

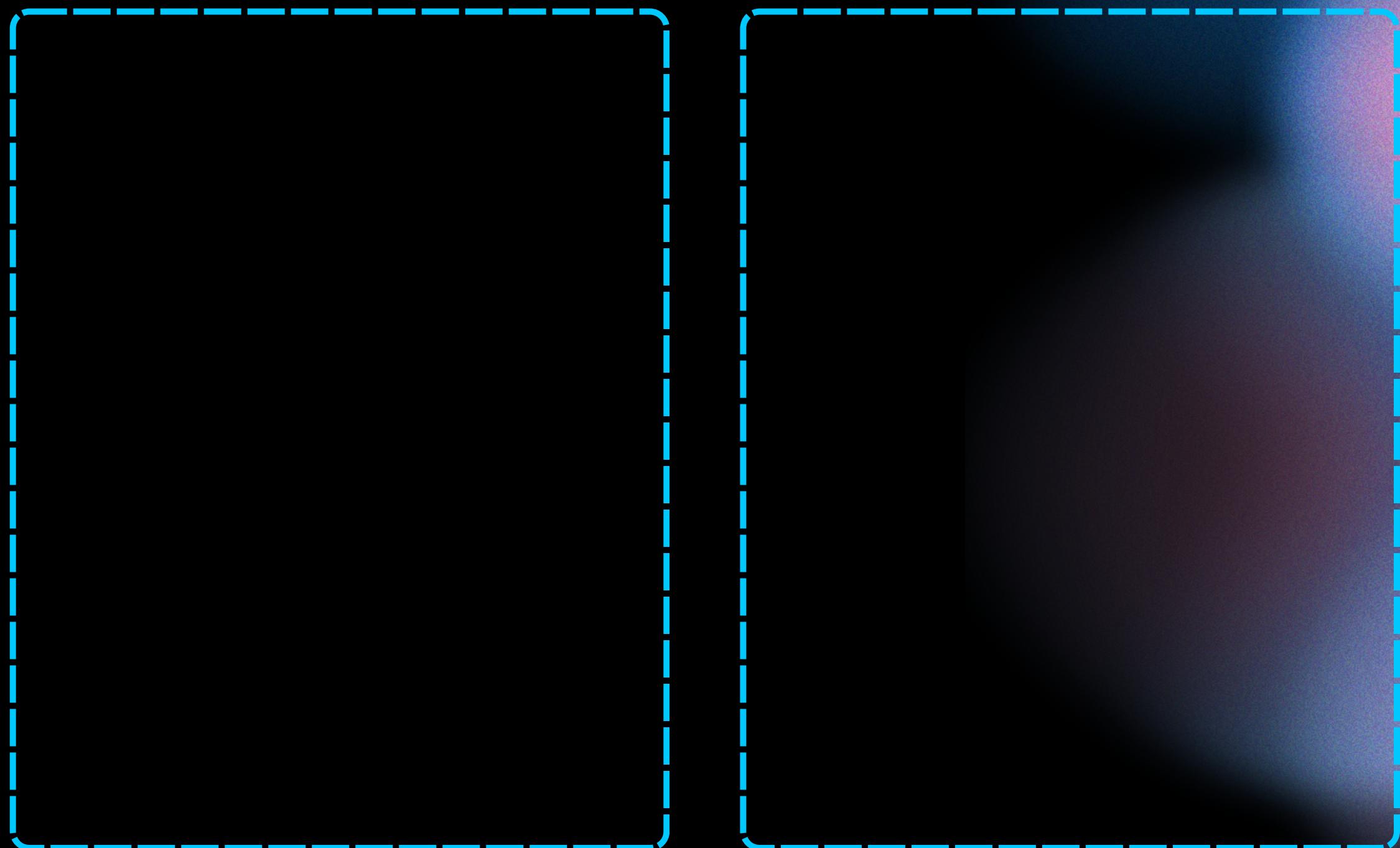
IT INNO HACK

LinkHouse

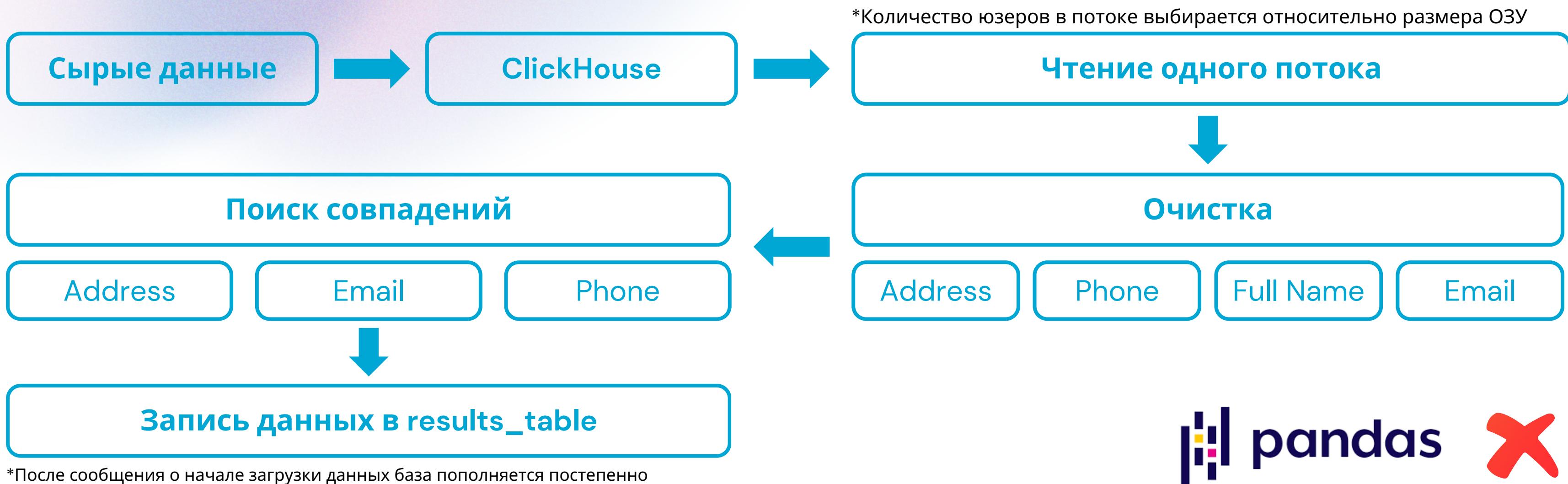
Команда БЕЗУМЦЫ



LinkHouse



Match Pipeline



Statistics

LinkHouse в цифрах

95%

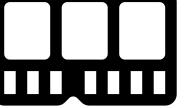
