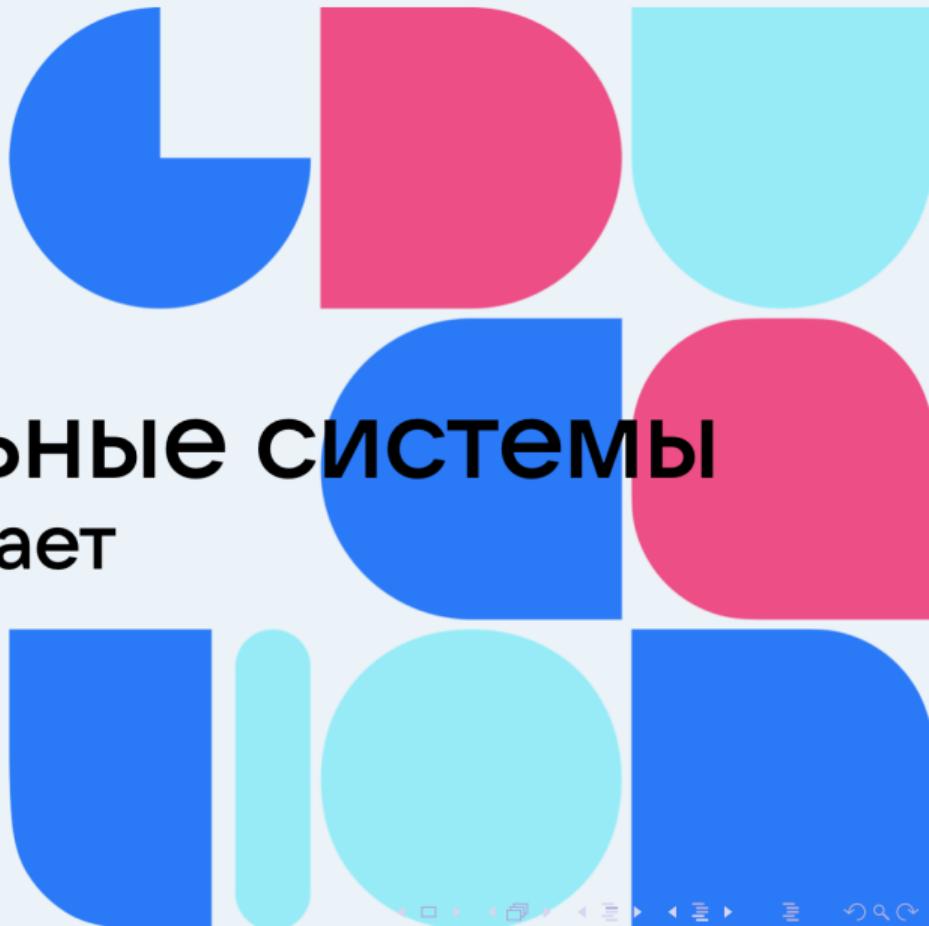


Рекомендательные системы и кто над ними работает

Николай Анохин

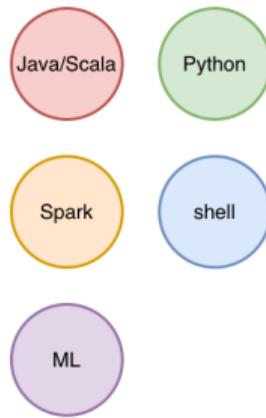


Обо мне

Академический опыт



Навыки

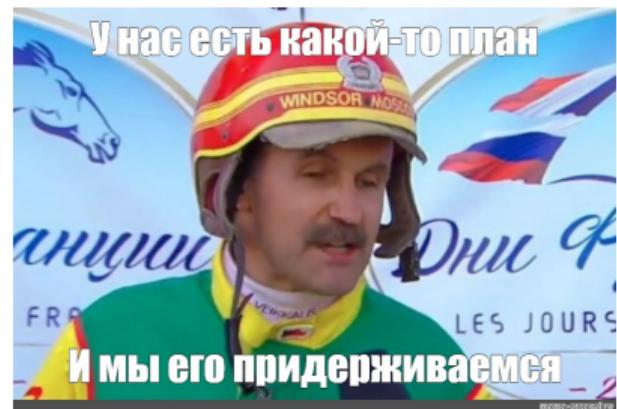


Индустриальный опыт



Из этой лекции вы узнаете

- Как устроена рекомендательная система (на примере рекомендаций музыки)
- Что нужно знать и уметь, чтобы ее создать (это сложно, но вам точно по силам)



Зачем нужны рекомендательные системы [RRSK10]



Кинопоиск



Зачем нужны рекомендательные системы [RRSK10]



- Бизнесу

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**
 - Увеличить продажи

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**

- Увеличить продажи
- Добиться большей лояльности

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**

- Увеличить продажи
- Добиться большей лояльности
- Улучшить пользовательский опыт

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**

- Увеличить продажи
- Добиться большей лояльности
- Улучшить пользовательский опыт

- **Пользователям**

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**

- Увеличить продажи
- Добиться большей лояльности
- Улучшить пользовательский опыт

- **Пользователям**

- Найти лучший товар

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**

- Увеличить продажи
- Добиться большей лояльности
- Улучшить пользовательский опыт

- **Пользователям**

- Найти лучший товар
- Найти **все** подходящие товары

Зачем нужны рекомендательные системы [RRSK10]



- **Бизнесу**

- Увеличить продажи
- Добиться большей лояльности
- Улучшить пользовательский опыт

- **Пользователям**

- Найти лучший товар
- Найти **все** подходящие товары
- Залипнуть

THE TINDER SWINDLER

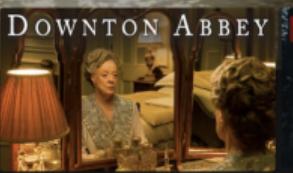
Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

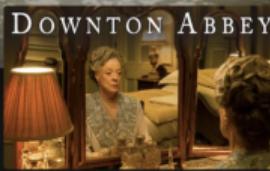
Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



