H F Labs CDI CONF 2024

«Фактор»: ФИО — это бизнес-ценность

Артем Калашманов
Аккаунт

Ольга Позднякова
Операционный директор

Содержание. **«Фактор»**

01 Что такое «Фактор»

02 Сложности работы с ФИО

03 Правила стандартизации

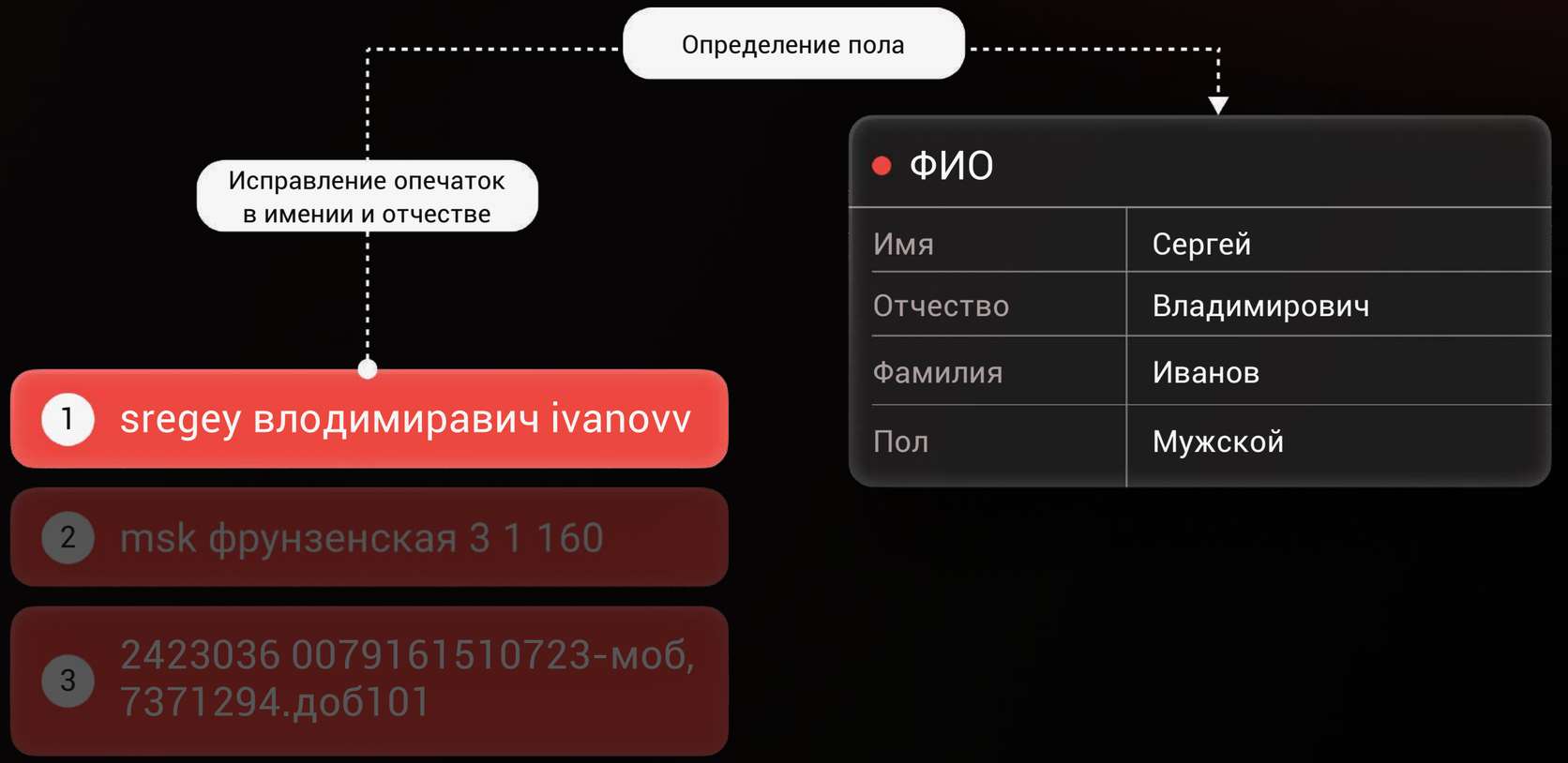
04 Справочники имен

05 Планы на развитие продукта

Что такое «Фактор»?

