

Loginom Data
Quality – НОВЫЕ
ВОЗМОЖНОСТИ



Решение Loginom Data Quality

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

Выход первой версии –
январь 2018 г.

Задача	Описание
Очистка	Исправление ошибок, опечаток, фиктивных данных
Обогащение	Дополнение полезными данными, заполнение пропусков, актуализация
Стандартизация	Унификация представления данных
Дедупликация	Обнаружение дублей и «похожих» записей, объединение их в группы
Формирование «золотой записи»	Создание профиля клиента, включающего актуальные данные всех систем

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

Телефон – пример очистки

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

Исходные данные

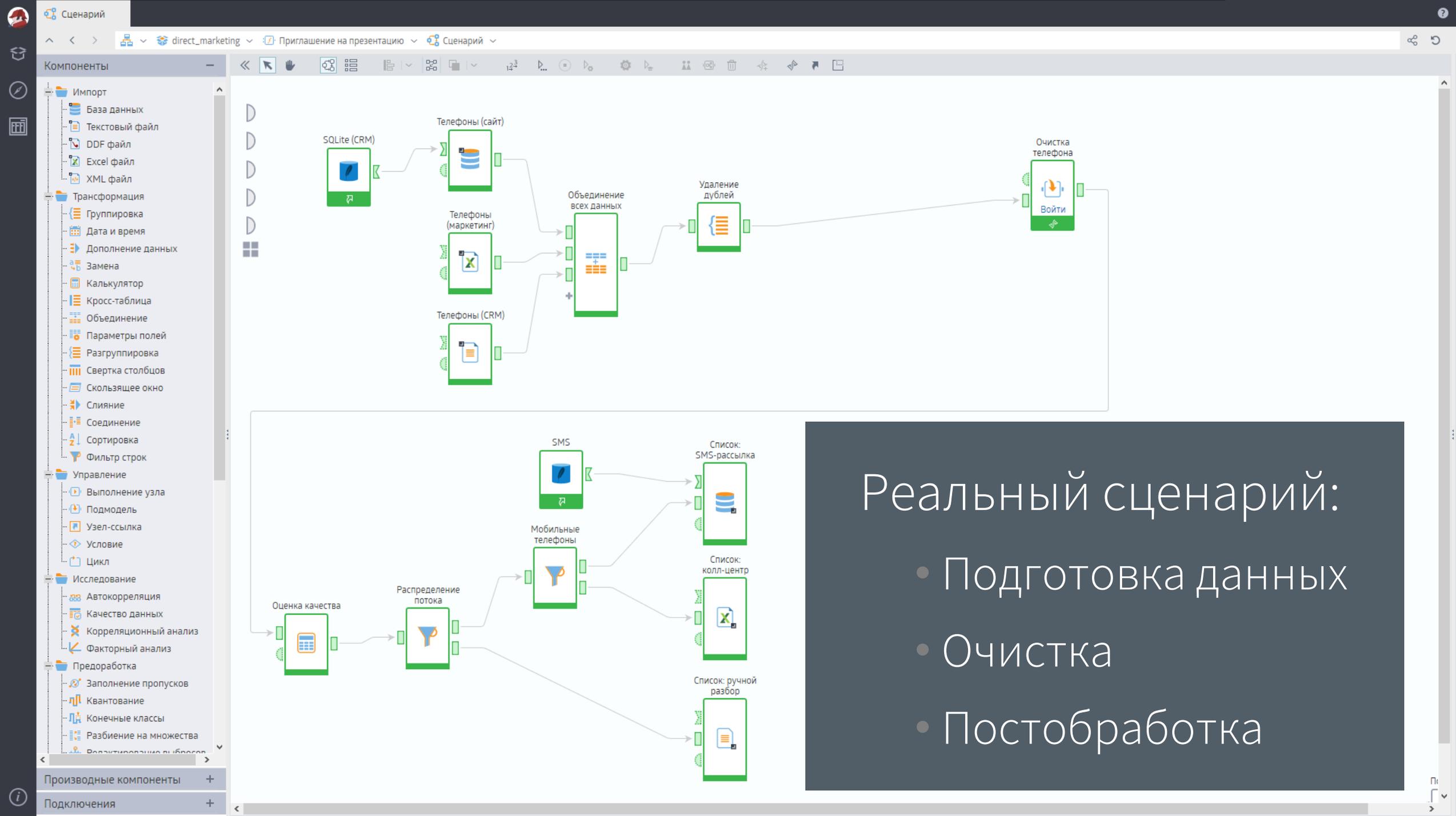
Результат обработки

Почтовый адрес – пример очистки

ab ID	252162
ab Исходный адрес	ЭНГЕЛЬС-23,УЛ.,АТКАРСКАЯ,24
ab Найденный адрес	413110, Саратовская область город Энгельс улица Аткарская д. 24
90 Доля найденных букв запроса	0,783
12 Масштаб адреса	4
ab Код КЛАДР	640000130000006
ab Ключ ФИАС	cd6746ba-1dab-4dc7-a86e-c84b71fdaf5b
ab Почтовый индекс	413110
0/1 Признак актуальности	Ложь
ab Полный адрес актуального АЭ	Саратовская Область Энгельсский Район Рабочий поселок Приволжский Улица Аткарская
ab Код КЛАДР актуального АЭ	640390000770037
ab Ключ ФИАС актуального АЭ	6272133e-3bb0-4dff-acee-8e0293387f20
ab Почтовый индекс актуально...	413110
ab Тип субъекта РФ	Область
ab Наименование субъекта РФ	Саратовская
ab Тип "города"	Город
ab Наименование "города"	Энгельс
ab Тип "улицы"	Улица
ab Наименование "улицы"	Аткарская

