

«Маскировщик»

Маскируем данные без потери смысла



Содержание

H F Labs

О компании	03
Зачем нужно маскирование?	05
Типовые подходы к обезличиванию	07
Как работает «Маскировщик»?	10
Типы данных, работа с GPT и не только	13
Маскирование согласий клиентов	19
Архитектура продукта	22

H F Labs

Лидер российского рынка обработки клиентских данных

Даем целостный взгляд на все составляющие бизнеса, в том числе на источники данных, авторство, качество, полноту и на потенциальное использование данных

2005

год основания

150

человек в штате



Полностью российская компания



Продукты входят в Реестр отечественного ПО

10

собственных
IT-решений

90+

проектов по управлению
качеством клиентских
данных

Раньше только стандартизировали, теперь еще и маскируем!

H F Labs

Под капотом
те же уникальные алгоритмы

Знаем специфику данных и изменяем их,
сохраняя качество и смысл

 ВТБ

 СБЕР

 Т БАНК

 Альфа Банк

 СОГАЗ

 РОСГОССТРАХ

 ИНГОССТРАХ

 Ростелеком

 OZON

 МЕГАФОН

Зачем нужно маскирование?

Чтобы избежать утечек чувствительных данных, компании ограничивают к ним доступ сотрудников

Еременк Сергей

21 июля 1960

Ванина, 1, Тамбов

6806 108771

8 926 118-12-12

mario@gmail.com

Антонов Сергей Андреевич

11 февраля 1961

Дорожная 5, Тамбов

6807 203771 8 926 311-89-84 elf@mail.ru

Типовые задачи



Доступ к данным

Доступ к боевым средам находится под строгим контролем, а тестовые среды открыты бóльшему количеству людей



Для тестирования

Качественно протестировать реальную систему можно только на максимально похожих данных



Построение моделей

Для построения аналитических моделей необходимы синтетические данные, свойства которых близки к исходным

H F Labs

Типовые подходы маскирования

Замена на «звездочки»

- Меняет тип данных
Алгоритм не годится для обезличивания чисел и дат
- Слабо защищает популярные имена и адреса
Даже с одной «звездочкой» популярные имена, отчества и города легко расшифровывать
- Убивает смысл данных
Теряется семантика, валидность, социально-демографические характеристики и связи по домохозяйствам

Фамилия, имя, отчество Абдюшев Павел Рашитович	→	А*****В П***л Р*****ч
Дата рождения 21.01.1979	→	2*.*1.1**9
Паспорт 6806 108711	→	6**61****1
Телефон 8 926 118-12-12	→	89*****12
Адрес Москва, Турчанинов пер., б, стр. 2	→	М****а, Т*****в пер., б, стр. *

Замена букв на буквы, цифр на цифры

До замены паспорта

62-летняя москвичка

После замены паспорта

50-летний конструкт неизвестного пола, с несуществующим паспортом и номером телефона

Утеряны данные о поле, возрасте, регионе, паспорте и номере,
Нарушены связи по ФИО и адресу,
Невозможно провести исследование или построить модель,
Тестировщики не смогут проверить — данные не проходят логику и формат.

Фамилия, имя, отчество

Еременко Наталья
Сергеевна

Нонингел Гпъпдбч
Мношнагп

Дата рождения

21.07.1961

11.02.1973

Паспорт

6806 108711

7187 315818

Телефон

8 926 118-12-12

9 037 229-23-23

Адрес

Тамбов,
ул. Советская, 11

Нуктаз,
ец. Зукогвяу, 57

Как работает «Маскировщик»?