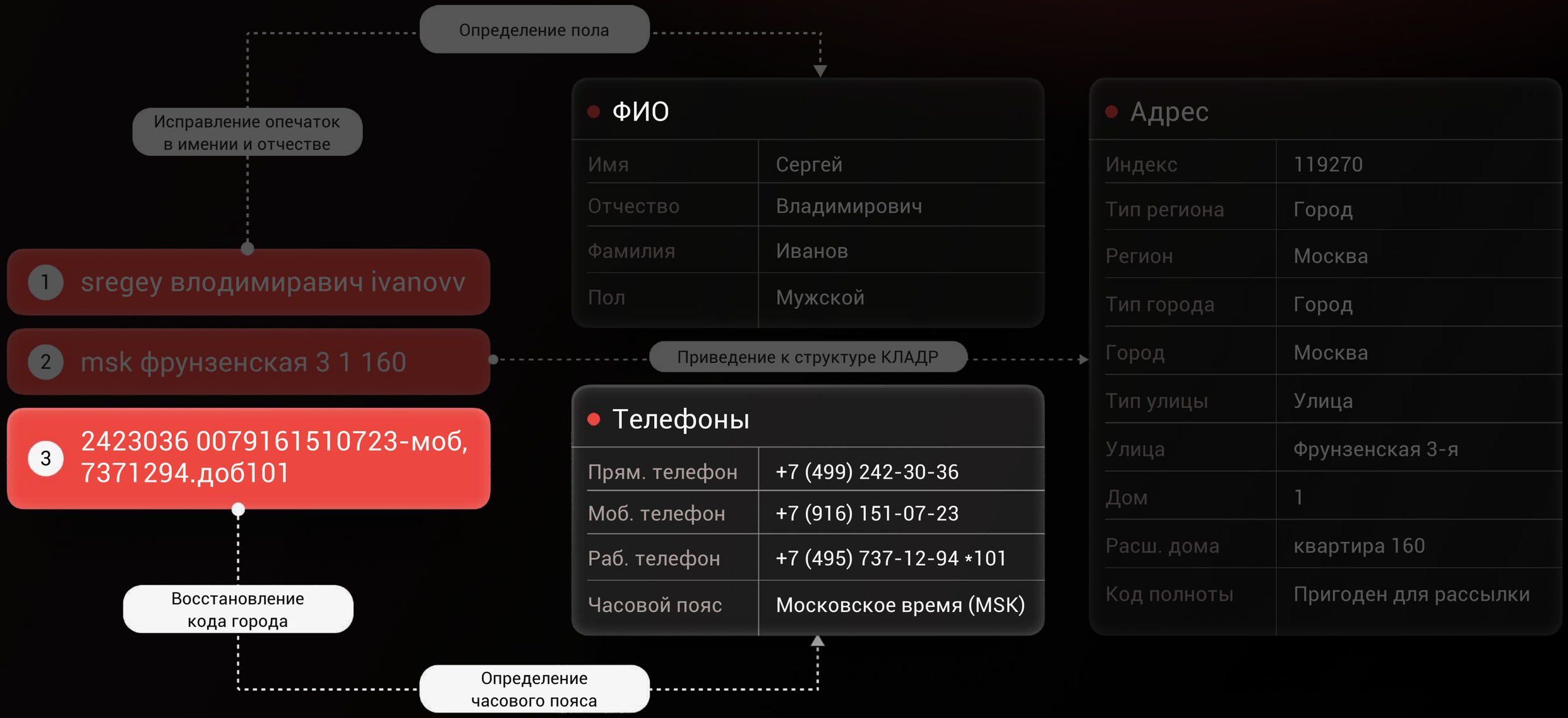
1. sregey влодимиравич ivanovv
2. msk фрунзенская 3 1 160
3. 2423036 00791615107 23-моб, 7371294.доб101

Что такое «Фактор»?