Исходные
данные

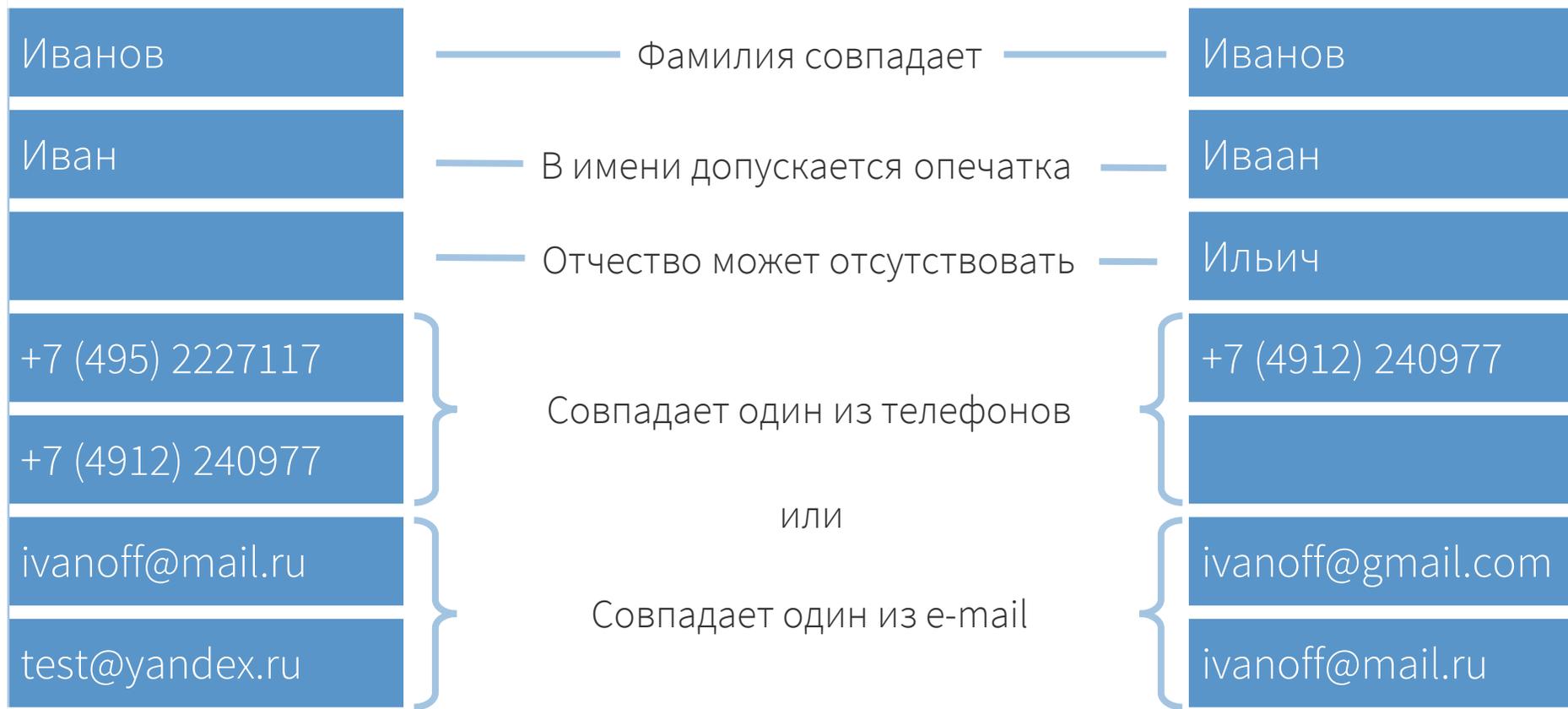
Результат
обработки



Реальный сценарий:

- Подготовка данных
- Очистка
- Постобработка

Пример стратегии дедупликации



Группа дублей

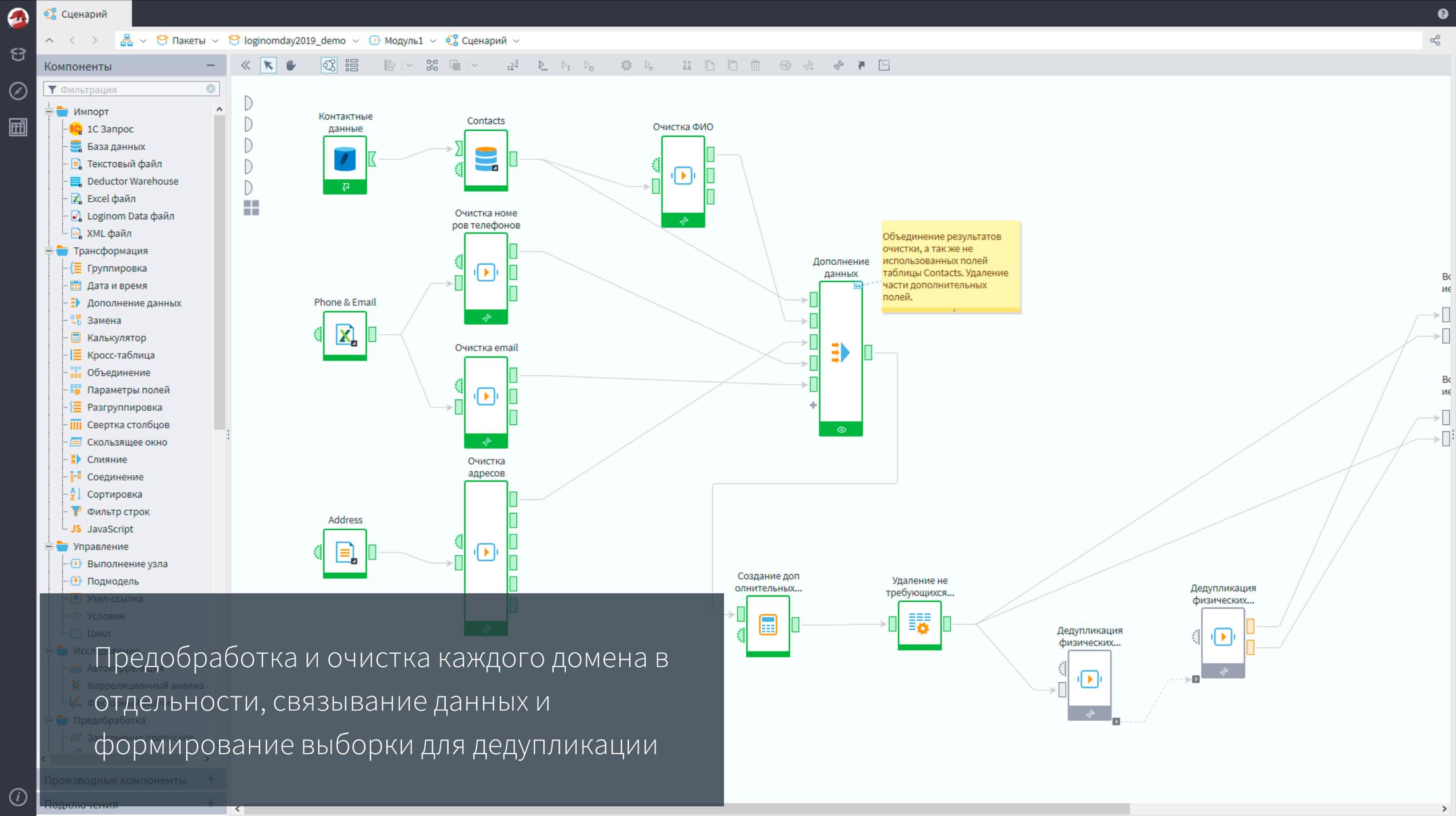
Фамилия	Имя	Отчество	Дата рождения	Телефон	Телефон	E-mail	E-mail
Иванов	Иван			+7 (495) 2227117	+7 (4912) 240977	ivanoff@mail.ru	test@yandex.ru
Иванов	Иваан	Ильич		+7 (4912) 240977		iii@yandex.ru	ivanoff@mail.ru
Иванов			15.12.1971	+7 (4912) 240699		ivanoff@mail.ru	