Trending Now



THE TINDER SWINDLER

1

Play

More Info



16+

TV Shows



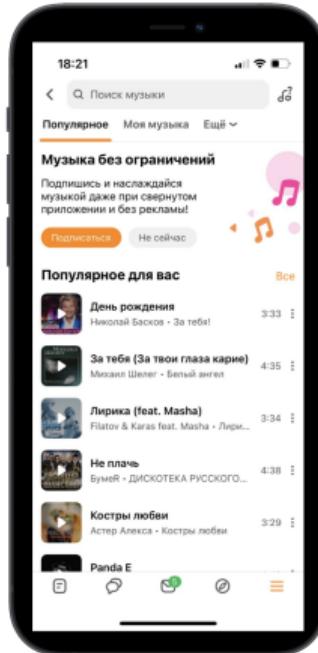
Trending Now



Рекомендации музыки

Задача

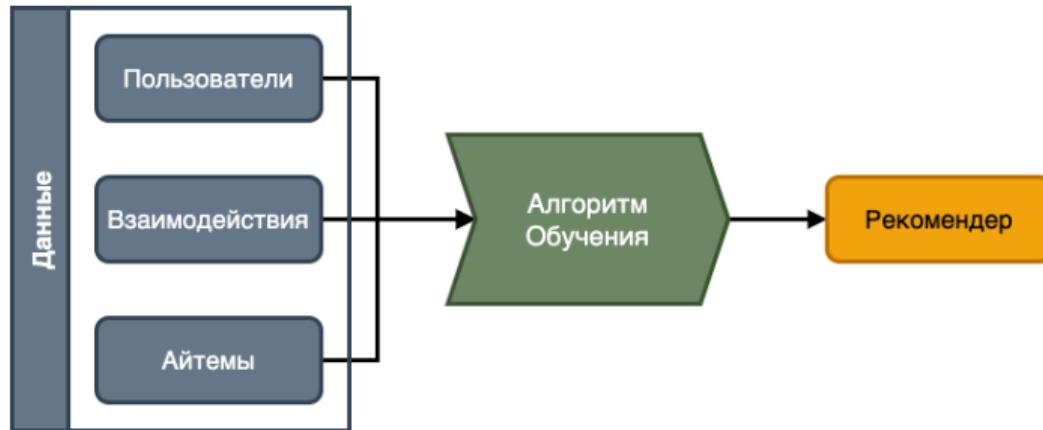
Рекомендовать пользователю **исполнителей**, так чтобы пользователь слушал их песни и **как можно дольше** оставался в приложении



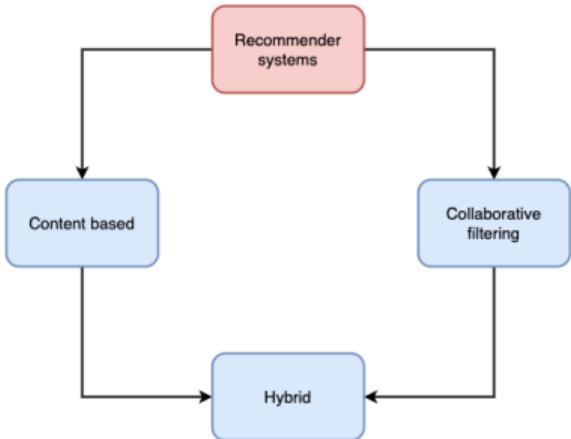
Алгоритм рекомендаций



Как строить рекомендации

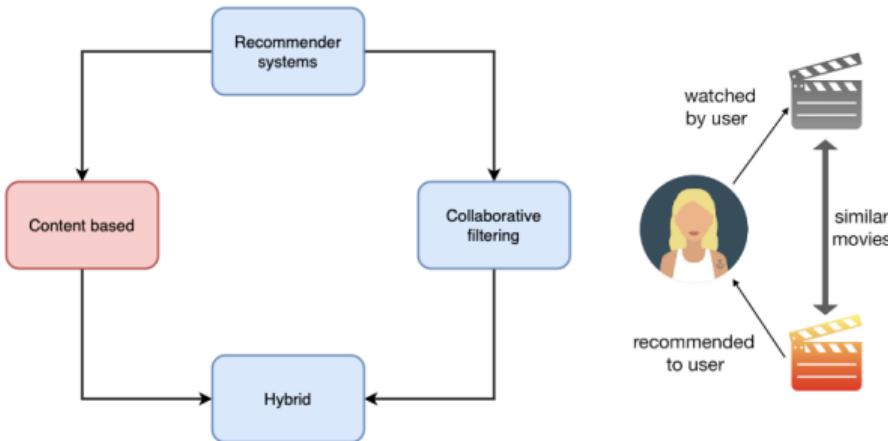


Зоопарк алгоритмов рекомендаций [Ali21]



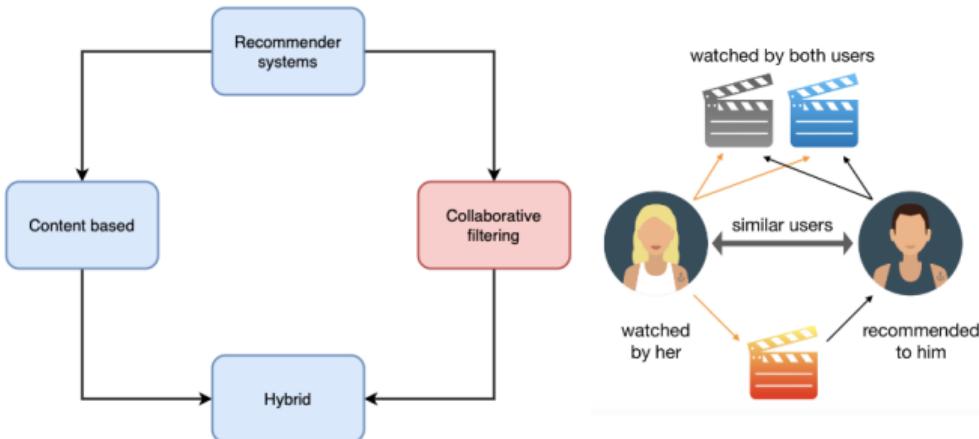
<https://shorturl.at/h5Pp5>

Зоопарк алгоритмов рекомендаций [Ali21]



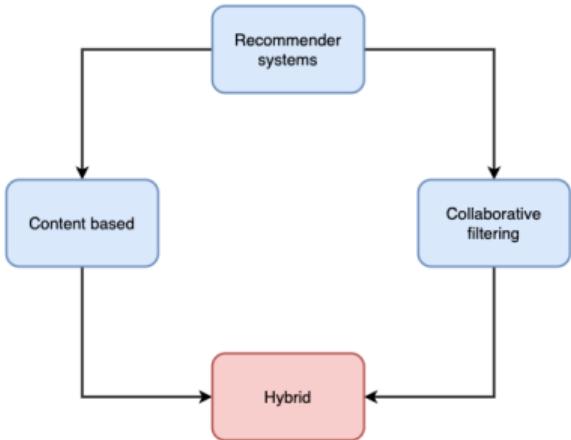
<https://shorturl.at/h5Pp5>

Зоопарк алгоритмов рекомендаций [Ali21]



<https://shorturl.at/h5Pp5>

Зоопарк алгоритмов рекомендаций [Ali21]



<https://shorturl.at/h5Pp5>

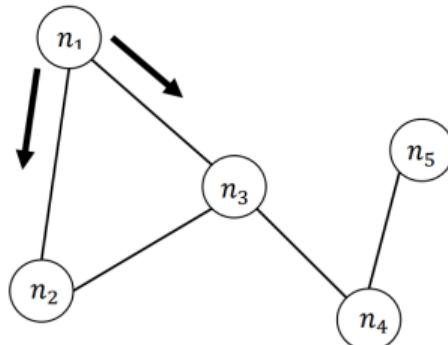
**ИЗВОЛЬТЕ
СКОМПИЛИРОВАТЬ**

ВАШ КОД СУДАРЬ

Personalized Page Rank (PPR) [PLCL19]

Начальное состояние

$$\vec{s}_0 = (1 \quad 0 \quad 0 \quad 0 \quad 0)$$



Вероятность перехода

$$\vec{p}_{n+1} = c \vec{s}_n \cdot P + (1 - c) \vec{s}_0$$

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$(1 - c)$ – вероятность рестарта

Где мы окажемся через бесконечное число шагов?