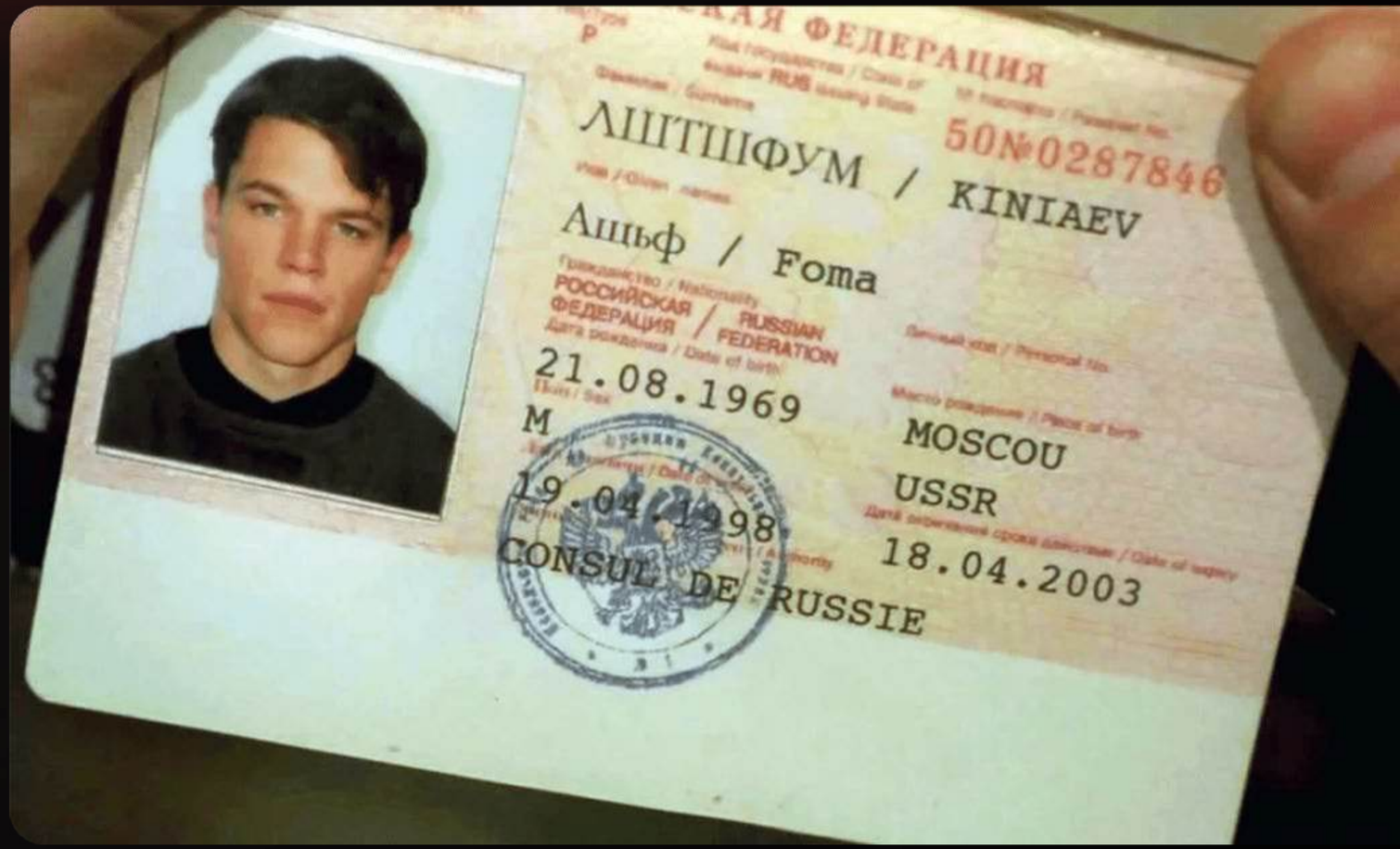


Что такое «Фактор»?

Что такое «Фактор»?



Что в имени
тебе моем?



Исправлять нельзя стандартизировать

Исправлять нельзя стандартизировать

1 Мацаян Саша Алясович	1 Мацаян Саша Алясович
2 Сафиуллин Ахмад Ахмад Оглы	2 Сафиуллин Ахмад Ахмад Оглы
3 Иванов Саша Петрович	3 Иванов Александр Петрович
4 Петров Егор Иванов	4 Петров Егор Иванович

Стандартизируй, но верни исходное

Результирующее ФИО	Предлагаемый вариант
1 Иванов Саша Петрович	1 Иванов Александр Петрович
Пол: мужской	Пол: мужской

Маркеры

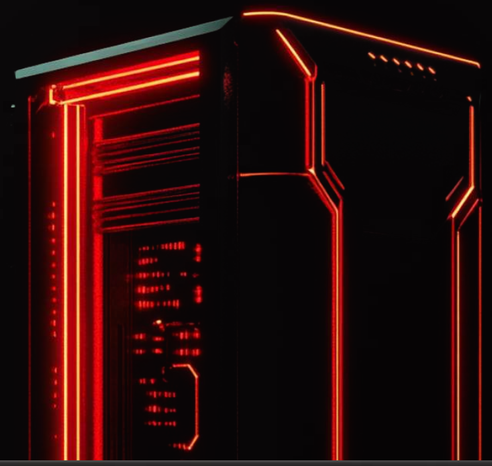
- Вернули исходное ФИО
- Предложены исправления
- Предложена замена на полное имя
- Пол корректный