Среднее сходство по
критерию Джаро-Винклера
для имен после матчей

>20 млн

Матчей пользователей для
публичной части датасета

~30 тыс. it/s

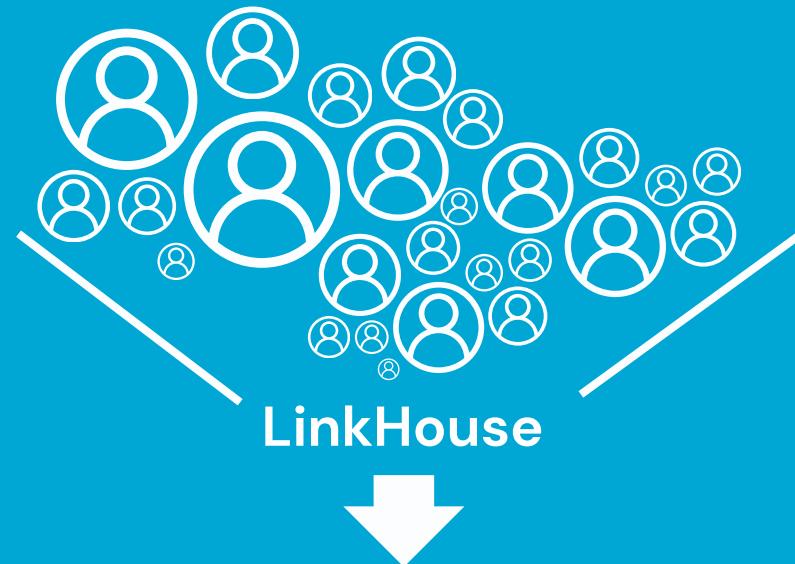
Скорость потоковой выгрузки
данных в ClickHouse