Строим граф исполнителей: создаем ребра I

Коля: Лепс, Газманов, Adele

Строим граф исполнителей: создаем ребра I

Коля: Лепс, Газманов, Adele

Задача

Есть K пользователей. Пользователь j случайным образом с равной вероятностью без возвращения выбирает n_j из N исполнителей. Сколько в среднем пользователей выберут каждую пару исполнителей?

Строим граф исполнителей: создаем ребра I

Коля: Лепс, Газманов, Adele

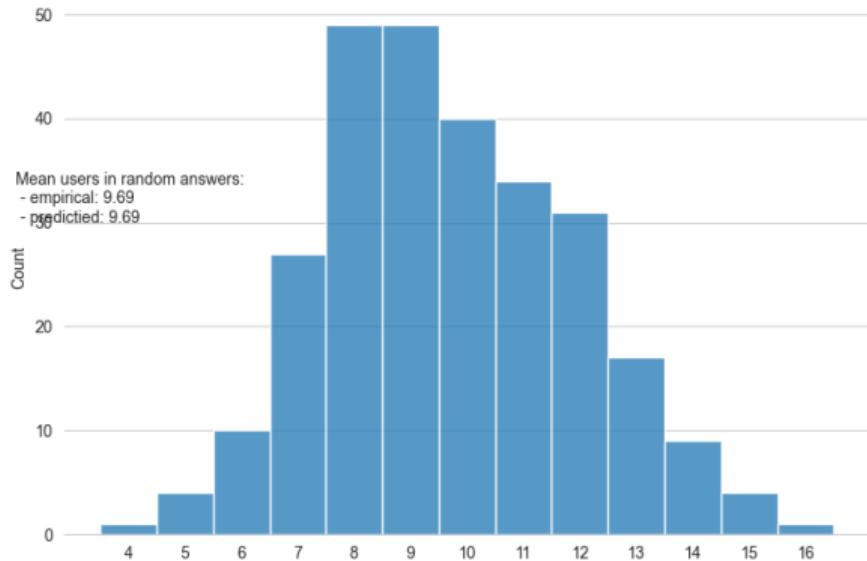
Задача

Есть K пользователей. Пользователь j случайным образом с равной вероятностью без возвращения выбирает n_j из N исполнителей. Сколько в среднем пользователей выберут каждую пару исполнителей?

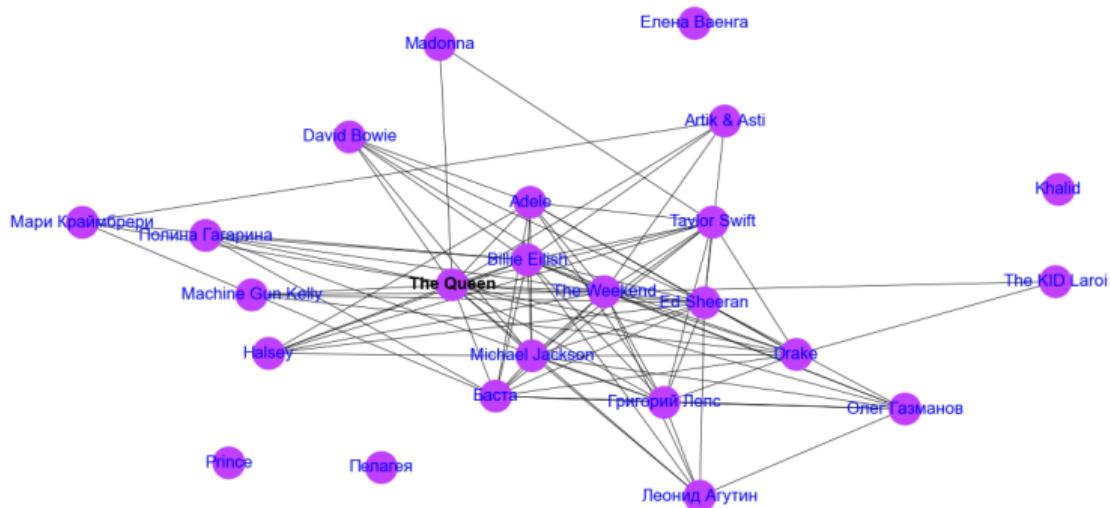
Ответ

$$E_1 + \dots + E_K = \frac{\sum_{j=1}^K n_j(n_j - 1)}{N(N - 1)}$$

Строим граф исполнителей: создаем ребра II



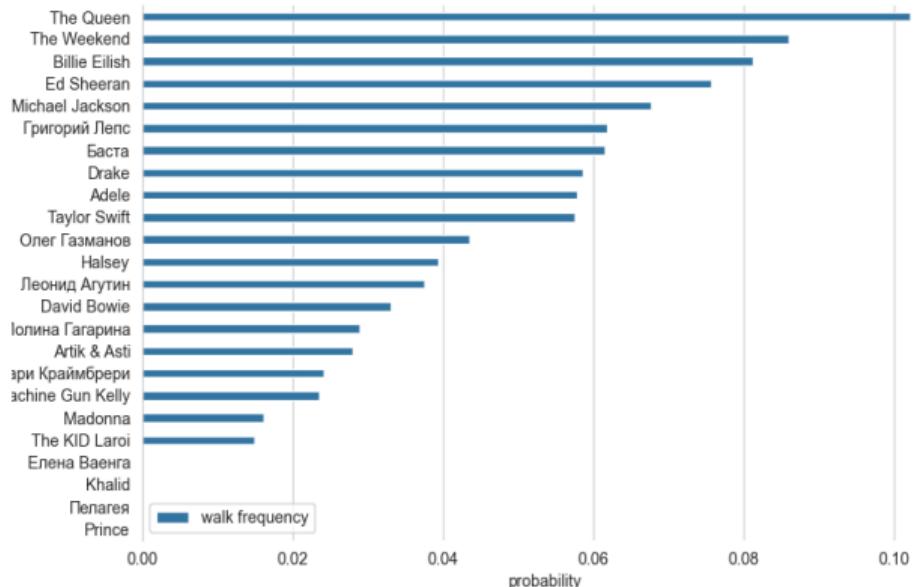
Строим граф исполнителей III



Случайные блуждания по графу исполнителей

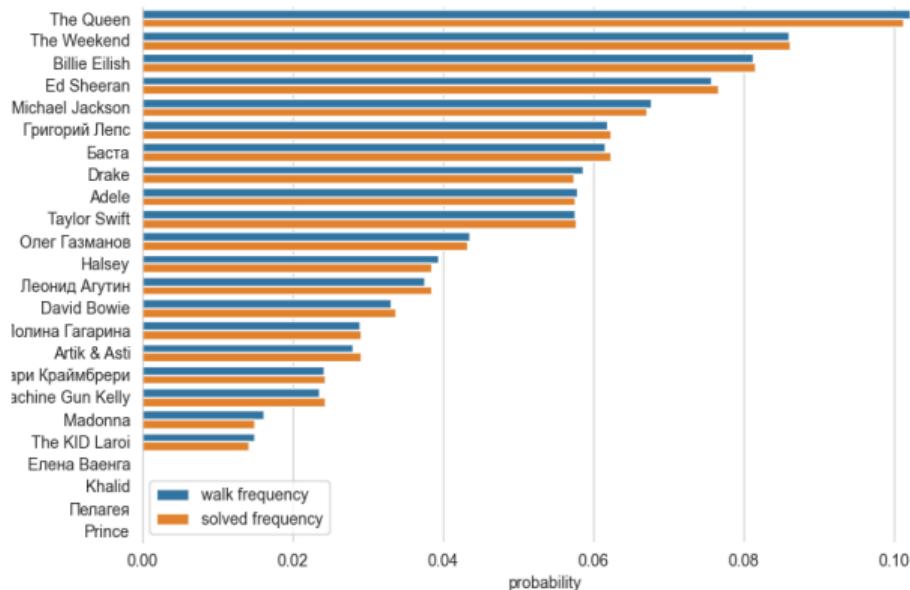
[Ссылка на визуализацию](#)

