

Отечественный self-service Data Preparation

Как переехать без компромиссов

Станислав Уколкин, старший менеджер данных T2

ДРУГИЕ
ПРАВИЛА.
НОВЫЙ
УРОВЕНЬ



24.10.24

Предыстория

Tableau Prep



Уход Tableau

Tableau ушёл с рынка, а вместе с ним пропала возможность продлять лицензию и на Tableau Prep



Миграция

Возникла потребность в миграции текущих процессов на альтернативное решение



Подбор решений

Отобрали наиболее подходящие инструменты с т.з. функциональности и UX.



Пилоты

Провели с текущими пользователями Tableau Prep пилотные реализации их бизнес-процессов на инструментах, набравших наибольшее кол-во баллов в скоринг-таблице целевых решений



Выбор

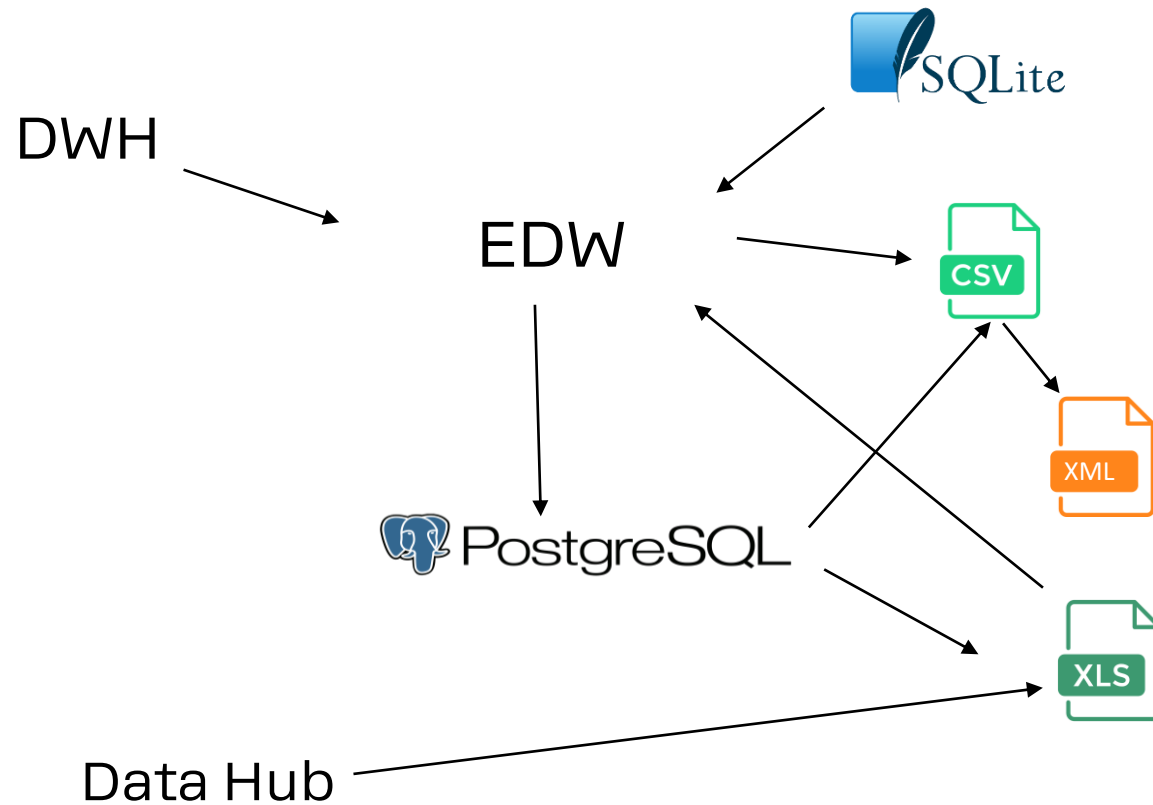
Выбрали Loginom как наиболее близкий по функциональности к Tableau Prep и собравший наибольшее кол-во позитивной обратной связи от пользователей