Когда это полезно

VIP-клиенты

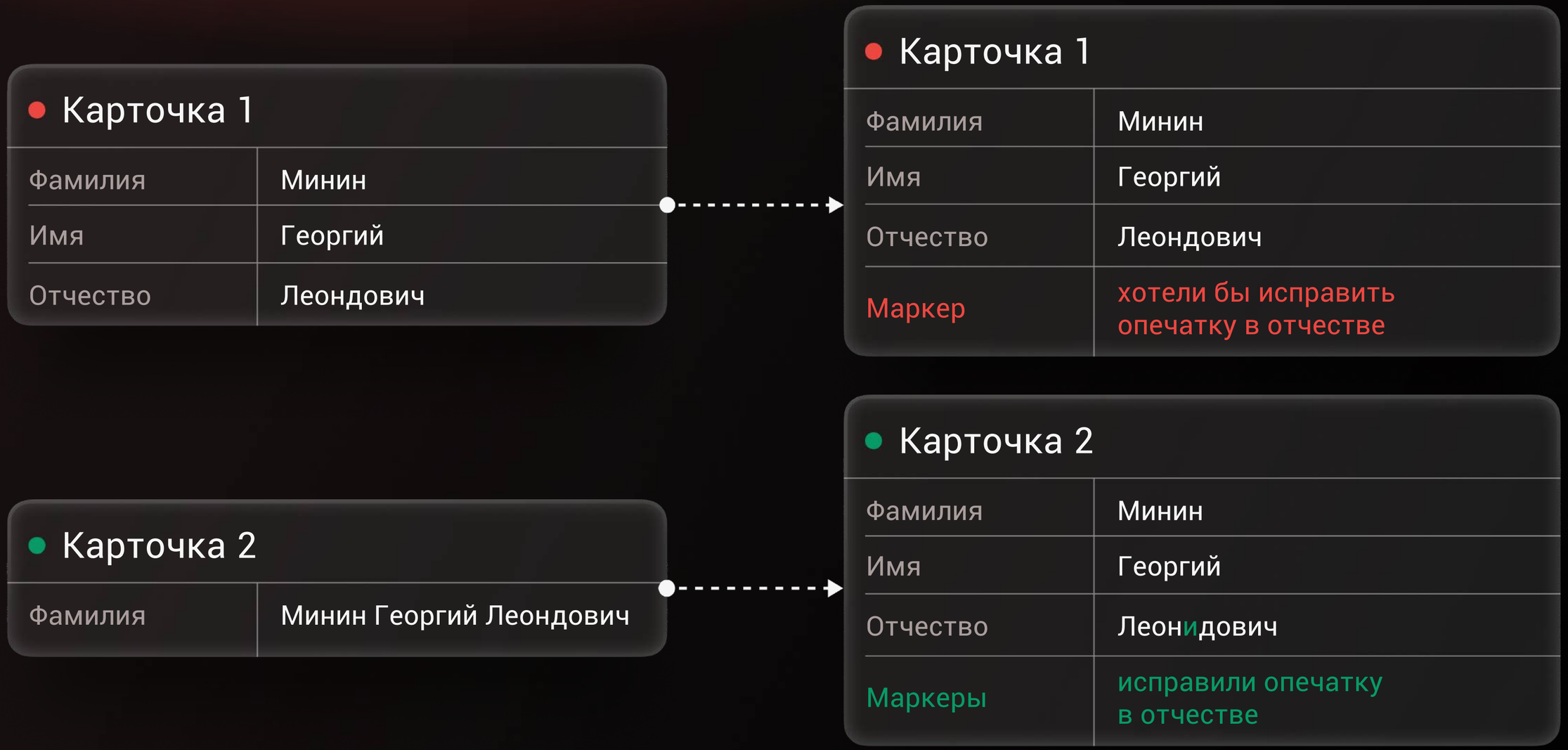
Когда это полезно

VIP-клиенты



Изменения допустимы
только в одной системе

Как работает «вернуть исходное ФИО»?



Тонкое место: поговорим о сложных именах

Тонкое место: поговорим о сложных именах

1	Тарковский Армений Александрович	1	Тарковский Арсений Александрович
2	Сорокина Владимир Георгиевич	2	Сорокин Владимир Георгиевич
3	Летов Игорь Игорь Сергеевич	3	Летов Игорь Сергеевич
4	Мацаян Саша Алясович	4	Мацаян Александр Алясович
5	Иванов Саша Петрович	5	Иванов Александр Петрович

Тюнинг стандартизации

Результирующее ФИО	Предлагаемый вариант
1 Мацаян Саша Алясович	1 Мацаян Александр Алясович
2 Иванов Саша Петрович	2 Иванов Александр Петрович

Маркеры

- Предложено исправление имени
- Предложена замена имени на полное
- Пол корректный