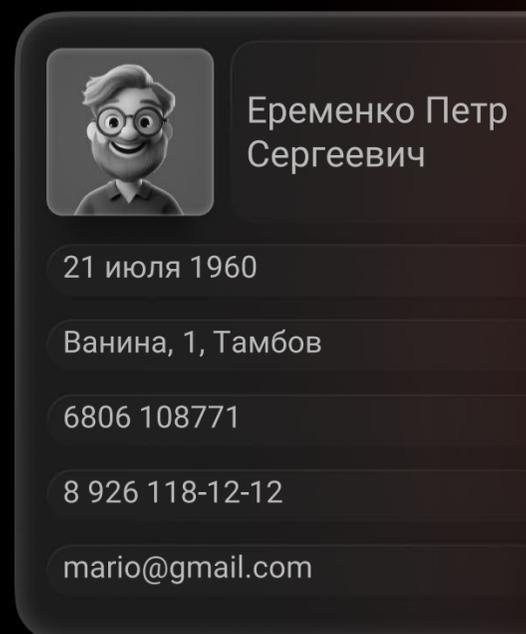
Физлица

Н

F

Labs

Сохранит родственные связи



Еременко Петр
Сергеевич

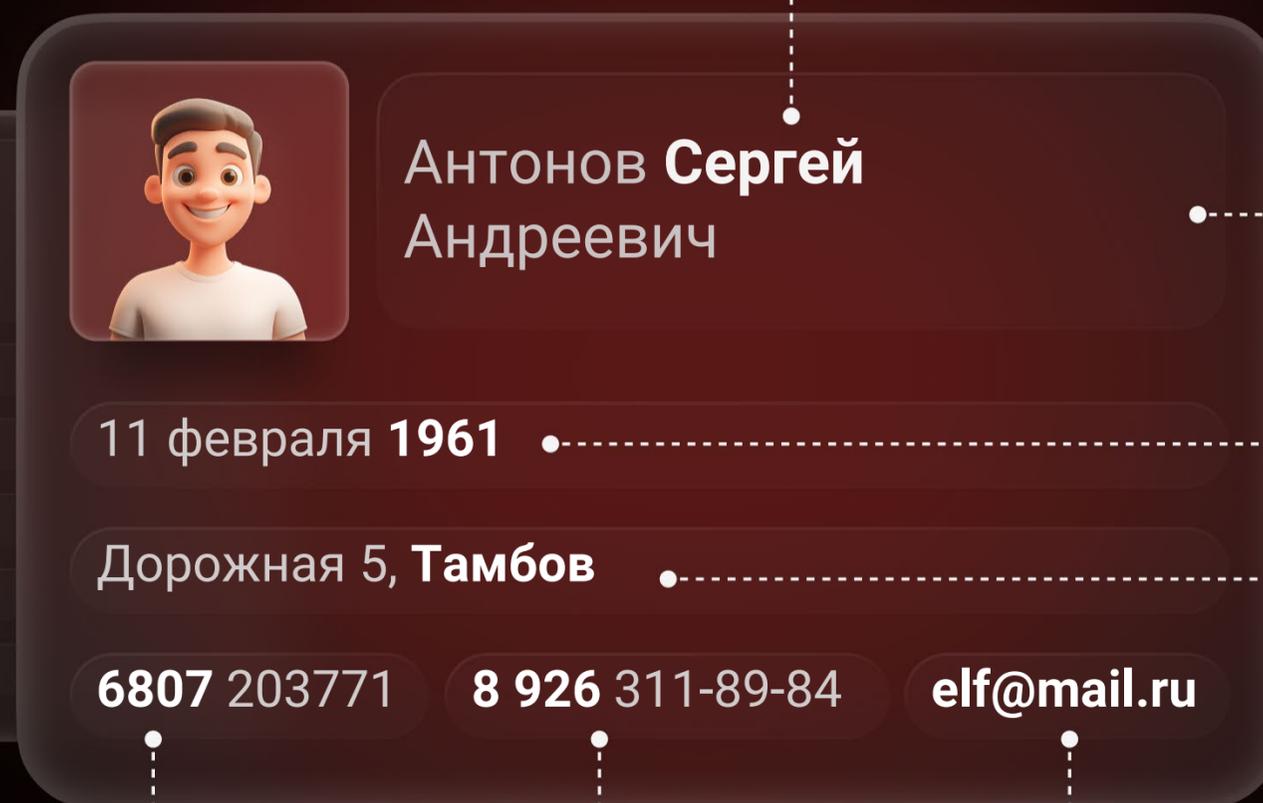
21 июля 1960

Ванина, 1, Тамбов

6806 108771

8 926 118-12-12

mario@gmail.com



Антонов Сергей
Андреевич

11 февраля 1961

Дорожная 5, Тамбов

6807 203771 8 926 311-89-84 elf@mail.ru

Сохранит гендерный
баланс

Сохранит соцдем-группы

Сохранит страну и регион

Серия паспорта учитывает
исходный регион

Оставит
оператора и страну

Понимает домены емейлов:
рабочие, личные, одноразовые

Как работает «Маскировщик»?

🏠 Домохозяйства

H

F

Labs

«Маскировщик» сохраняет аналитику по местам регистрации: например, Ивановы — родственники, живущие по одному адресу. Отчество меняется синхронно с именем отца, адрес — консистентно.