Рекомендации на основе блужданий



Рекомендации на основе точного решения

$$\vec{r} = c\vec{P} + (1 - c)\vec{s} \rightarrow \vec{r} = (1 - c)\vec{s}(I - cP)^{-1}$$



ML Инженер



Продакшен



Сервис рекомендаций

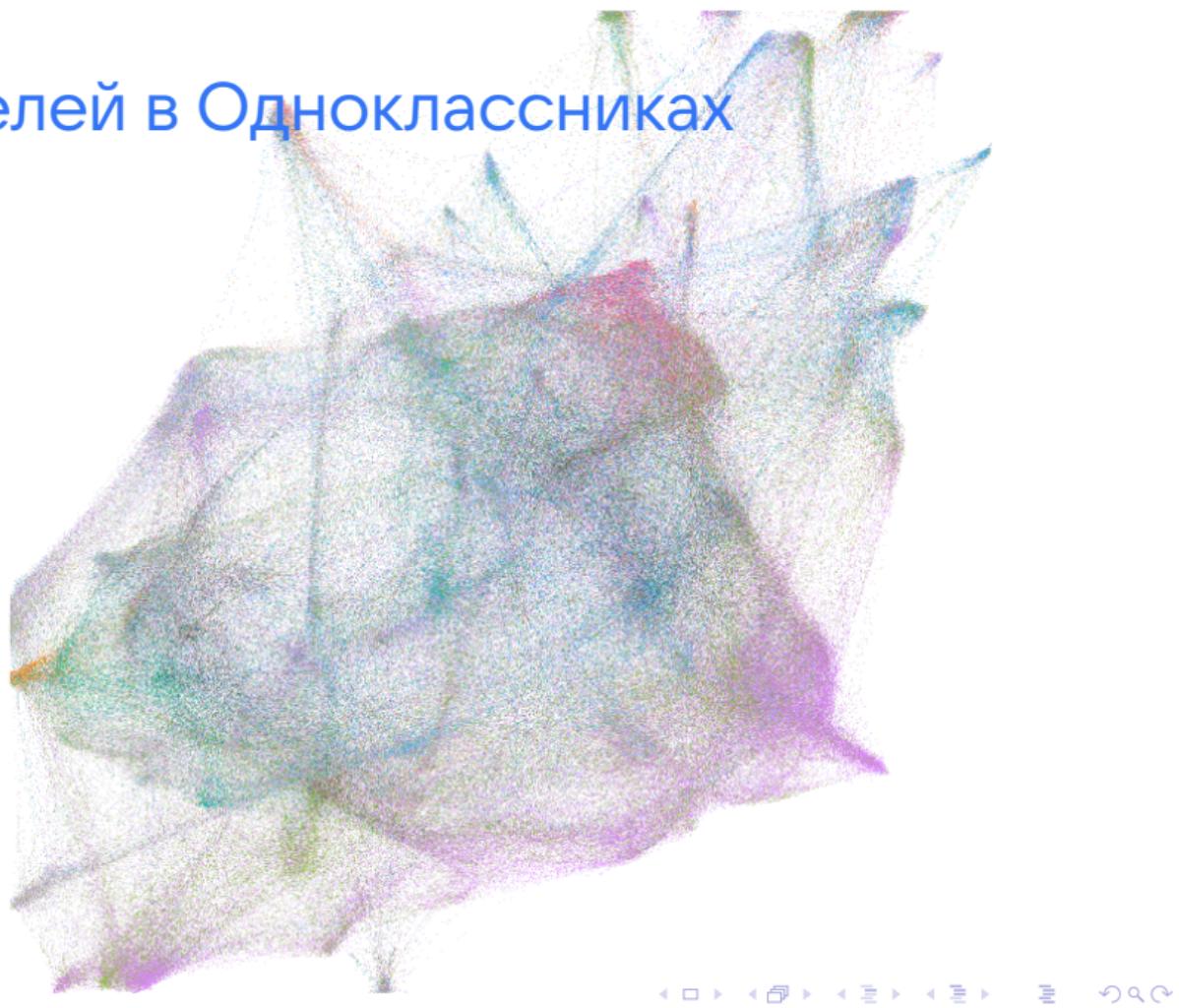


Граф исполнителей в Одноклассниках

Параметры графа

20000 вершин

750000 ребер



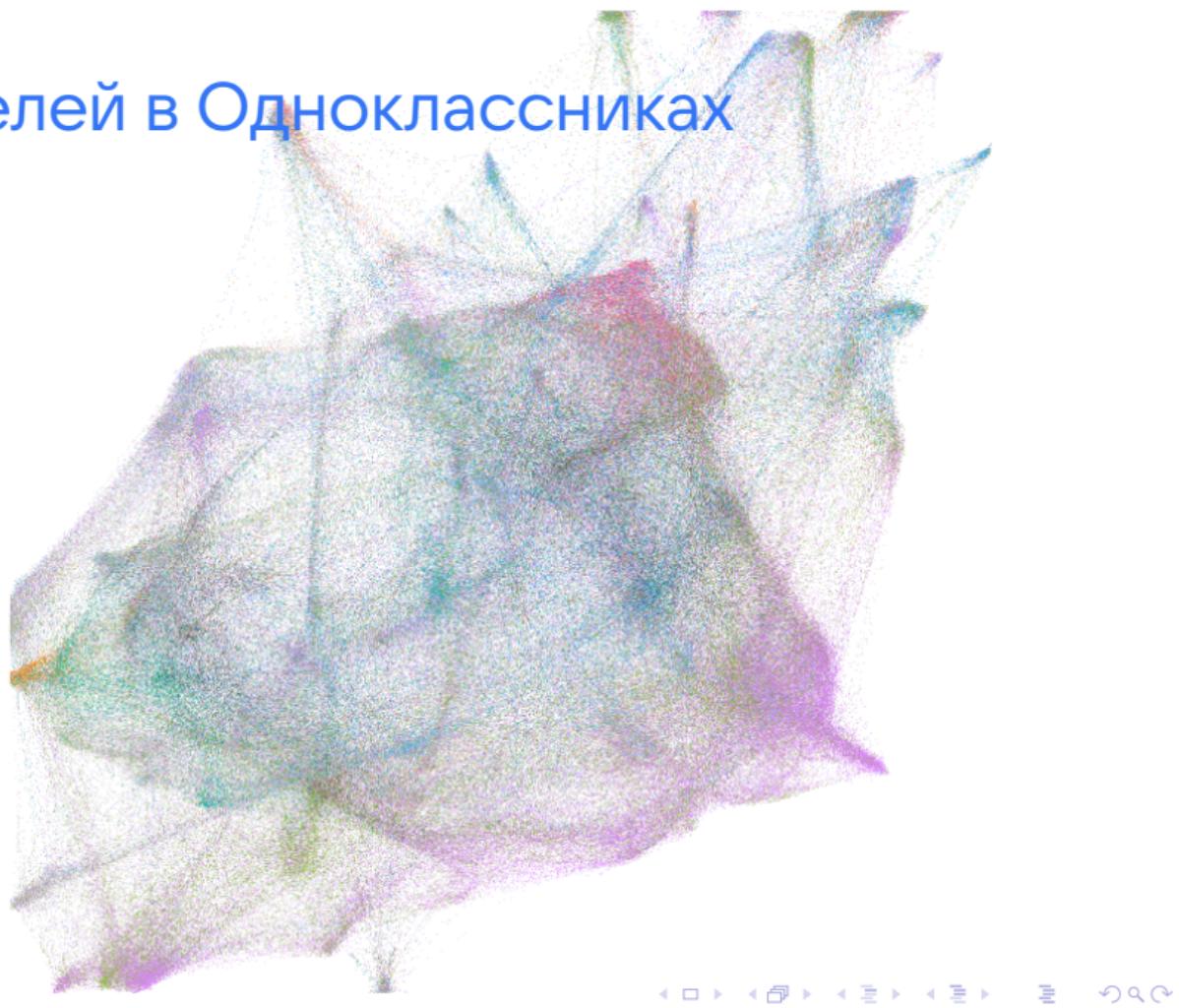
Граф исполнителей в Одноклассниках

Параметры графа

20000 вершин

750000 ребер

Память



Граф исполнителей в Одноклассниках

Параметры графа

20000 вершин

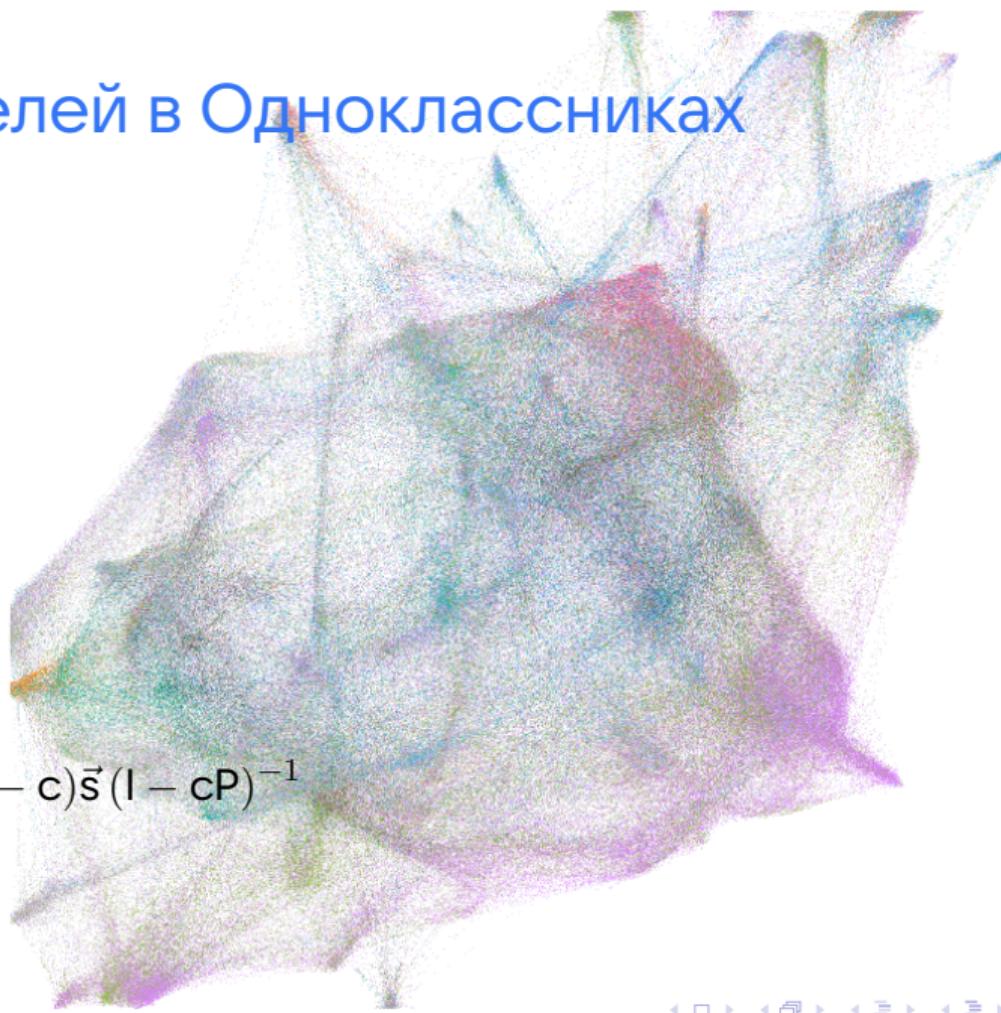