- Tableau Prep был включен по умолчанию в поставку Tableau
- Простой и понятный интерфейс для подготовки данных к визуализации
- Возможность сбора данных из хранилищ и других систем в корп.контуре
- Функциональность работы с CSV, Excel, txt файлами и выгрузка в них

Что такое self-service data preparation

Инфраструктура пользователя Data Office T2

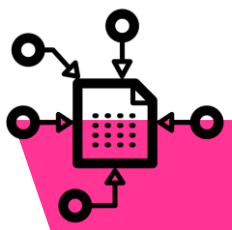


- ! У каждого пользователя уникальный набор источников и процесс подготовки данных
- Очередь на разработку
- Нулевая управляемость процессом со стороны DO
- Трудоемкость описания текущего бизнес-процесса



Self-service data preparation

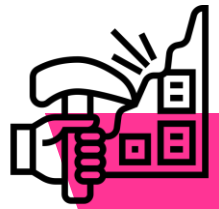
глазами пользователя



Загрузка данных из БД, файлов и веб-сервисов

Сбор данных

Также можно сгенерировать данные



Визуализация, статистический анализ, поиск аномалий

Исследование данных

Оценка пригодности, качества и структуры исходных данных



Удаление дубликатов, корректировка форматов

Очистка и проверка

Очистка данных от разного рода «мусора»



Преобразование к целевому формату, добавление справочников

Трансформация и обогащение

Улучшение качества и повышение ценности данных



Выгрузка данных в БД, файл или веб-сервис

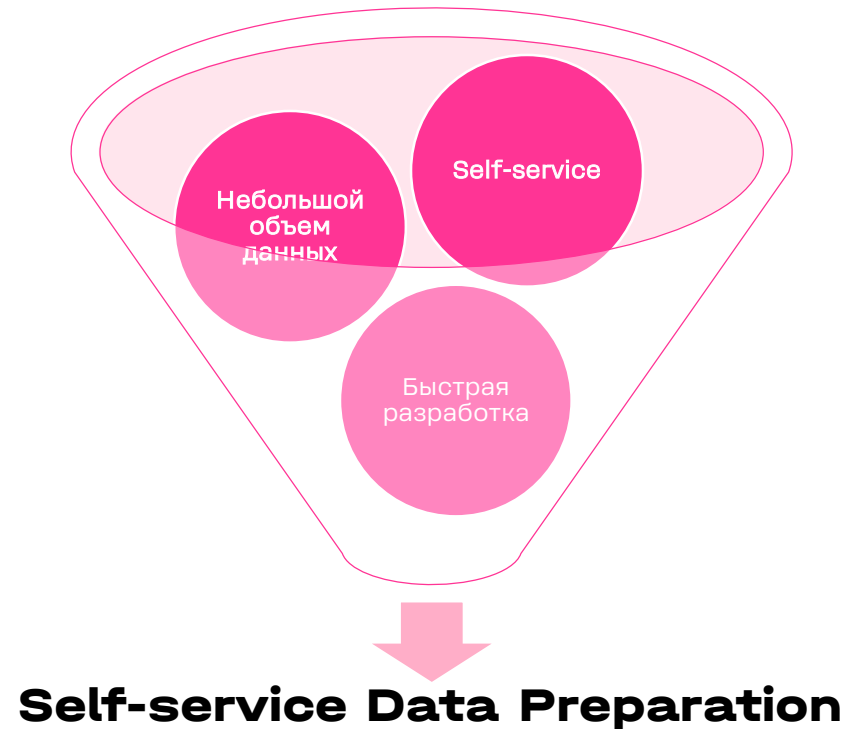
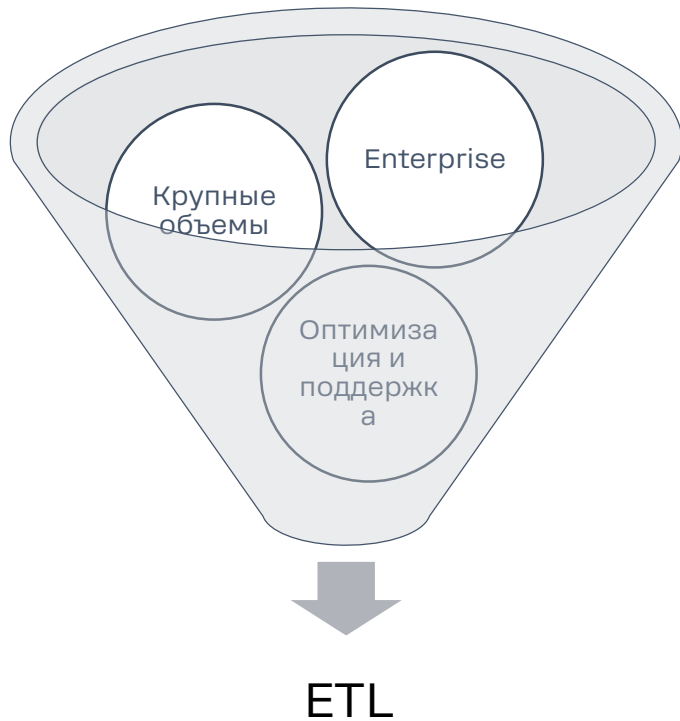
Проверка и публикация