Больше маркеров богу стандартизации

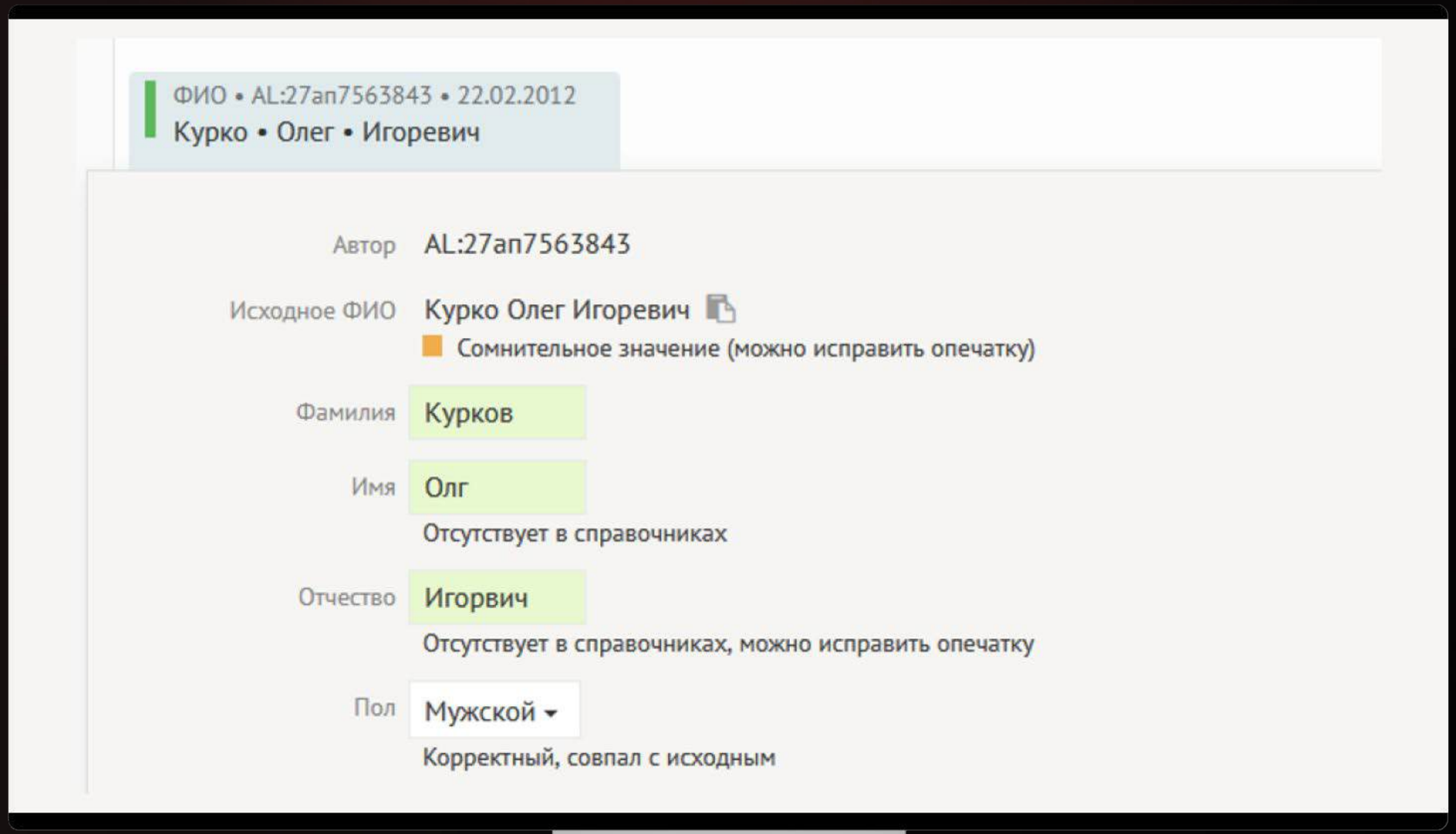
Было

1 ФИО исправлено

Больше маркеров богу стандартизации

Было	Стало
1 ФИО исправлено	1 Исправили опечатку по Левенштейну
	2 Исправили опечатку по справочнику
	3 Заменили на полный компонент

Как это выглядит



Чем подробные маркеры полезны бизнесу

```
select * from dev_customer.fio f where f.surname_qc like 'UNNATURAL';
```

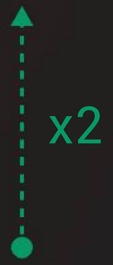
io 1

	id_fio	hid_fio	surname	name	patronymic	qualitycode	surname_qc	firstname_qc	patronymic_qc	gender_qc
1	1 230	1 230	Карачай	Абу Имран	[NULL]	NOT_SURE	UNNATURAL	UNNATURAL	EMPTY	UNDEF
2	3 986	3 986	Эмин	Азери	Ильяс оглы	NOT_SURE	UNNATURAL	UNNATURAL		UNDEF
3	26 986	26 986	Сойтекин	Али	[NULL]	NOT_SURE	UNNATURAL	UNNATURAL	EMPTY GARBAGE	UNDEF
4	9 524	9 524	Одина	Восидзода	Муродбеги	NOT_SURE	UNNATURAL	UNLIST	UNLIST UNNATURAL	UNDEF
5	10 552	10 552	Курбанмагомедов	Ибрагимов	Дадаш	NOT_SURE	UNNATURAL	UNLIST	FEW_UNIQUE_LETTERS UNLIST UNNATURAL	UNDEF
6	11 732	11 732	Абдаллах	Яхья	[NULL]	NOT_SURE	UNNATURAL	UNNATURAL	EMPTY	UNDEF
7	13 095	13 095	Сень	Ли Ень	[NULL]	NOT_SURE	UNNATURAL	UNLIST UNNATURAL	EMPTY	UNDEF
8	28 461	28 461	Эмир	Хаттаб	[NULL]	NOT_SURE	UNNATURAL	UNNATURAL	EMPTY	UNDEF
9	15 997	15 997	Осман	Ухеида	Сулейман	NOT_SURE	UNNATURAL	UNLIST	UNLIST UNNATURAL	UNDEF
10	17 265	17 265	Бок	Бек Им	[NULL]	NOT_SURE	UNNATURAL	UNLIST UNNATURAL	EMPTY	UNDEF
11	21 915	21 915	Васильевич	Роман	[NULL]	NOT_SURE	UNNATURAL	UNNATURAL	EMPTY	UNDEF

Больше качественных данных

Справочник имен

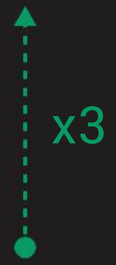
98 тыс.



~~46 тыс.~~

Справочник отчеств

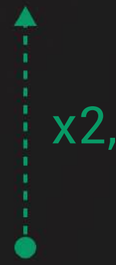
266 тыс.