Единичные атрибуты	Значение
Фамилия	Иванов
Имя	Иван
Отчество	Ильич
Дата рождения	15.12.1971

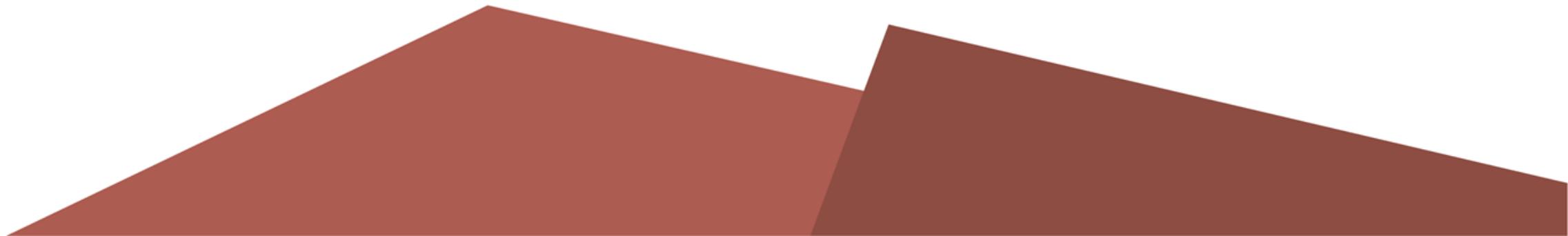
Множественные атрибуты	Значение
Телефон 1	+7 (495) 2227117
Телефон 2	+7 (4912) 240977
Телефон 3	+7 (4912) 240699
E-mail 1	ivanoff@mail.ru
E-mail 2	test@yandex.ru
E-mail 3	iii@yandex.ru

«Золотая» запись



Предобработка и очистка каждого домена в отдельности, связывание данных и формирование выборки для дедупликации

ЧТО НОВОГО



Направления развития

1. Улучшение разбора российских почтовых адресов
2. Упрощение настройки дедупликации
3. Инкрементное автообновление базы решения на стороне клиента

Очистка российских почтовых адресов

- Улучшение качества разбора данных
- Выделения уточняющих объектов (военные части, комнаты, секции, помещения...)
- Изменение логики формирования результатов очистки адресов (белая, серая и черная зона)
- Отчеты для контроля качества разбора