 **4 ГБ**  2 мин
озу

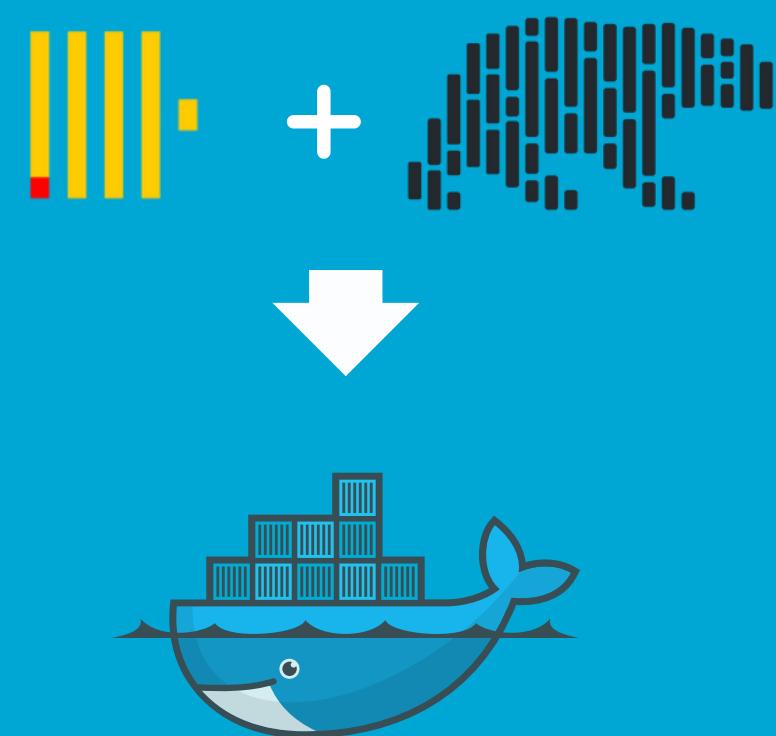
Нужно для поиска дубликатов
среди 10 млн пользователей

LinkHouse

Алгоритмы матчинга,
адаптированные под большие
данные “из коробки”



Docker-решение,
оптимизированно
подстраивающееся под ОЗУ



Docker ⚓ / Worker

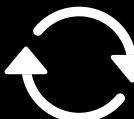
Запуск решения “из коробки”
одной командой