~~81 тыс.~~

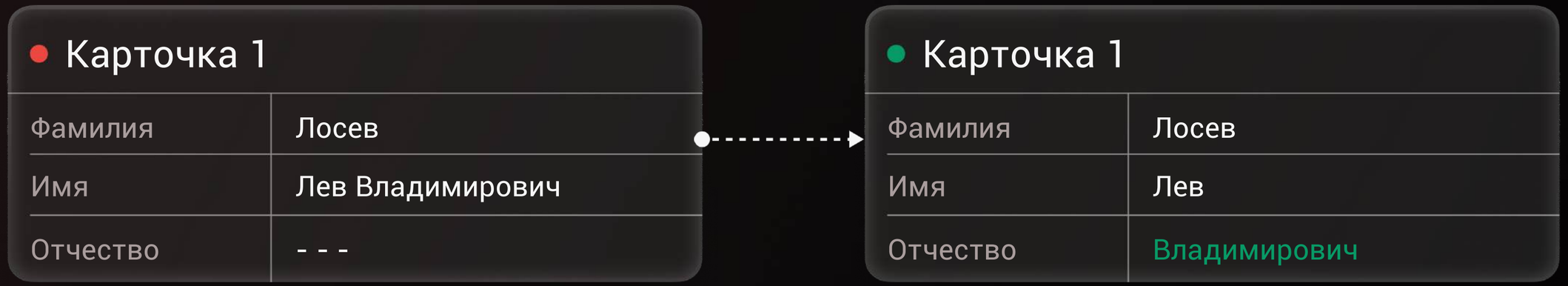
Справочник фамилий

1076 тыс.