Иванов Петр Ильясович

21.07.1961

Тамбов, Советская, 11

Петров Андрей Фатихович

11.02.1961

Тамбов, Астраханская, 5

Иванов Ильяс Петрович

05.03.1988

Тамбов, Советская, 11

Петров Фатих Андреевич

11.08.1987

Тамбов, Астраханская, 5



Какие задачи решает



Есть MVP по безопасному обмену данными с GPT-агентами, маскированию текстов

- Безопасное обращение данных внутри компании,
- Передача качественных данных для тестирований,
- Кратное снижение рисков утечки ПДн,
- Обучение моделей.



Что маскирует



Учитываем специфику систем. Знаем, что важно при обезличивании АБС банков, MDM-систем, 1С

- Базы данных,
- Сегменты клиентов полностью или по условию,
- Обезличивание по истечению срока использования ПДн.



Решение внедрено



Идет процесс внедрения в 1 крупном банке и 2 ДЗО Сбербанка

- 3 ДЗО Сбербанка,
- «Платферрум»,
- Обезличивание сегмента данных в Банке «Открытие»,
- Обезличивание систем Страховой «Пульс».

H F Labs



«Маскировщик»

С какими типами данных работает «Маскировщик»?

H F Labs

Другие типы данных могут быть обработаны с использованием предустановленной логики замен. Поможем подобрать подходящие варианты базовой замены или разработать новые решения при необходимости

ФИО

01.

Даты
рождения

02.

Документы
физлиц

03.

Данные
юрлиц

04.

Адреса

05.

Номера
телефонов

06.

Емейлы

07.

ИНН

08.

СНИЛС

09.

Банковские
карты

10.

Автомобили

11.

Банковские
счета

12.

Безопасность «Маскировщика»

H F Labs

Можно настроить случайный подбор замен в сессии или сохранение консистентности

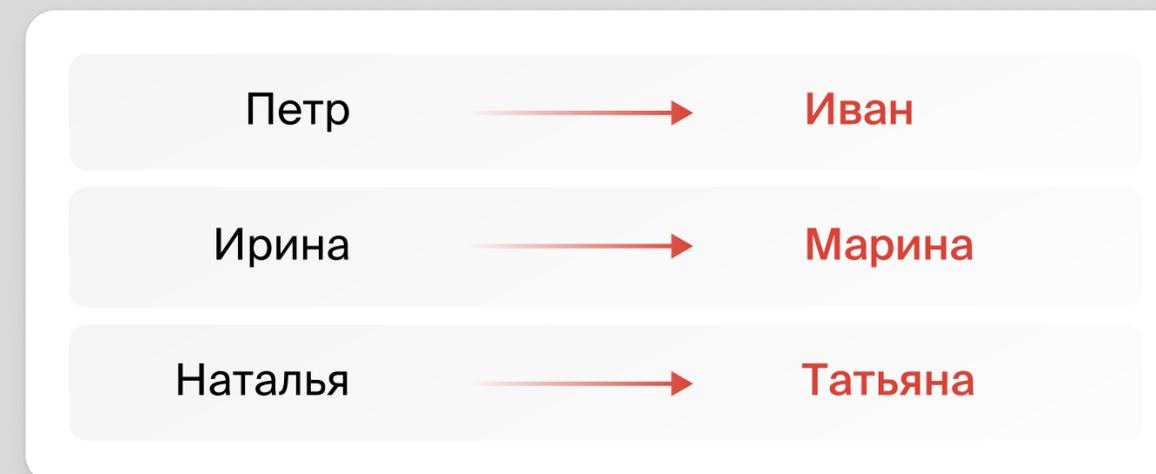
 Пары замен хранятся в виде хешей — набора цифр и букв. Восстановить конкретного человека по этим хешам невозможно

 Набор замен создается каждый раз заново, но его можно сохранить для повторного маскирования нескольких стендов, чтобы поддержать между ними согласованность данных

11 НОЯБРЯ



15 ДЕКАБРЯ



*> Нам нужно одинаково
маскировать несколько
баз данных*

**«Маскировщик»
это умеет**



Сохраняем консистентность!

«Маскировщик» бережет консистентность данных, что особенно важно для тестирования интеграций между несколькими базами данных.