750000 ребер

Память

$O(N_{\text{nodes}}^2)$

Сложность

Прямое решение $\vec{r} = (1 - c)\vec{s} (I - cP)^{-1}$



Граф исполнителей в Одноклассниках

Параметры графа

20000 вершин

750000 ребер

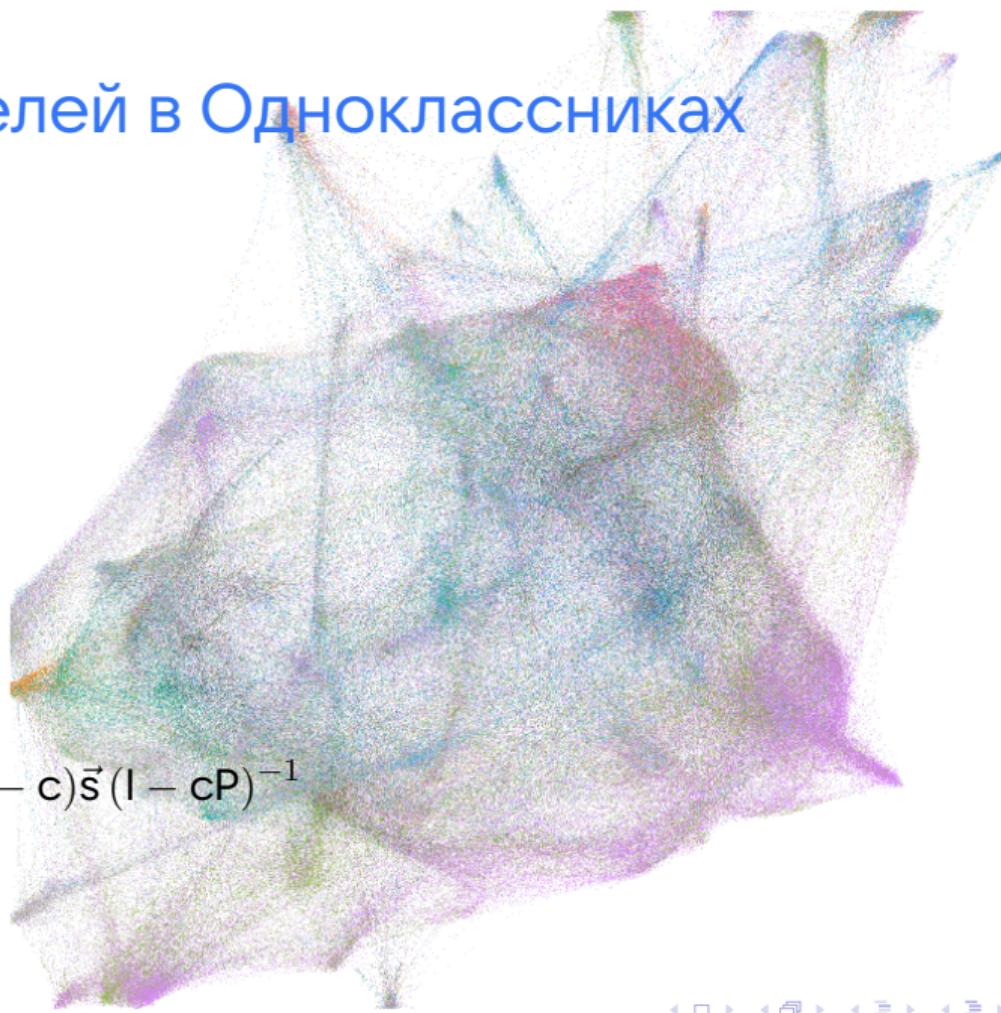
Память

$O(N_{\text{nodes}}^2)$

Сложность

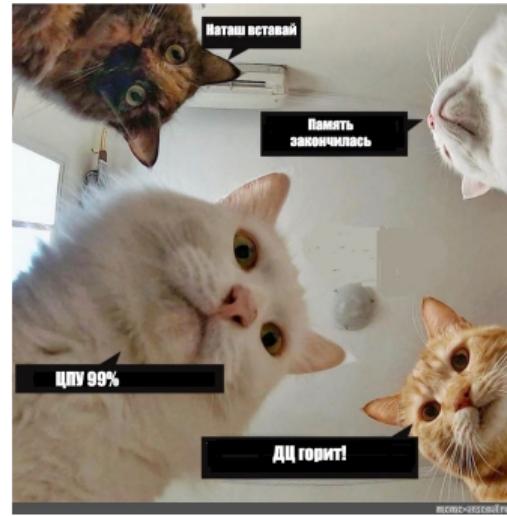
Прямое решение $\vec{r} = (1 - c)\vec{s} (I - cP)^{-1}$

$O(N_{\text{nodes}}^3)$



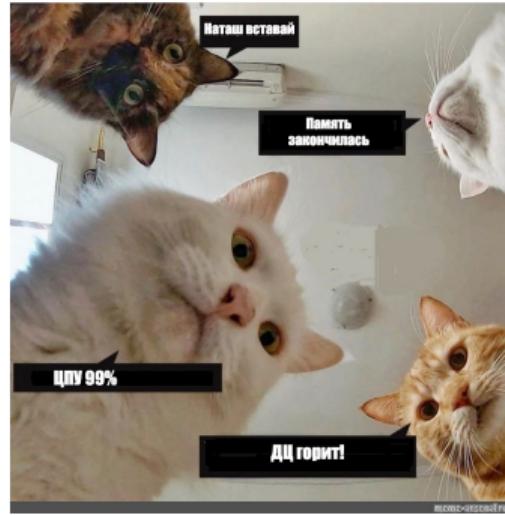
Оптимизируем вычисление рекомендаций [PLCL19]

- Сэмплирование монте-карло



Оптимизируем вычисление рекомендаций [PLCL19]