Очистка почтовых адресов РФ

Сырой адрес	Было	Стало
Зерноград 2а Шукшина	Ростовская область Зерноградский район город Зерноград улица им Шукшина	347740 , Ростовская область Зерноградский район город Зерноград улица им Шукшина д. 2А
450000 Республика Башкортостан УФА бульвар Т ЯНАБИ стр. д. 47 кв. 97А	-	450043 , республика Башкортостан город Уфа бульвар Тухвата Янаби д. 47, стр. А, кв. 97
196211 Г САНКТ-ПЕТЕРБУРГ ПР-КТ ЮРИЯ ГАГАРИНА д. 22 строение 1 корп. кв. 18	196211, город Санкт-Петербург проспект Юрия Гагарина	196211, город Санкт-Петербург проспект Юрия Гагарина д. 22, стр. 1, кв. 18
ул. Мира, д. 11 , офис "Адвокаты", г. Пермь, Пермский край, 614022	614022 , Пермский край город Пермь улица Мира д. 11	614095 , Пермский край город Пермь улица Мира д. 11, оф. "Адвокаты"
117452 г Москва Г. МОСКВА Б-Р. ЧОНГАРСКИЙ дом 25А корп. В/Ч кв. 01972	117452, город Москва бульвар Чонгарский д. 25А, корп. В, кв. 1972	117452, город Москва бульвар Чонгарский д. 25А, в/ч 01972

Сырой адрес	Результат очистки	Зона
СОЧИ, ул. ШОССЕЙНАЯ д. 5В кв. 25, 354037	354037, Краснодарский край город Сочи улица Шоссейная д. 5В, кв. 25	Белая
614000;Пермский край;Город ПЕРМЬ;Улица СЕМФИРОПОЛЬСКАЯ;д. 86	614089, Пермский край город Пермь улица Симферопольская д. 86	
644117 обл Омская Г. ОМСКУЛ. ЗМОЛОДЕЖНАЯ дом 62 корп. СЕК. 5 кв. 64	644117, Омская область город Омск улица Молодежная 3-я д. 62, кв. 64	
414009 Астрахань Атарбекова 17 12	414009, Астраханская область город Астрахань улица Атарбекова д. 17, кв. 12	Серая
	414009, Астраханская область город Астрахань площадь Атарбекова д. 17, кв. 12	
	414009, Астраханская область город Астрахань переулок Атарбекова д. 17, кв. 12	
край Краснодарский Г. КРАСНОДАР ПР-КТ. И-М. ПИСАТЕЛЯЗНАМЕНСКОГО дом 14 кв. 157	350065, Краснодарский край город Краснодар проспект им писателя Знаменского д. 14, кв. 157	
Самара в/ч 12454	443000, Самарская область город Самара в/ч 12454	
Калужская область Д ПЕРЕДОЛЬ улица НЕТ стр. д. 10 кв.	249160, Калужская область Жуковский район деревня Передоль д. 10	
2 а ул. Усть-Курдюмская	410018, Саратовская область город Саратов улица Усть-Курдюмская д. 2А	Черная
10А ул.м-он Солопова		
10-Я КРАСНОАРМЕЙСКАЯ,22А		
Всеволожск нет Промзона Кирпичный завод		

Настройки дедупликации

Было

Сложная, но гибкая настройка стратегий с применением любых атрибутов

Формирование произвольных стратегий дедупликации

Стало

Преднастроенные стратегии дедупликации для физических и юридических лиц

Жесткие стратегии – формирования золотых записей, мягкие – нахождение потенциальных дублей

```
{
  "options": [{
    "strategy_1": [{
      "strategy_name": "Стратегия 1",
      "condition": "group_1 AND group_2 AND group_3 AND group_4",
      "id_name": "ID",
      "manual_processing": true,
      "data": [{
        "group_1": [{
          "name_1-1": "last_name",
          "ifnull": false,
          "mistakes_max": 1,
          "percentage": 15
        }],
        "group_2": [{
          "name_2-1": "name",
          "ifnull": false,
          "mistakes_max": 0,
          "percentage": 0
        }],
        "group_3": [{
          "name_3-1": "middle_name",
          "ifnull": true,
          "mistakes_max": 0,
          "percentage": 0
        }],
        "group_4": [{
          "name_6-1": "date",
          "ifnull": false,
          "mistakes_max": 0,
          "percentage": 0
        }],
      ]
    }],
  ]
}
```