Если сохранить кэш замен, можно:



Маскировать консистентно разные стенды. Последовательно и параллельно, в любое время, с любой физической структурой источников



Инкрементально домаскировать изменения в уже маскированную ранее базу

Ассоциация Больших Данных проверила риск-модели деобезличивания

H F Labs

АБД и HFLabs провели совместное исследование по оценке рисков повторной идентификации.

Специалисты моделировали кибератаки с целью получения ПДн из датасетов. Результаты исследования показали, что «Маскировщик» позволяет сохранить баланс качества и безопасности данных.

Гибкая настройка «Маскировщика»

Позволяет самостоятельно настроить баланс качества и допустимой безопасности в зависимости от целей использования маскированных данных.



[i Подробнее об исследовании](#)

«Маскировщик» обеспечивает безопасную работу с GPT-агентами

H F Labs

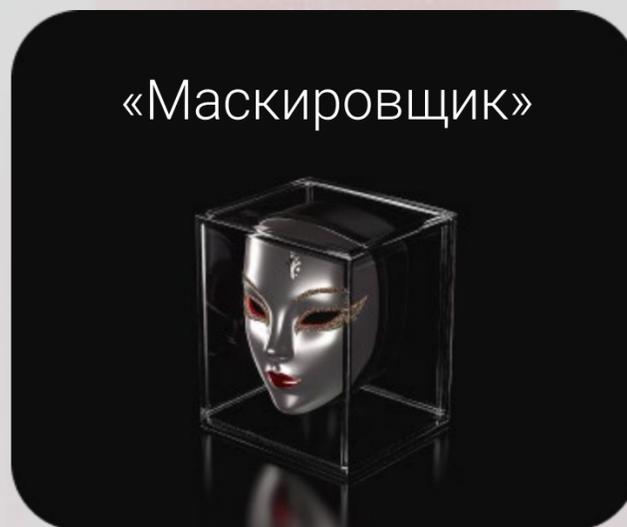
«Маскировщик» умеет самостоятельно определять нужные для маскирования данные и обеспечивать безопасность при работе с ChatGPT



Вопрос
с оригинальными
данными



«Маскировщик»



Вопрос
с замененными
ПД



Ответ
с возвращенными
оригинальными
значениями



Ответ может
содержать
замененные
значения



Пример работы с GPT

mask support

20 марта 2025

О чем текст ниже? Я , Орлов Игорь Владимирович, документ, удостоверяющий личность 7010 № 256423, выдан ТП УФМС Москвы, зарегистрированный (ая) по адресу: Москва ул Мира дом 13 кв 33,согласен (а) на обработку моих персональных данных: (фамилия, имя, отчество; дата рождения; контактный телефон (дом., мобильный, рабочий); адрес проживания; место работы (учебы), должность, ученая степень, ученое звание, адрес организации, информацию о банковских счетах) Администрацией Томской области с целью обработки материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры.

Автор текста, Орлов Игорь Владимирович, даёт согласие на обработку своих персональных данных Администрацией Томской области. Цель обработки — рассмотрение материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры. В согласии указаны персональные данные, которые будут обработаны: Ф. И. О., дата рождения, контактные телефоны, адрес проживания, место работы (учёбы), должность, а также информация об организации и банковских счетах автора.

Отправить

20 марта 2025

О чем текст ниже? Я , Орлов Игорь Владимирович, документ, удостоверяющий личность 7010 № 256423, выдан ТП УФМС Москвы, зарегистрированный (ая) по адресу: Москва ул Мира дом 13 кв 33,согласен (а) на обработку моих персональных данных: (фамилия, имя, отчество; дата рождения; контактный телефон (дом., мобильный, рабочий); адрес проживания; место работы (учебы), должность, ученая степень, ученое звание, адрес организации, информацию о банковских счетах) Администрацией Томской области с целью обработки материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры.

Маскирование завершено

О чем текст ниже? Я , Волков Григорий Айдарович, документ, удостоверяющий личность 5463 № 787580, выдан ТП УФМС Москвы, зарегистрированный (ая) по адресу: **РОССИЯ, ГОРОД МОСКВА, ПРОЕЗД ТРОЕКУРОВСКИЙ**,согласен (а) на обработку моих персональных данных: (фамилия, имя, отчество; дата рождения; контактный телефон (дом., мобильный, рабочий); адрес проживания; место работы (учебы), должность, ученая степень, ученое звание, адрес организации, информацию о банковских счетах) Администрацией Томской области с целью обработки материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры.