- Сэмплирование монте-карло
- Оптимизированное прямое решение



Оптимизируем вычисление рекомендаций [PLCL19]

- Сэмплирование монте-карло
- Оптимизированное прямое решение
- Оптимизированный power iteration



Backend Разработчик



Данные для рекомендаций

• • • • •

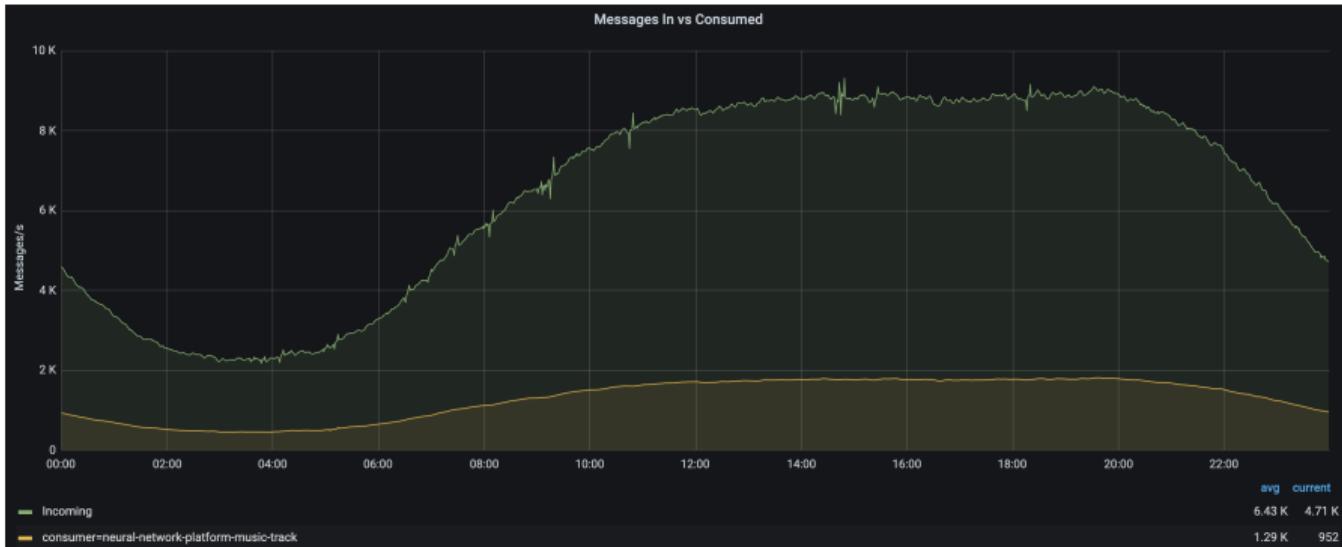


Транспорт данных между сервисами



- Хранение данных для анализа

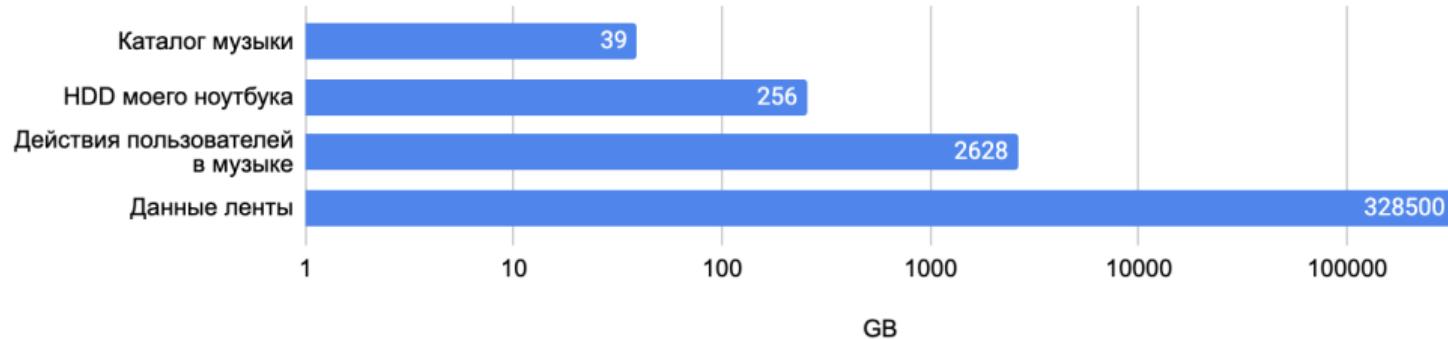
Транспорт данных между сервисами



- Хранение данных для анализа
- Потоковая обработка данных (Spark Streaming, Apache Samza)