Выходные	Имя	Назначение	Значение
0/1 Стратегия 1А	Strategy1A	Не задано	true
0/1 Стратегия 2А	Strategy2A	Не задано	true
0/1 Стратегия 3А	Strategy3A	Не задано	true
0/1 Стратегия 4А	Strategy4A	Не задано	true
0/1 Стратегия 5А	Strategy5A	Не задано	true
0/1 Стратегия 1М	Strategy1M	Не задано	true
0/1 Стратегия 2М	Strategy2M	Не задано	true
0/1 Стратегия 3М	Strategy3M	Не задано	true

Изменения способа
настройки дедупликации

Преднастроенные стратегии дедупликации

Физические лица:

- 5 автоматических (жестких) стратегий
- 3 ручных (мягких)

Юридические лица:

- 5 автоматических (жестких) стратегий
- 3 ручных (мягких)

StrategyList	GroupID	ID	Phone1	Email1	FirstName	MiddleName	LastName	DocSeries	DocNumber	Address
Стратегия 2A	a-22-41-09102019-114852	22	+7 (927) 9807728	ferrumlena@mail.ru	Екатерина	Владимировна	Котова	6525	392289	
		41		ferrumena@mail.ru	Екатерина	Витальевна	Котова	6525	392289	
	a-78-79-09102019-114852	78	+7 (912) 9708511	shap314@mail.ru	Евгений	Алексеевич	Суворов	7120	464665	305007, Курская область город Курск улица Сумская д. 37, ко...
		79	+7 (951) 3264245		Евгений	Алексеевич	Суворов	7120	464665	363246, республика Северная Осетия - Алания Алагирский р...
Стратегия 4A	a-33-34-09102019-114852	33	+7 (929) 7113339	asta.aniki@gmail.com	Анастасия	Алексеевна	Тимухина	6530	588514	140055, Московская область город Котельники микрорайон ...
		34	+7 (937) 3582277	asta.aniki@gmail.com	Анастасия	Александровна	Тимухина			

Пример стратегии дедупликация по жестким стратегиям для автоматической обработки

Атрибуты	Расстояние Дамерау-Левенштейна	Непротиворечивость
Страна	0	True
Имя	2/25	False
Фамилия	2/25	False
Инициалы	0	False
Тип документа	0	False
Серия документа	0	False
Номер документа	0	False

Куб

Пакеты loginomday2019_demo Модуль1 Сценарий Корректировка имен и меток полей Визуализаторы Куб

Факты

StrategyList	GroupID	ID	Phone1	Email1	FirstName	MiddleName	LastName	BirthDate	DocSeries	DocNumber	Address	
Стратегия 2М	m-38-39-09102019-114852	38	+7 (922) 4712988	astrolesya@mail.ru	Максим	Андреевич	Руднев	18.07.1995 ...				
		39	+7 (922) 4712988		Максим	Алексеевич	Руднев		6515	752343	353201, Краснодарский край Динской район станица	
	m-65-66-09102019-114852	65	+7 (495) 2227117		Николай		Сорокин	16.04.1995 ...				
		66	+7 (495) 2227117	remstroyelit@bk.ru	Николай	Кондратьевич	Сорокин	26.09.1999 ...				
	m-91-93-09102019-114852	91	+7 (919) 3006312	stellamn287@ramble...	Дарья	Дмитриевна	Сергеева					350062, Краснодарский край город Краснодар ули
		93	+7 (919) 3006312	slvlawer@yandex.ru	Дарья	Дмитриевна	Сергеева	09.05.1978 ...	8027	711607	350062, Краснодарский край город Краснодар ули	
Стратегия 3М	m-16-25-09102019-114852	16	+7 (495) 2122365	vibra4ka@mail.ru	Ирина	Ахмедовна	Исупова	23.07.1991 ...	6527	437914	196211, город Санкт-Петербург проспект Юрия Гаг	
		25	+7 (963) 4690112	nat_74@inbox.ru	Ирина		Исупова	23.07.1991 ...	6514	326540	183032, Мурманская область город Мурманск ули	
	m-73-77-09102019-114852	73	+7 (924) 1132477		Ольга	Викторовна	Кобзева	30.12.1983 ...	6522	798871	450050, республика Башкортостан город Уфа улиц	