Отправка в chat gpt завершена

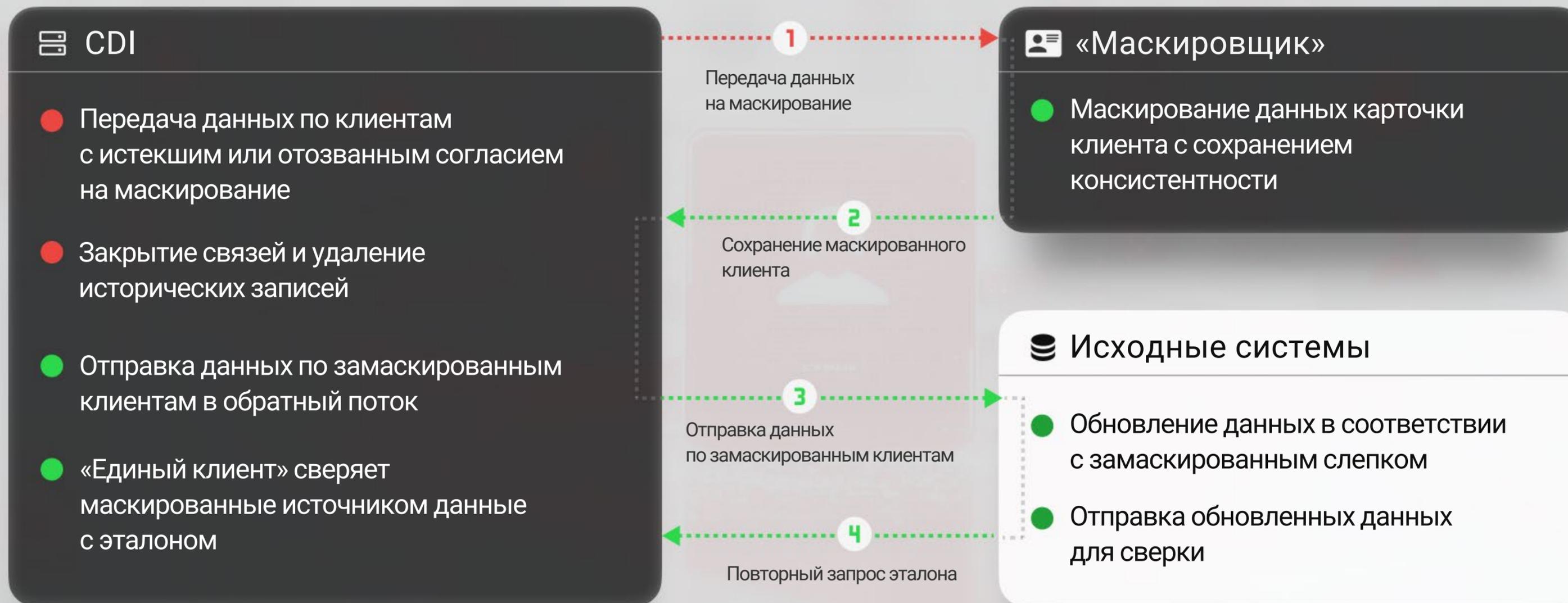
Автор текста, Волков Григорий Айдарович, даёт согласие на обработку своих персональных данных Администрацией Томской области. Цель обработки — рассмотрение материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры. В согласии указаны персональные данные, которые будут обработаны: Ф. И. О., дата рождения, контактные телефоны, адрес проживания, место работы (учёбы), должность, а также информация об организации и банковских счетах автора.

Демаскирование завершено

Автор текста, **Орлов Игорь Владимирович**, даёт согласие на обработку своих персональных данных Администрацией Томской области. Цель обработки — рассмотрение материалов на соискание премии Томской области в сфере образования, науки, здравоохранения и культуры. В согласии указаны персональные данные, которые будут обработаны: Ф. И. О., дата рождения, контактные телефоны, адрес проживания, место работы (учёбы), должность, а также информация об организации и банковских счетах автора.

Маскирование согласий клиентов

H F Labs



И что в итоге?