Результаты работы публикуются или выгружаются для дальнейшего использования



Что такое self-service data preparation

Отличие от ETL (Informatica, Nifi и пр.)



Решили не просто найти замену, а создать новый процесс и дать пользователям больше возможностей для работы с данными

Требования к инструменту

- Кросс-системная работа с данными
- Нет требования к знанию SQL
- Простота и удобство использования
- Быстрая проверка гипотез на основе данных
- Интерфейс планирования запусков

Выбор целевого решения



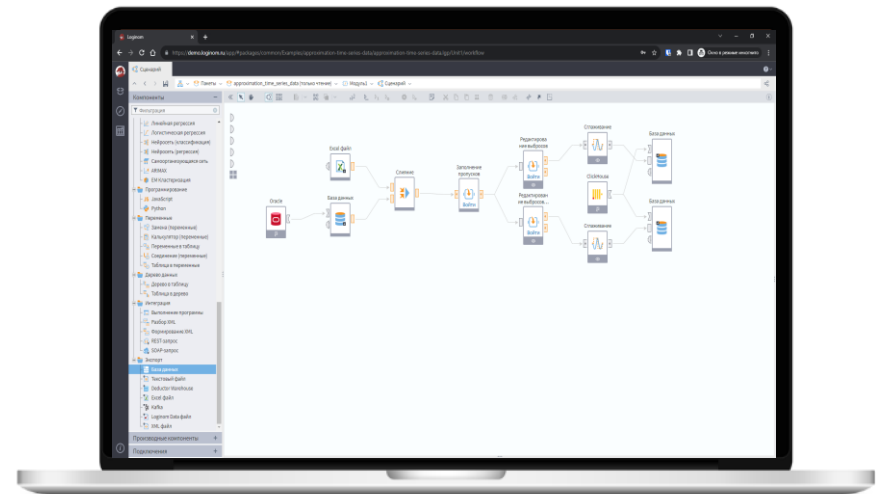
84,71%

ИТОГ

1. Общие	8,5
2. Интеграция	18,00
3. Преобразование данных	4,00
4. Управление потоком данных	7,00
5. Производительность	6,00
6. Экспорт данных и API	12,50
7. Администрирование	5,00
8. Простота внедрения	9,50
9. Дополнительно	6,50
10. Общий балл	68,50

