

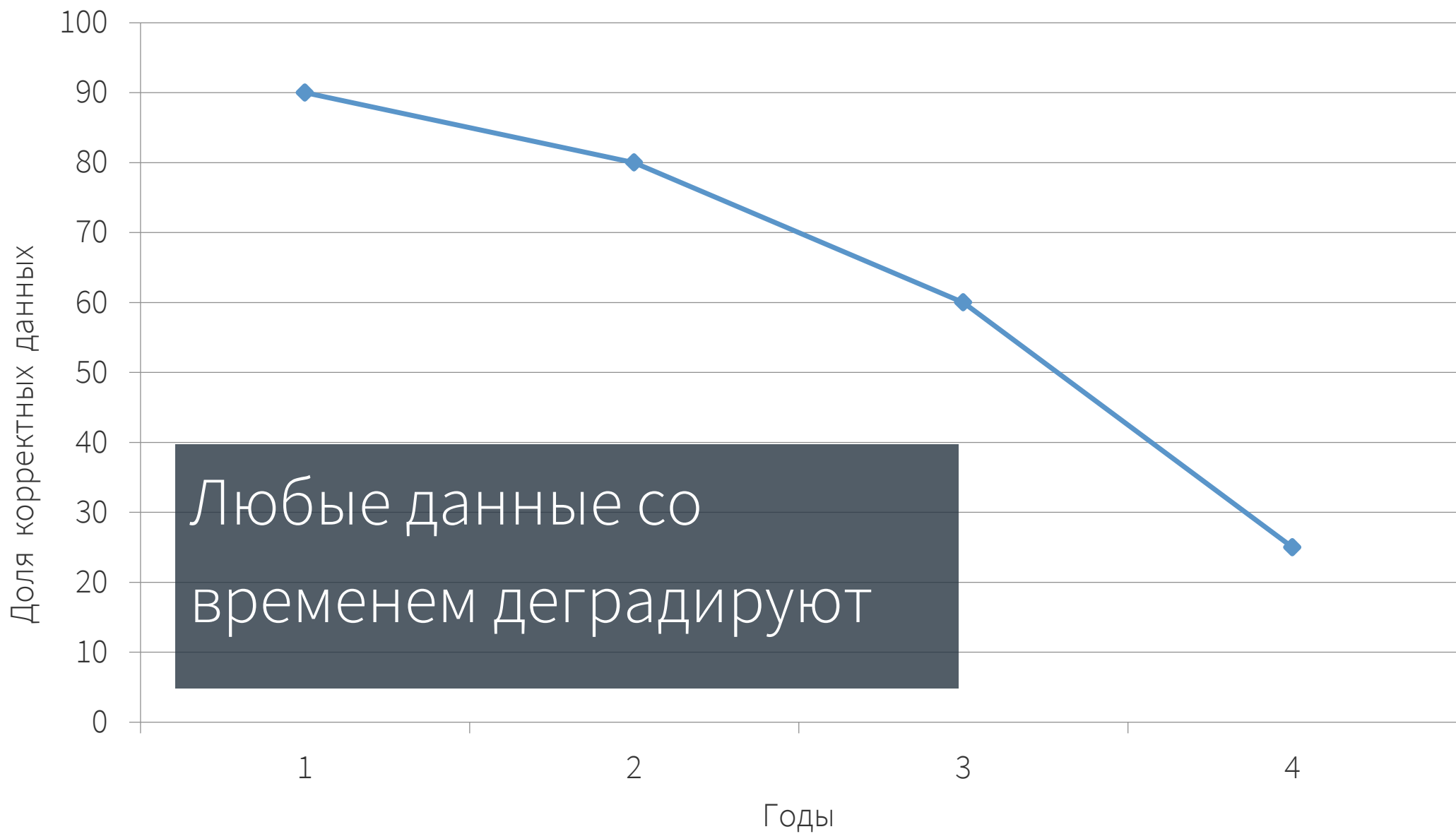


Loginom Data Quality: очистка данных



Люди ошибаются, поэтому проблемы с качеством данных будут всегда:

- Пропуски
- Опечатки
- Дубли и противоречия
- Фиктивные сведения
- Устаревшие данные



Последствия:

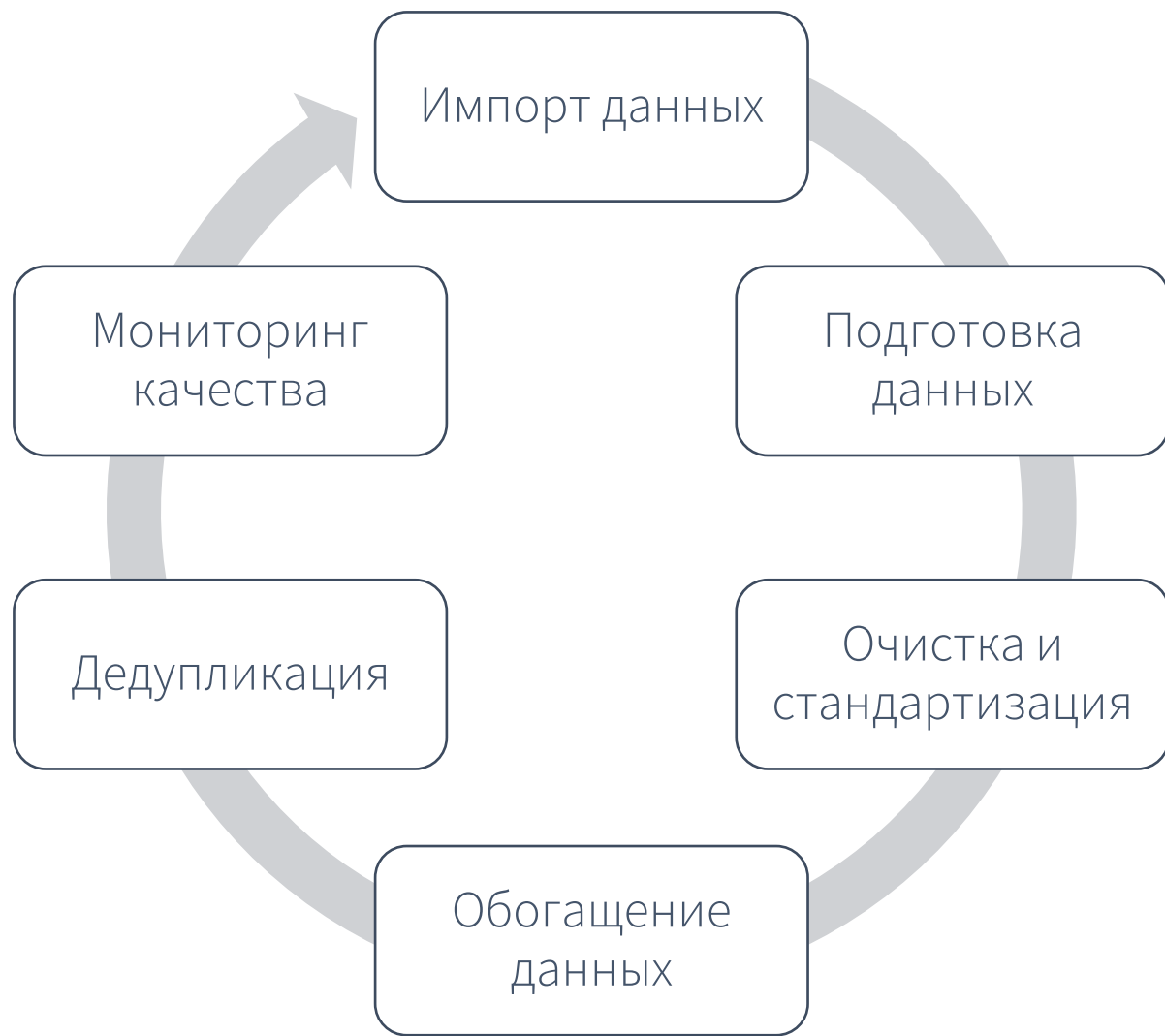
- Безадресный маркетинг
- Раздражающие коммуникации
- Потерянные контакты
- Отток клиентов
- Искаженная отчетность

| Поле | Значение | Ошибка |
|---------|--|--------------------------|
| Имя | Сергей | Первая буква - латинская |
| Фамилия | Петрович | Неверное поле |
| Страна | Лимония | Нет такой страны |
| Город | Мсква | Опечатка |
| Телефон | 0000000 | Фиктивный номер |
| E-mail | 1@siteforspam.com | Нет такого домена |
| Адрес | Новые Васюки 10 | Нет адреса |
| Паспорт | АБВГДЕЖЗ | Не стандарт |