H F Labs



Из базы данных, содержащих записи о миллионах реальных людей

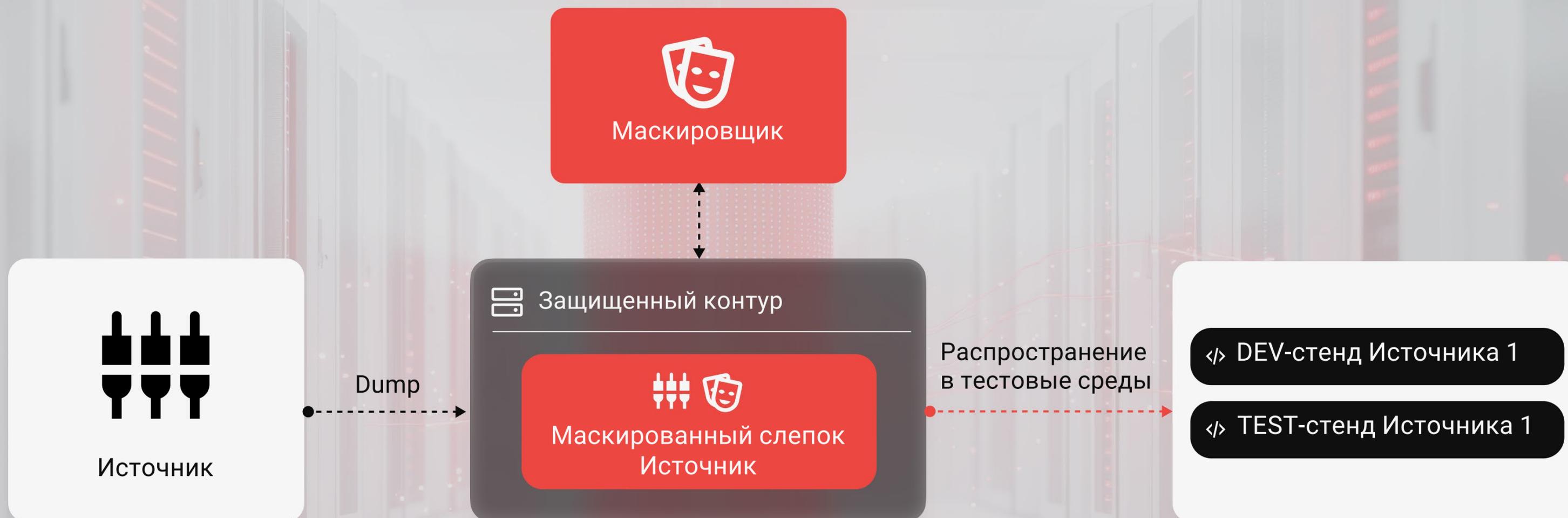
i Сохраняет родственные связи, качество данных, соцдем и географическое распределение



«Маскировщик» делает ровно такую же, неотличимую для человека базу данных людей, только нереальных.