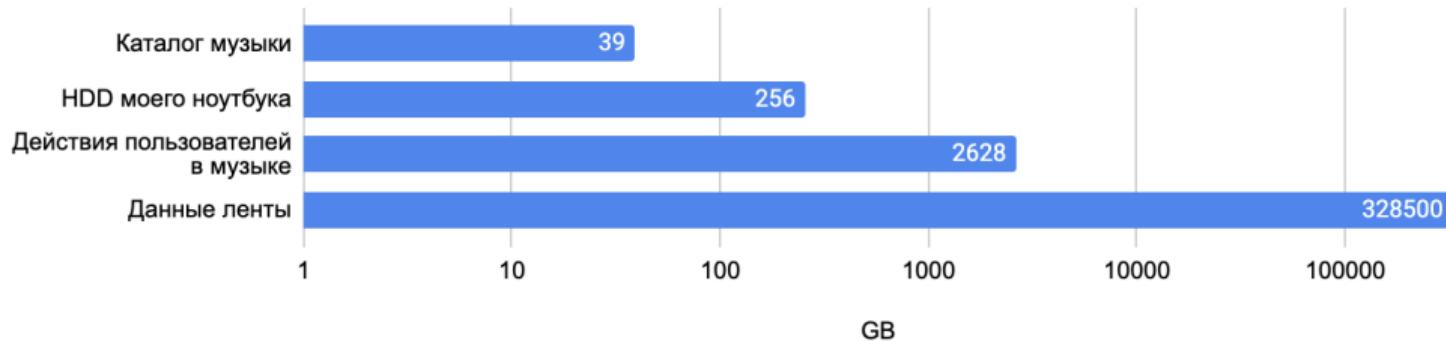
Хранение и анализ данных

Размер с репликацией



Хранение и анализ данных

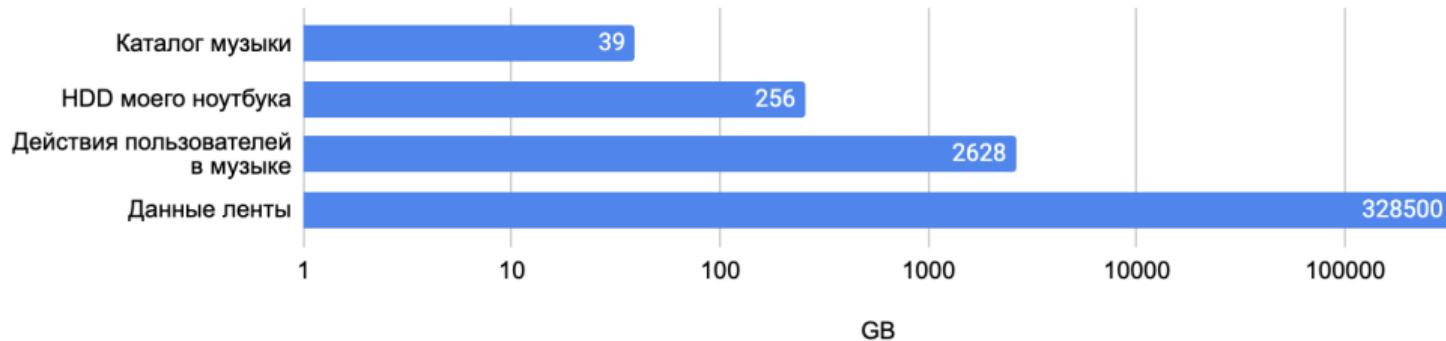
Размер с репликацией



- Распределенная файловая система: Hadoop HDFS

Хранение и анализ данных

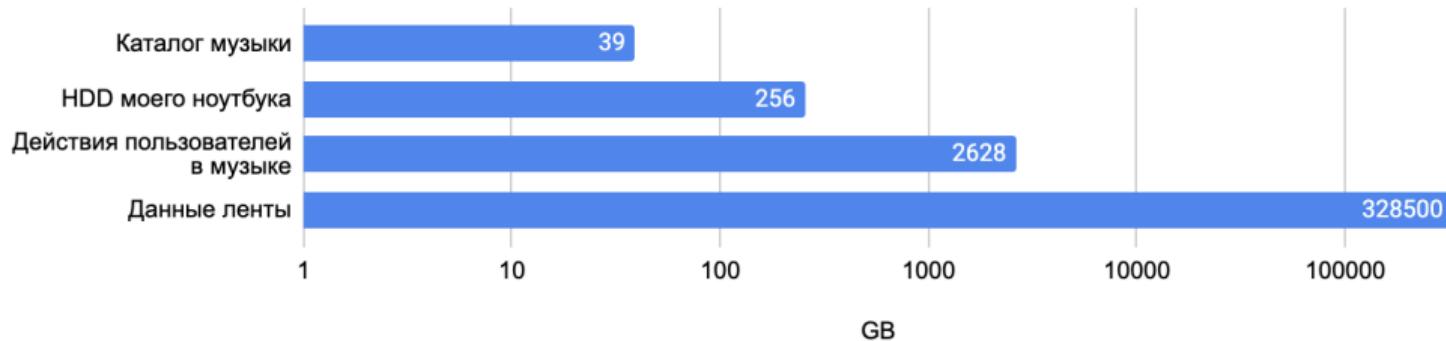
Размер с репликацией



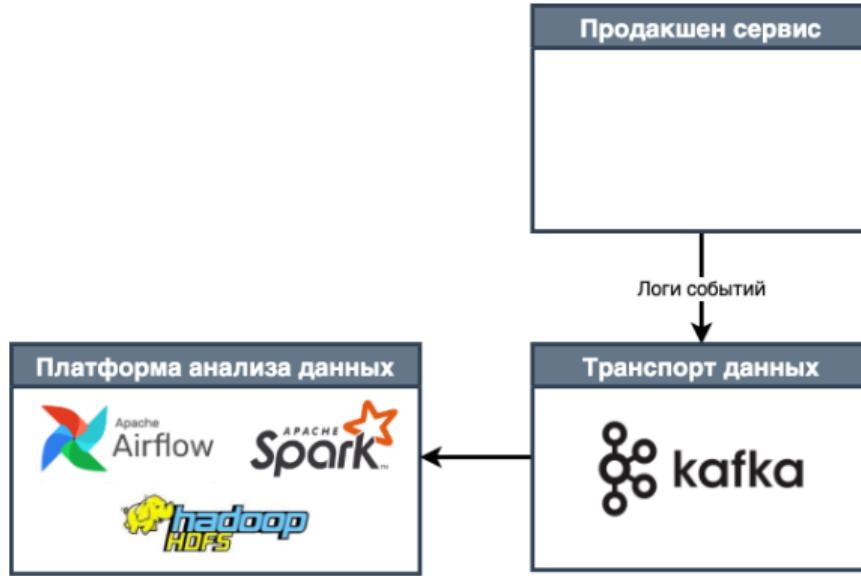
- Распределенная файловая система: Hadoop HDFS
- Анализ данных: Apache Spark