Решение Loginom Data Quality

Автоматически
исправляет ошибки,
приводит к
стандартному виду,
восстанавливает и
обогащает
клиентские данные




Loginom Data Quality

поддерживает

все этапы

очистки данных

| Очищаемые домены | Физ. лица | Юр. лица |
|------------------------|-----------|----------|
| Фамилия, имя, отчество | ● | |
| Название организации | | ● |
| Почтовый адрес | ● | ● |
| Телефоны | ● | ● |
| Электронная почта | ● | ● |
| Удостоверения личности | ● | |
| Реквизиты организаций | | ● |
| Даты | ● | ● |
| Банковские реквизиты | | ● |

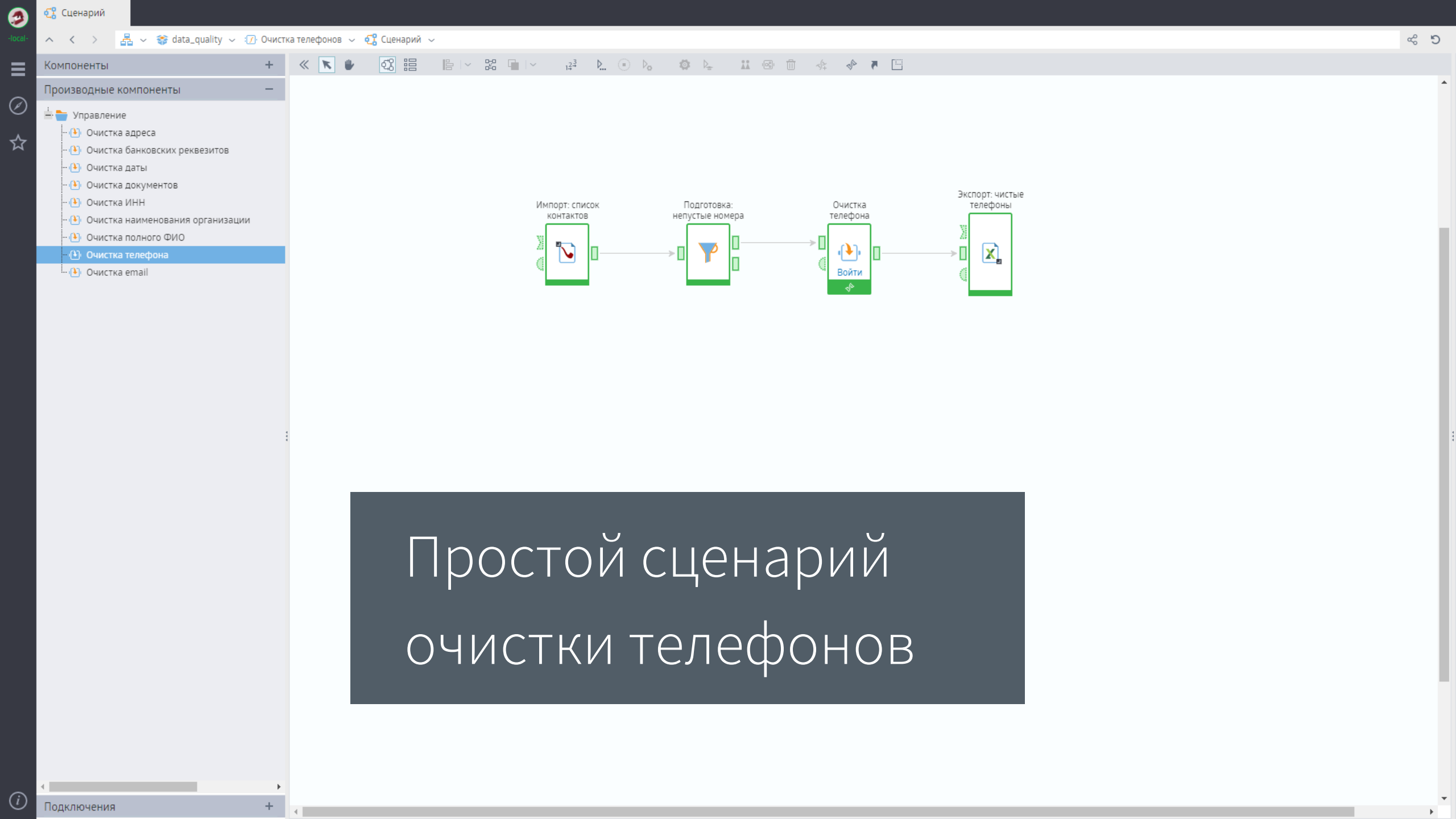


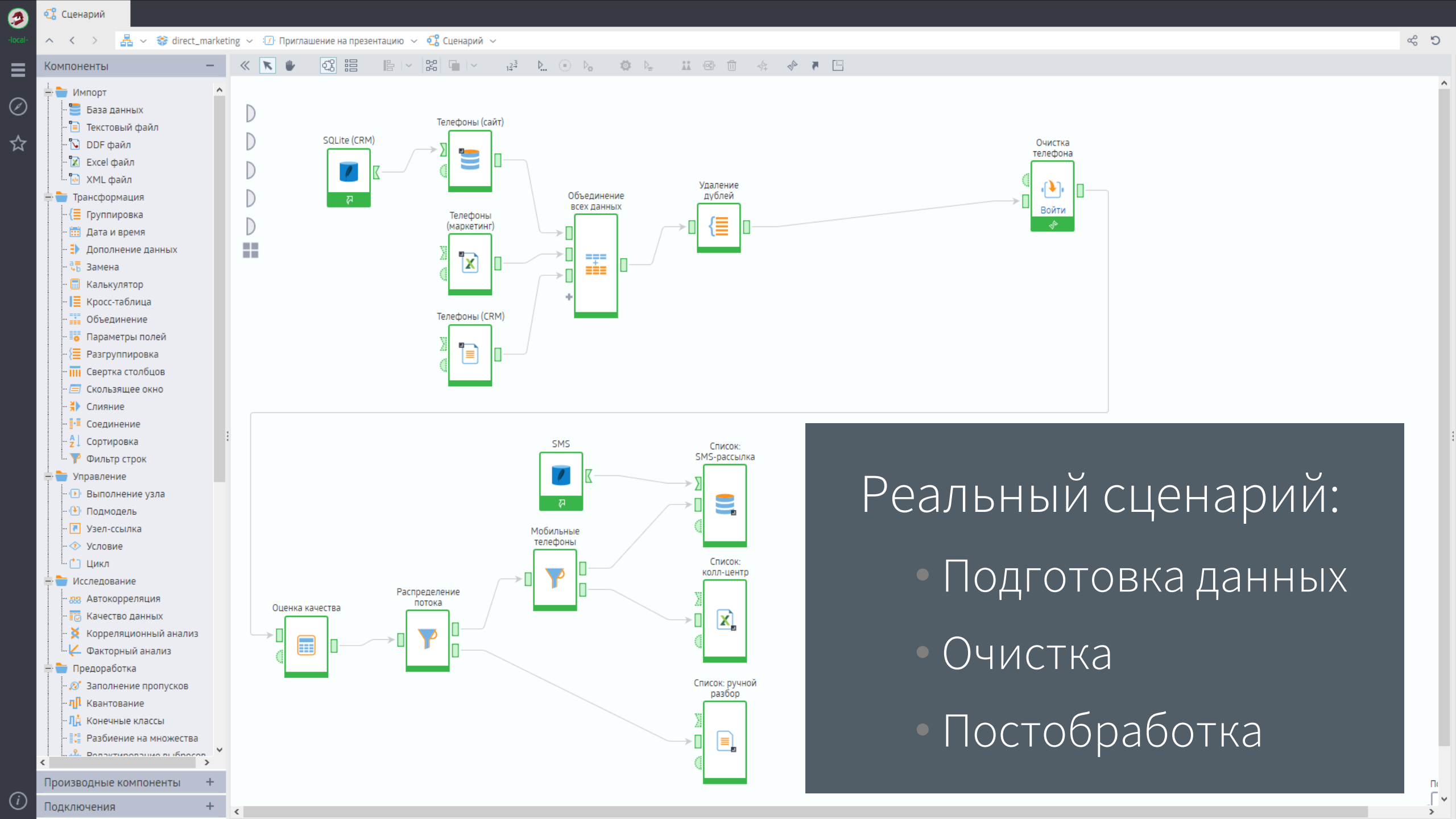
Состав решения

1. Компоненты Loginom для каждого домена
2. Подготовленные справочники
3. Документация
4. Демонстрационный пример

Пример очистки ФИО







Реальный сценарий:

- Подготовка данных
- Очистка
- Постобработка

| | |
|-----------------------------|--|