без лишних действий!

```
bezumec@hack:~ $ docker compose up
```



Возможность подключиться к
внешнему ClickHouse



Потоковая загрузка результата

Умная очистка после каждого этапа
матча, не храним таблицы целиком



Возможность быстрой обработки
крупных датасетов

сэмплинг

подстраивание под размеры ОЗУ

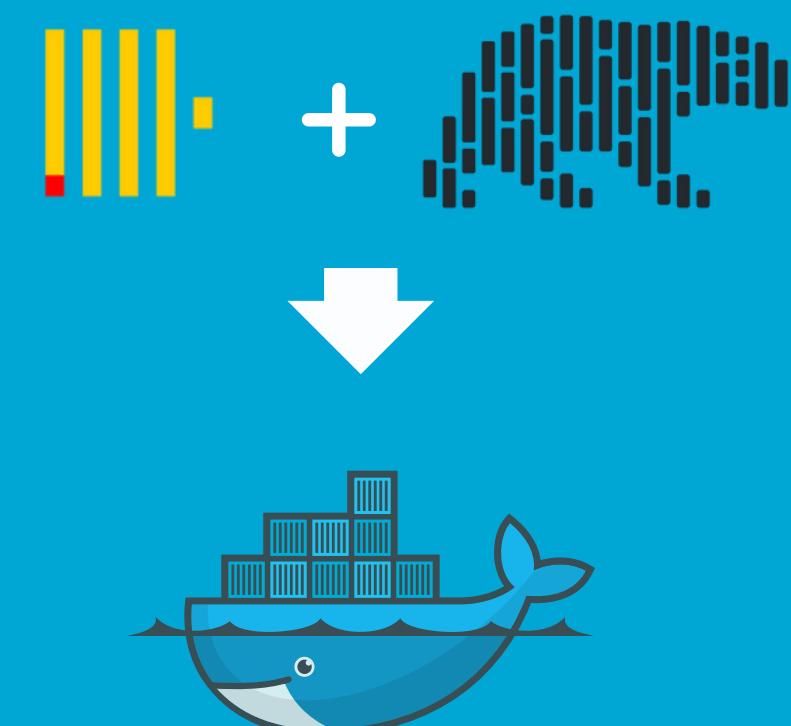
> 40 млн
пользователей

LinkHouse

Алгоритмы матчинга,
адаптированные под большие
данные “из коробки”



Docker-решение,
оптимизированно
подстраивающееся под ОЗУ



Интерфейс,
позволяющий удобно
просматривать результаты

Управление базами данных и поиск по
профилям клиентов

Таблица 1
table_dataset1
8639816 записей
2024-09-28 01:08:50
Таблица 2
table_dataset2
9366907 записей
2024-09-28 01:08:50
Таблица 3
table_dataset3
10134448 записей
2024-09-28 01:08:50

LinkHouse

Единый профиль клиента через устранение дубликатов

Найдется все...

Может занять некоторое время.

Создано для демонстрации. Более точный результат при запуске скриптов локально.

Добавить базу

Web Panel

03 ОБЪЕДИНЕНИЕ ПРОФИЛЕЙ:

Создать единый профиль клиента, объединяя данные из разных источников.



Обеспечить минимизацию потерь информации и разрешение конфликтующих данных.

ДЕЙНТЬЕР БАТИР КУРБОНАЛИЕВИЧ

Найдется все...



Колонки	table_dataset1	table_dataset2
uid	bf8adc63-532c-476c-9bfd-db8f58beae35	a4b7f703-0a8d-46eb-ad47-c859aa22b443
full_name	ДЕЙНТЬЕР БАТИР КУРБОНАЛИЕВИЧ	ДЕЙНТЬЕР БАТТР КУРБОНАЛИЕВ Ч
email	batir_dejnter5683836@example.ru	
address	клх Печенга, алл. Короткая, Дом 19 Квартира 7/8, 948139	клх Печенга, алл. Короткая, д. 19 к. 7/8, 948139
sex	m	
birthdate	1991-11-09	1991-11-09
phone	8 614 269 93 49	8 614 269 93 49
first_name		БАТТР
middle_name		КУРБОНАЛИЕВ Ч
last_name		ДЕЙНТЬЕР

Добавить базу

Поиск профилей по ФИО, номеру и email

Управление базами данных

Объединение профилей на одной странице

Колонки

uid

181e2bcf-c60c-44ac-9ee3bc

full_name

ТОШБОЛТАНГАЛЖАМРЕЛИУ

email

aleksej_toshbolta@mail.ru

address

к. Домодедово, ул. Мира, д. 4

sex

м

birthdate

2005-06-11



FastAPI

ПОТЕНЦИАЛ



Алгоритмы интегрированы в fastAPI, что делает систему более масштабируемой



FastAPI