Хранение и анализ данных

Размер с репликацией



- Распределенная файловая система: Hadoop HDFS
- Анализ данных: Apache Spark
- Контроль выполнения: Apache Airflow



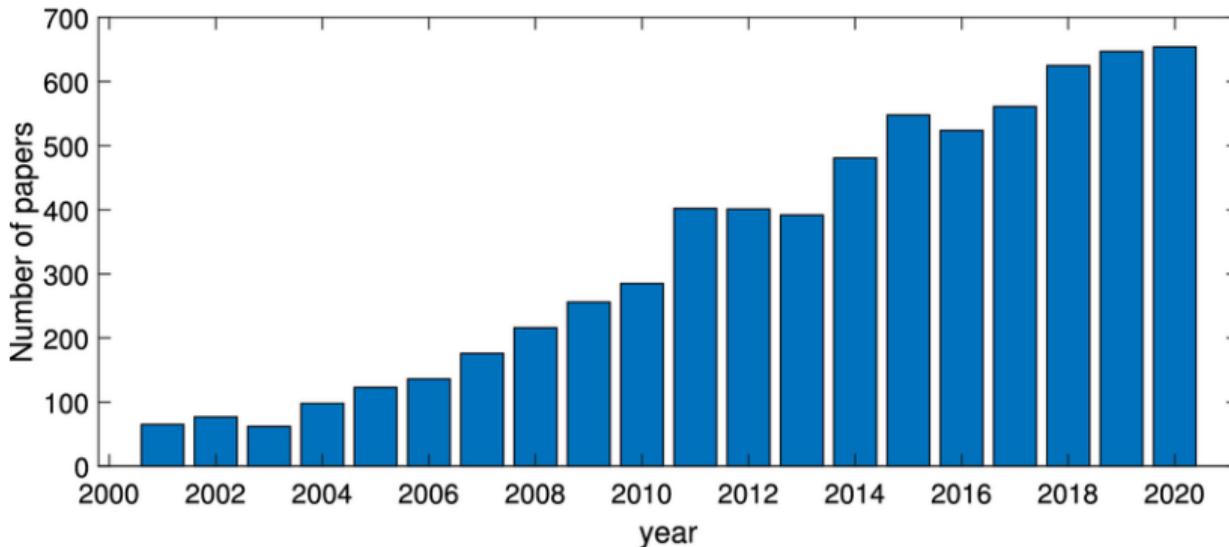
Big Data Инженер



Наука о рекомендациях



Количество статей с recommender system" в названии



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря
 - Разнообразие рекомендаций



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря
 - Разнообразие рекомендаций
 - Reinforcement learning



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря
 - Разнообразие рекомендаций
 - Reinforcement learning
- Пользовательский опыт



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря
 - Разнообразие рекомендаций
 - Reinforcement learning
- Пользовательский опыт
 - Объяснения рекомендаций



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря
 - Разнообразие рекомендаций
 - Reinforcement learning
- Пользовательский опыт
 - Объяснения рекомендаций
 - Настройки рекомендаций



Актуальные задачи в науке о рекомендациях

- Новые алгоритмы
- Счастье пользователей
 - Долгосрочный эффект
 - Многосторонние маркетплейсы
 - Честность рекомендаций
- Выйти из пузыря
 - Разнообразие рекомендаций
 - Reinforcement learning
- Пользовательский опыт
 - Объяснения рекомендаций
 - Настройки рекомендаций
 - Оптимизация UI



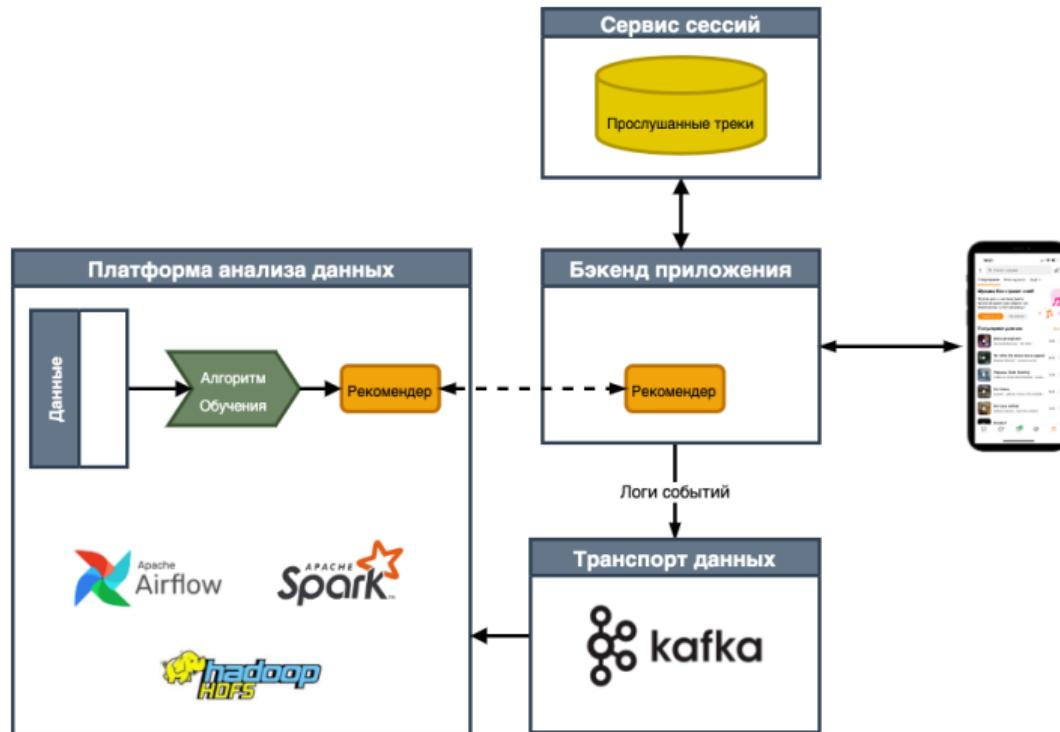
Учёный



Итоги



Готовая рекомендательная система



Разработчики приложений

Менеджер проектов

Тестировщик

ML инженер

Фронтенд разработчик

Бэкенд разработчики

Big Data
инженер

Директор по
проектам

Ученых у
нас нет :(



Становитесь частью IT вместе с VK Education



Портал

Пройти обучение на курсах,
получить знания и опыт



Сообщество

Участвуйте в проектах
и стройте карьеру в IT

Спасибо за внимание!

Николай Анохин

t.me/anokhinn



Лекция



t.me/mlvok



Литература I

-  Iman Ali, Ai powered search and recommendation system, 2021.
-  Sungchan Park, Wonseok Lee, Byeongseo Choe, and Sang-goo Lee, A survey on personalized pagerank computation algorithms, IEEE Access **PP** (2019), 1–1.
-  Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, Recommender systems handbook, 1st ed., Springer-Verlag, Berlin, Heidelberg, 2010.