| ab ID | 8 |
| ab Исходный номер телефона | 8 908 92 24933 справочная |
| ab Очищенный номер телеф... | +7 (908) 9224933 |
| ab Добавочный номер теле... | |
| ab Тип номера телефона | Мобильный |
| ab Остаток строки | справочная |
| ab Код страны | 7 |
| ab Код ABC/DEF | 908 |
| ab Номер телефона | 9224933 |
| ab Страна | Россия |
| ab Регион | Свердловская область |
| ab Город | |
| ab Оператор | ООО "ЕКАТЕРИНБУРГ-2000" |
| ab Часовая зона 1 | МСК+2 |
| ab Часовая зона 2 | московское время плюс 2 часа |
| ab Часовая зона 3 | UTC+5 |
| ab Лог | В номере 11 цифр и первая 8, она заменяется на 7 |
| 7 Дата очистки | 20.10.2017 18:00:20 |


Было

Стало

Проблемы
дедупликации:
противоречия
и неполнота
данных

| Домен | Персона 1 | Персона 2 |
|---------------|--|--|
| ФИО | Иванов Иван Ильич | Иван Ильич |
| Адрес | | г. Рязань, Новая 53в |
| Телефон | +7 (4912) 24-09-77 | |
| Дата рождения | 1971 г. | 15 декабря |
| E-mail | ivanoff@mail.ru | ivanoff@gmail.com |
| Место работы | BaseGroup Labs | BGL |
| Источник | CRM-система | Facebook |

Возможности дедупликации:

1. Настройка стратегий поиска дублей
 2. Задание сочетаний полей, определяющих дубли
 3. Накопление групп дублей
 4. Нечеткий поиск с заданной точностью
 5. Обработка пропусков
- 

Создание «золотой записи» не входит в состав решения, она формируется в рамках проекта с учетом:

- Доверия к источнику данных
- Точного соответствия значений
- Частичного соответствия
- Нечеткой логики
- Кросс-проверок

| Домен | Записей/сек. |
|------------------------|--------------|
| Фамилия, имя, отчество | 540 |
| Название организации | 262 |
| Почтовый адрес | 18 |
| Телефоны | 714 |
| Электронная почта | 9 091 |
| Удостоверения личности | 1 785 |
| ИНН | 2 272 |
| Даты | 7 142 |
| Банковские реквизиты | 562 |

Набор:

100 000 записей

Режим:

Пакетная обработка

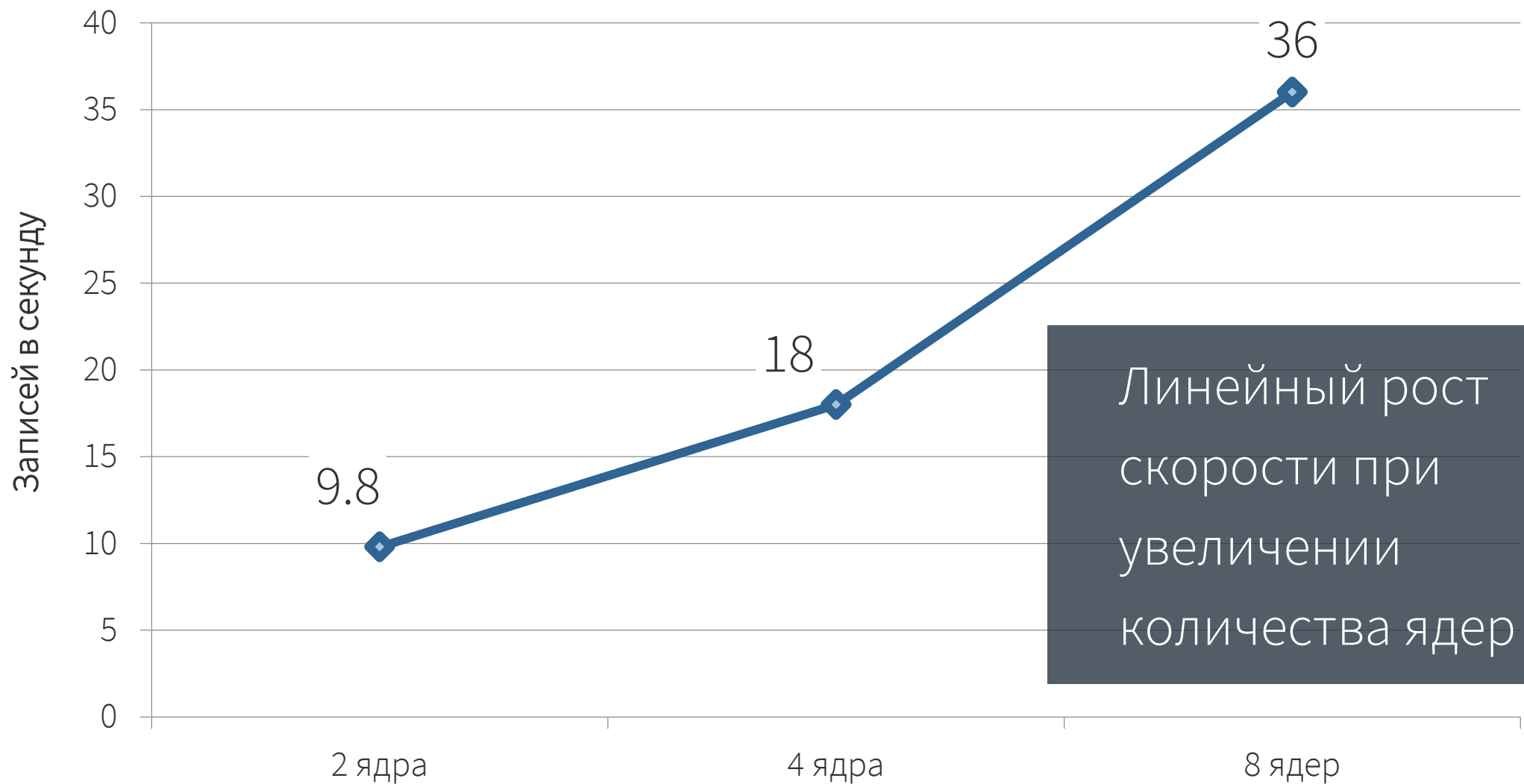
Сервер:

Windows Server 2012 R2

x64/ Intel Xeon E5, 4 ядра

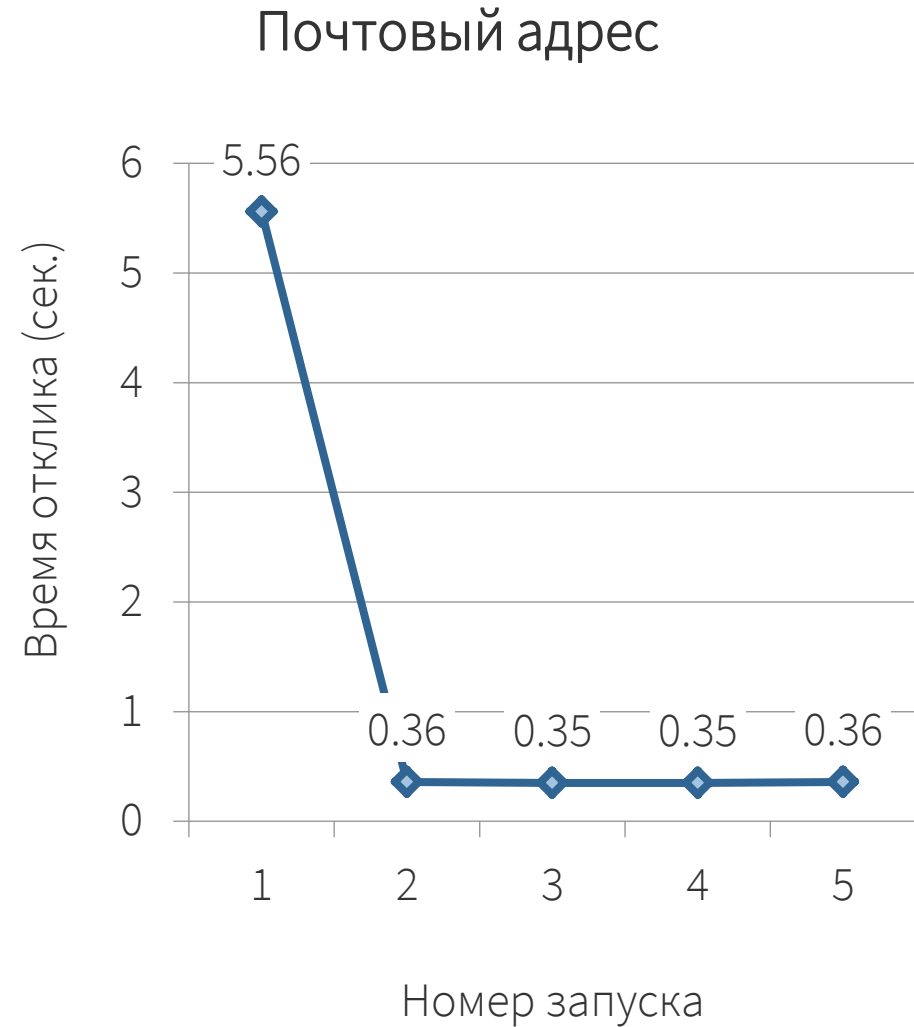
2.60 GHz/16 Gb/1 Tb

Пример: Очистка почтовых адресов



Веб-сервис:

1. Работает медленнее из-за накладных расходов
2. Сильнее зависимость от сетевой инфраструктуры
3. Нужно время на первое обращение – «разогрев»



| Поставляемые справочники | Записей |
|---------------------------|-------------|
| Почтовые адреса: до улиц | 1 400 000 |
| Почтовые адреса: дома | 26 000 000 |
| Фамилии | 700 000 |
| Имена | 10 000 |
| Отчества | 10 000 |
| Операторы связи/регионы | 380 000 |
| Недействительные паспорта | 100 000 000 |
| E-mail домены | 33 000 |

50 Гб

поставляемых
справочников

Вариант

Назначение


Пакетная
обработка

1. Первичная очистка данных
2. Регулярная регламентная очистка
3. Эпизодическая очистка больших наборов данных

Веб-сервис


1. Online очистка
 2. Очистка небольших наборов данных
 3. Интеграция с другими системами
-

Обеспечение качества данных –
постоянный процесс:

- Входной контроль
 - Регулярная очистка
 - Управление мастер-данными
- 

Содержание проекта

1. Интеграция с IT-системами
2. Специфичная пред- и постобработка
3. Интеграция с мастер-данными
4. Настройка регулярной очистки
5. Мониторинг качества данных



Быстрый
запуск

1 час – от получения данных
до первого результата

Гарантии
качества

Исправление сотен ошибок,
актуальные справочники

Визуальный
конструктор

Интеграция и настройка
логики без кодирования

loginom.ru