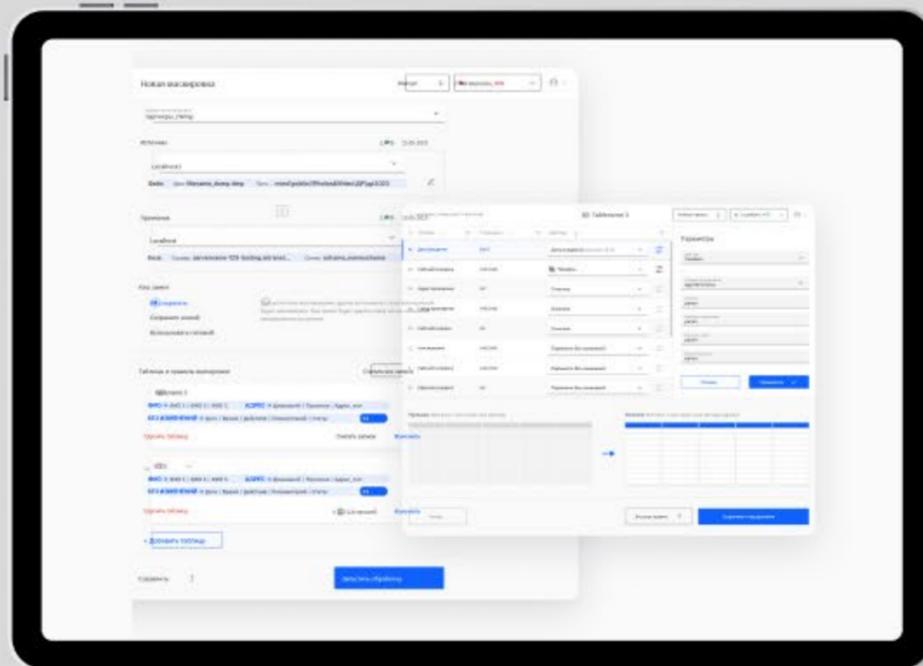
Тестовый стенд под ключ

H F Labs

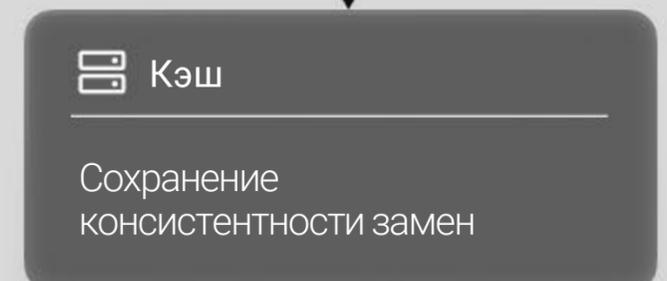
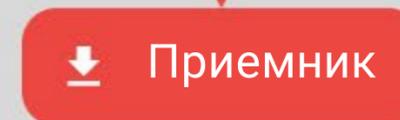
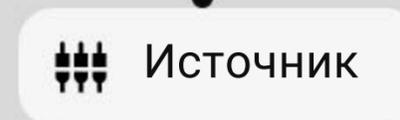
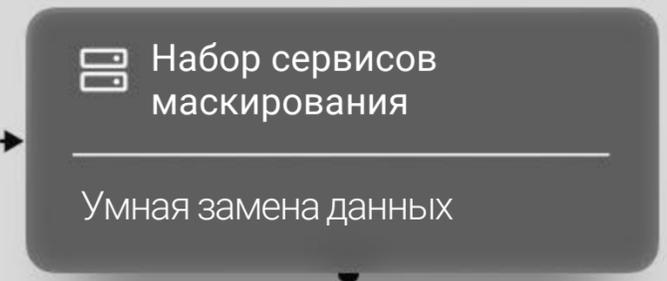
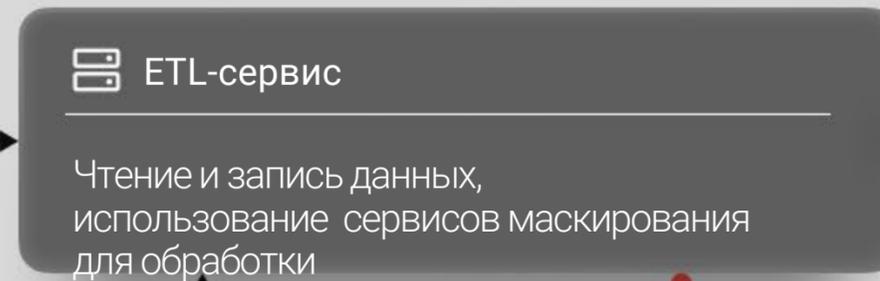


Как это работает технически?

H F Labs



Удобный веб-интерфейс позволяет настраивать конфигурацию маскирования, а также смотреть историю сессий маскирования. Разграничен по правам доступа