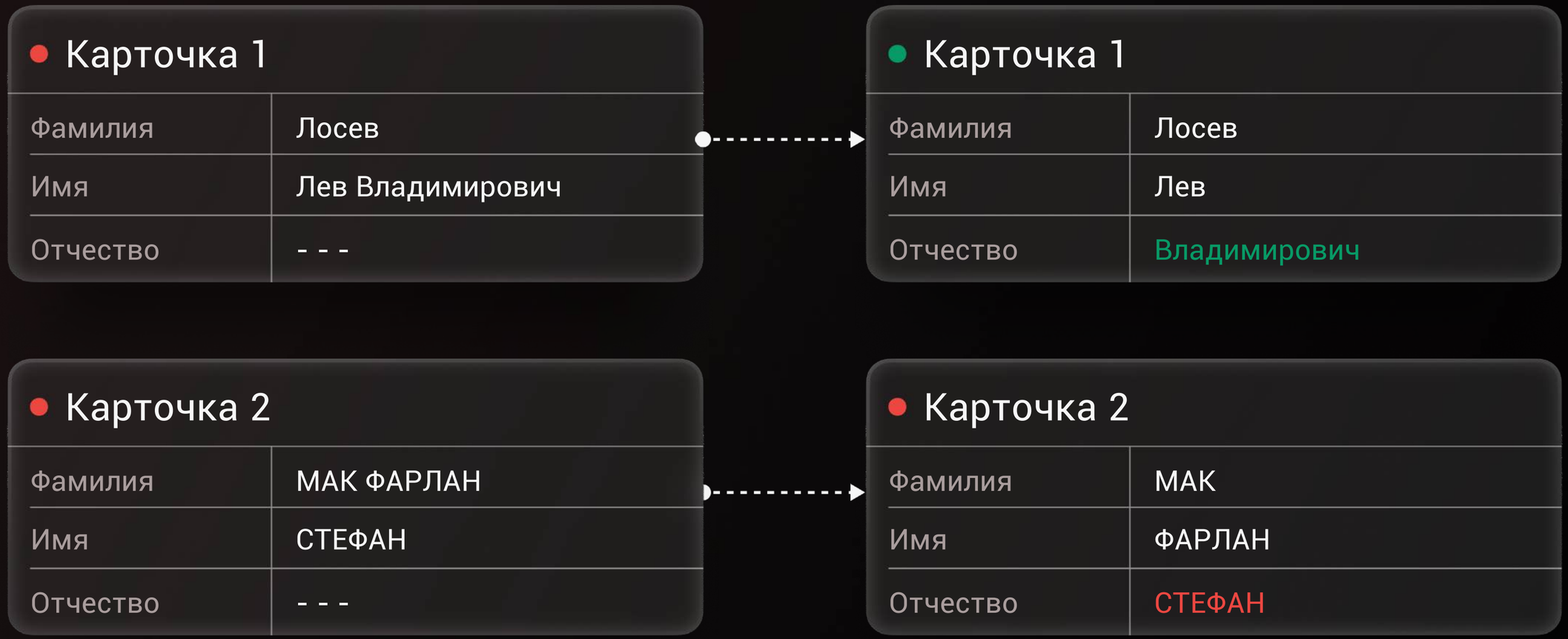


~~437 тыс.~~

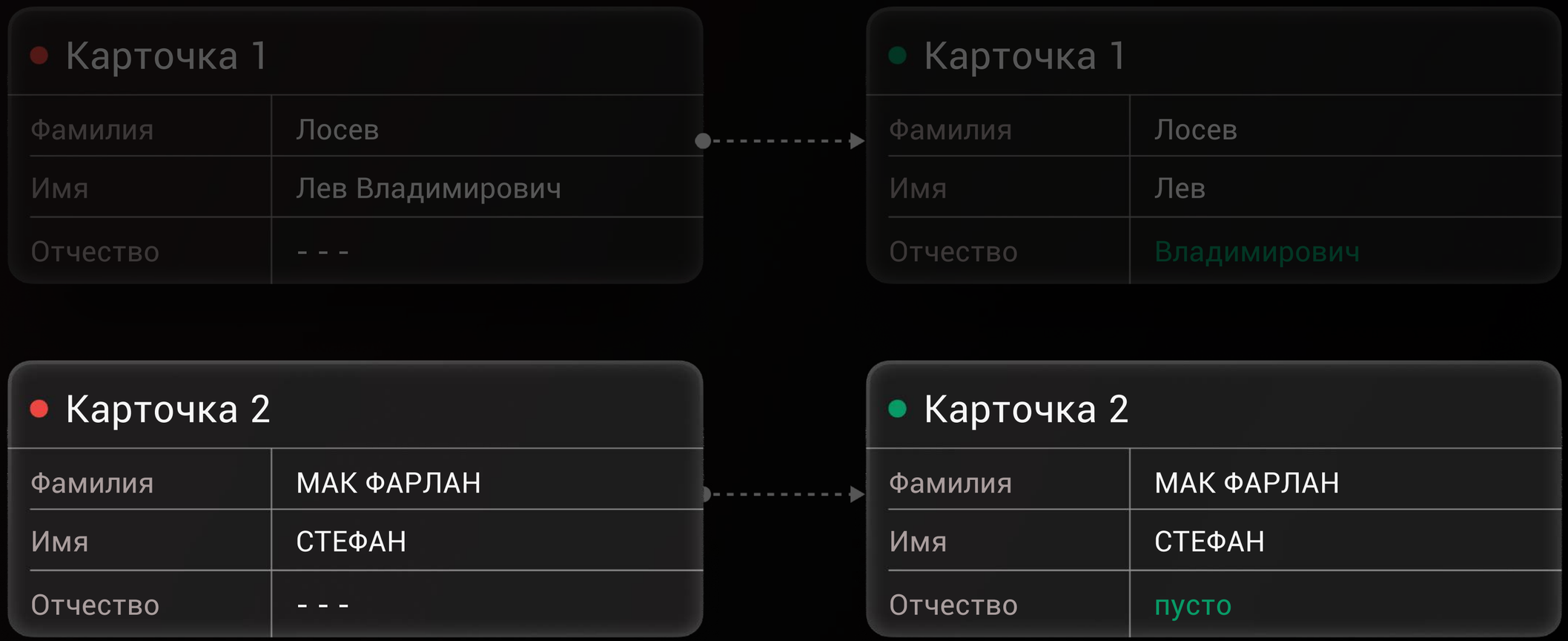
Не заполняем ради заполнения



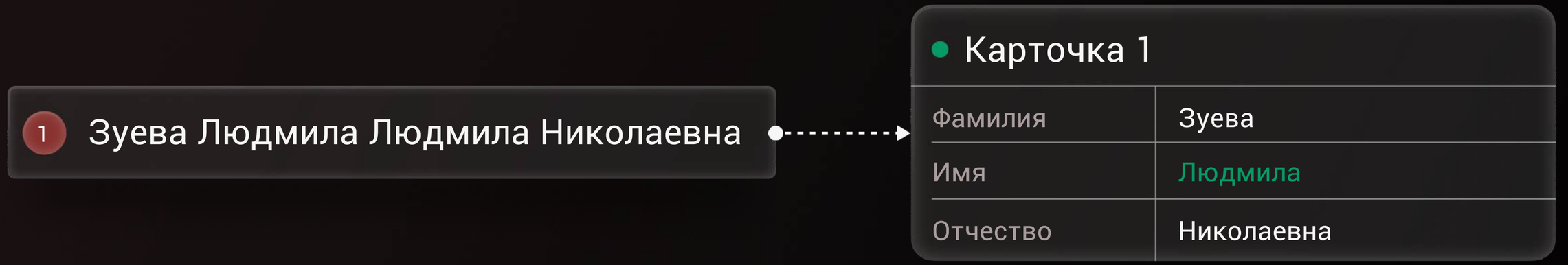
Не заполняем ради заполнения



Не заполняем ради заполнения



Два раза не повторяю, не повторяю



Два раза не повторяю, не повторяю

1 Зуева Людмила Людмила Николаевна

Карточка 1	
Фамилия	Зуева
Имя	Людмила
Отчество	Николаевна

2 Галина Галина Петровна

Карточка 2	
Фамилия	пусто
Имя	Галина
Отчество	Петровна

Два раза не повторяю, не повторяю

1 Зуева Людмила Людмила Николаевна

Карточка 1	
Фамилия	Зуева
Имя	Людмила
Отчество	Николаевна

2 Галина Галина Петровна

Карточка 2	
Фамилия	Галина
Имя	Галина
Отчество	Петровна

Это фамилия или отчество?

Было

1 ОЛЬГА СЕРГЕЕВИЧ

Карточка

Фамилия	пусто
Имя	Ольга
Отчество	Сергеевич

Пол: неизвестен

Стало

2 ОЛЬГА СЕРГЕЕВИЧ

Карточка

Фамилия	Сергеевич
Имя	Ольга
Отчество	пусто

Пол: женский

Пара слов о сложных суффиксах

Было

1 МЕХТИ-ЗАДЕ НАЗИМ АГА-ОГЛЫ

Карточка

Фамилия	МЕХТИ-ЗАДЕ
Имя	НАЗИМ АГА
Отчество	пусто

Стало

2 МЕХТИ-ЗАДЕ НАЗИМ АГА-ОГЛЫ

Карточка

Фамилия	МЕХТИ-ЗАДЕ
Имя	НАЗИМ
Отчество	АГА-ОГЛЫ

Fuf! Ошибка раскладки

Что дальше?