Пример стратегии дедупликация по мягким стратегиям для ручной обработки

Атрибуты	Расстояние Дамерау-Левенштейна	Непротиворечивость
Фамилия	0	False
Инициалы отчества	0	False
Телефоны (Phone1, Phone2, Phone3)	0	False

Пример формирования мастер-записи

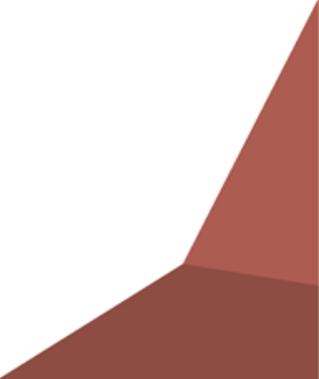
- Единичные атрибуты (ФИО, Адрес, паспорт...): информация из первой записи. Если поле пустое, то из следующей записи.
- Множественные атрибуты (телефоны, e-mail): Формируются уникальные значения, выбираются первые три записи, остальные отбрасываются.

Название	Значение
#	5
ab ID	78
ab GroupID	a-78-79-09102019-114852
ab StrategyList	Стратегия 2А
ab LastName	Суворов
ab FirstName	Евгений
ab MiddleName	Алексеевич
ab FirstNameIni...	Е
ab MiddleNameIni...	А
ab Email1	shap314@mail.ru
ab Email2	
ab Email3	
ab Phone1	+7 (912) 9708511
ab Phone2	
ab Phone3	
ab DocType	Паспорт
ab DocSeries	7120
ab DocNumber	464665
31 BirthDate	01 Октябрь 1993
ab BirthYear	1993
ab Country	РФ
ab Address	305007, Курская область город Курск улица Сумская д. 37, ко...
31 CreationDate	27.02.2015 18:49:38
31 UpdateDate	31.07.2015 18:49:38

Название	Значение
#	6
ab ID	79
ab GroupID	a-78-79-09102019-114852
ab StrategyList	Стратегия 2А
ab LastName	Суворов
ab FirstName	Евгений
ab MiddleName	Алексеевич
ab FirstNameIni...	Е
ab MiddleNameIni...	А
ab Email1	
ab Email2	
ab Email3	
ab Phone1	+7 (951) 3264245
ab Phone2	
ab Phone3	
ab DocType	Паспорт
ab DocSeries	7120
ab DocNumber	464665
31 BirthDate	22 Январь 1985
ab BirthYear	1985
ab Country	РФ
ab Address	363246, республика Северная Осетия - Алания Алагирский ра...
31 CreationDate	10.05.2009 18:49:38
31 UpdateDate	07.04.2011 18:49:38

Название	Значение
#	1
ab ID	79
ab GroupID	a-78-79-09102019-114852
ab LastName	Суворов
ab FirstName	Евгений
ab MiddleName	Алексеевич
ab FirstNameIni...	Е
ab MiddleNameIni...	А
ab Email1	shap314@mail.ru
ab Email2	
ab Email3	
ab Phone1	+7 (951) 3264245
ab Phone2	+7 (912) 9708511
ab Phone3	
ab DocType	Паспорт
ab DocSeries	7120
ab DocNumber	464665
31 BirthDate	22 Январь 1985
ab BirthYear	1985
ab Country	РФ
ab Address	363246, республика Северная Осетия - Алания Алагирский рай...
31 UpdateDate	07.04.2011 18:49:38
31 CreationDate	10.05.2009 18:49:38

Инкрементное автообновление

1. Обращение к веб-сервису для получения обновления
 2. Определение разницы между базами клиента и последней версии LDQ
 3. Формирование diff и передача клиенту
 4. Накатывание изменений клиенту
- 

Выгоды инкрементного обновления

1. Возможность клиенту выбрать комфортный режим обновления базы
 2. Снижение более чем в 2000 раз объемов передаваемых данных
 3. Обновление баз данных в 6 раз быстрее
- 

loginom.ru