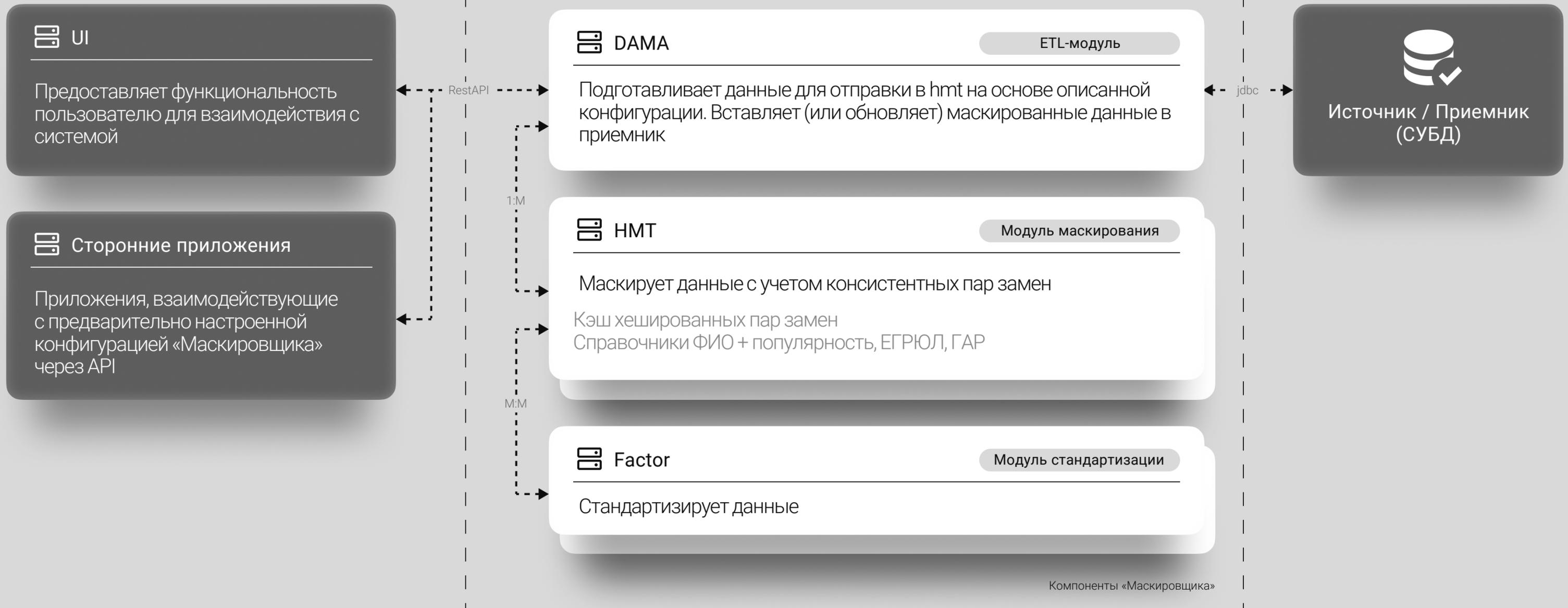


Базы данных на чтение и запись. Из коробки сразу доступны PostgreSQL, MS SQL, Oracle. Можно подключить дополнительные коннекторы



Архитектура «Маскировщика»

H F Labs



Архитектура «Маскировщика»

H F Labs

UI

Предоставляет функциональность пользователю для взаимодействия с системой

Сторонние приложения

Приложения, взаимодействующие с предварительно настроенной конфигурацией «Маскировщика» через API

DAMA

Подготавливает данные для отправки в hmt на основе описанной конфигурации. Вставляет (или обновляет) маскированные данные в приемник

ETL-модуль

Volume
Внутреннее хранилище

HMT

Маскирует данные с учетом консистентных пар замен

Модуль маскирования

Volume
Справочники
Кэш хешированных пар замен

Источник / Приемник (СУБД)

HMT

Модуль стандартизации

Стандартизирует данные

HMT

Модуль маскирования

Маскирует данные с учетом консистентных пар замен

HMT

Модуль маскирования

Маскирует данные с учетом консистентных пар замен

HMT

Модуль стандартизации

Стандартизирует данные

Volume

Справочники
Индексы

Volume

Справочники

Volume

Справочники

Volume

Справочники
Индексы

Cloud infrastructure

Этапы работы Пилот

1 Выбираем набор таблиц, в том числе из разных систем, чтобы подтвердить консистентность

2 Согласуем модель и выбираем оптимальный сервис маскирования для нужных бизнес-сценариев

3 Маскируем

4 Оцениваем результат и скорость

🕒 Длительность: 2–3 недели

✦ **Преимущества:** в результате пилота можно сразу проверить не только маскируемые данные, но и корректную работу приложений, связанных с маскированными объектами, автотесты, интеграцию

1 Масштабируем модели маскирования на все нужные объекты и стенды

2 Помогаем продумать варианты интеграции с учетом имеющихся процессов

3 Обучаем менять модель, добавлять новые поля и т.д.

? А после внедрения?

Мы всегда рядом. Вы будете самостоятельно добавлять новые объекты и поля, но если возникнут вопросы или потребуется добавить системы и внести изменения — мы поддержим, проконсультируем или заберем задачу под ключ

H F Labs

Этапы работы
Внедрение

H F Labs



Спасибо за внимание!



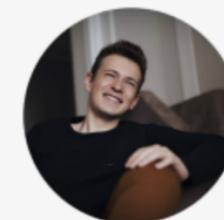
«Маскировщик»
на сайте HFLabs



Ольга Сердобинцева
Владелец продукта «Маскировщик»
✉ olgaser@hflabs.ru



Екатерина Илюхина
Директор по развитию бизнеса
✉ ei@hflabs.ru



Кирилл Ушаков
Менеджер по продажам
✉ kirillush@hflabs.ru