1. Расширение справочников мусора

(Радио, Умер и т. д.)

2. Склейка компонентов, случайно разделенных пробелом

(Лерм онтов Михаил Юрь евич)

3. Контекст ФИО и уменьшительно-ласкательные компоненты

(Миша Оганесян)

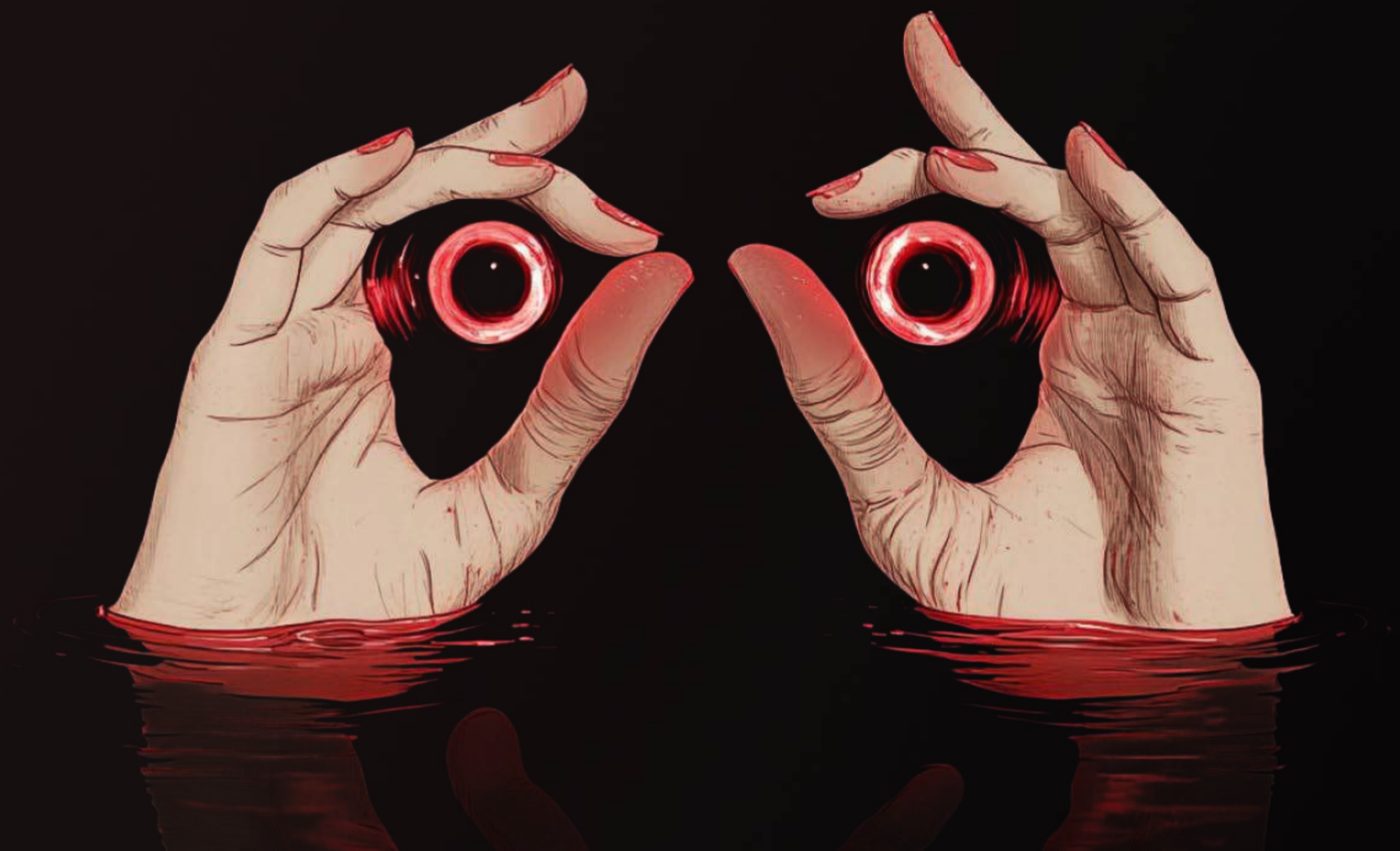
4. Новые маркеры

Популярность ФИО

Гарантированные дубликаты

Негарантированные дубликаты

Трудно разбирать
руками и глазами



Но мы умнее!

id	Фамилия	Имя	Отчество	ДР
1	Абамелек	Есения	Франковна	1996-06-17
2	Эмир	Бурухтан	Второй-второй	1996-06-17
3	Эмир	Бурухтан	Второй-первый	1996-06-17
4	Травин	Максим	Анатольевич	1995-07-10
5	Травин	Максим	Анатольевич	1995-05-05
6	Травин	Максим	Анатольевич	1985-08-10
N	...	...	...	...

H F Labs CDI CONF 2024

Задавайте ваши вопросы

Артем Калашманов
Аккаунт
artemk@hflabs.ru

Ольга Позднякова
Операционный директор
olgap@hflabs.ru