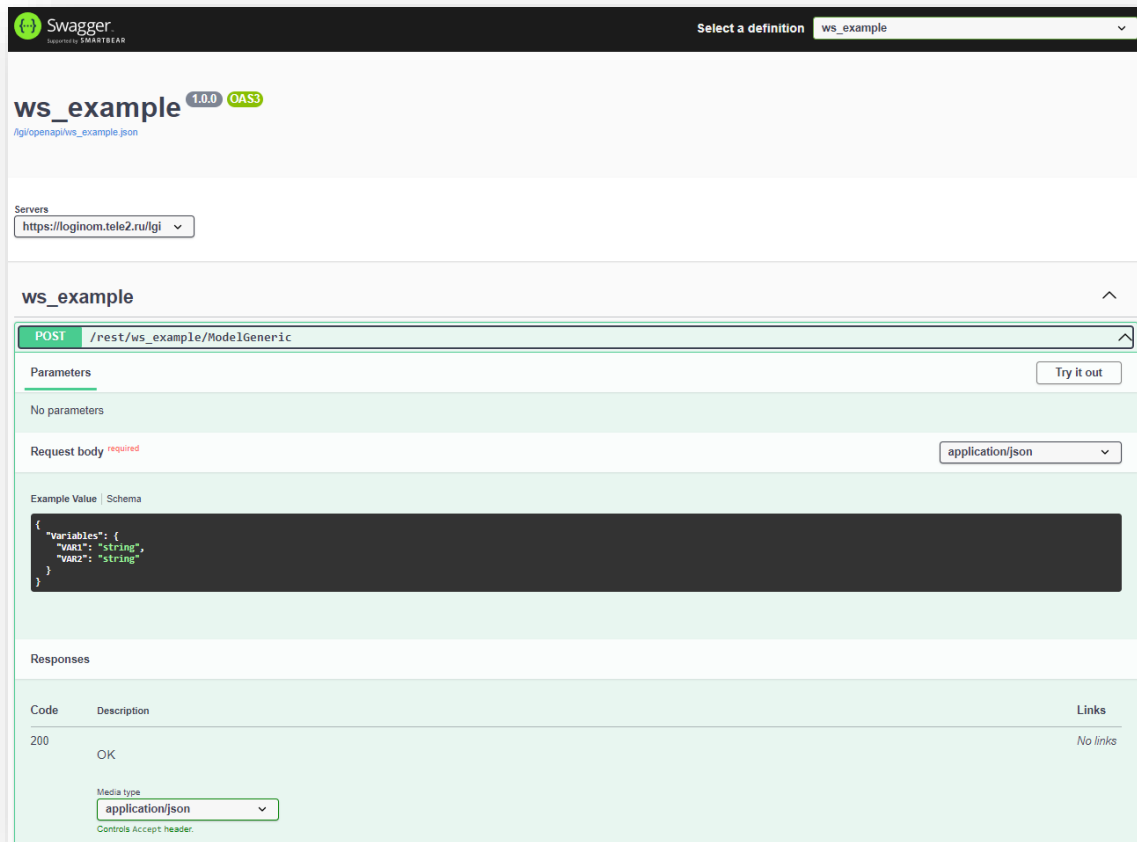


- **Покрывает функциональность Preq**
(умеет работать с основными корпоративными БД и файлами, удобный интерфейс)
- **Быстрая реакция поддержки**
(стабильно быстрые и качественные ответы по тикетам даже в рамках пилота)
- **Качественное сопровождение**
(каталог курсов, сообщество, демо-примеры и мероприятия)
- **Аккредитованное ПО**
зарегистрирован в реестре отечественного ПО



Logiном как web-сервис

Запуск по событию через API



- **Максимальная гибкость**
Встраивание логики Logiном в любой из шагов бизнес-процесса
- **Запуски по событию**
Возможность запустить пакеты Logiном из пользовательских скриптов
- **Передача данных для ИТ**
Альтернатива передаче файлов и данных для выгрузки через сетевые папки для ИТ

Дополнительная функциональность

Инструмент оркестрации данных - Cronicle

Plugin: Shell Script

Plugin Parameters

Script Source

```
#!/bin/sh  
# Enter your shell script code here  
/opt/loginom/server/BatchLauncher -Package=/stanislav.ukolkin/test_acc.lgp
```

Add Date/Time Stamps to Log

Interpret JSON in Output

Select the plugin parameters for the event.

Timing: Daily

Timezone: Europe/Moscow

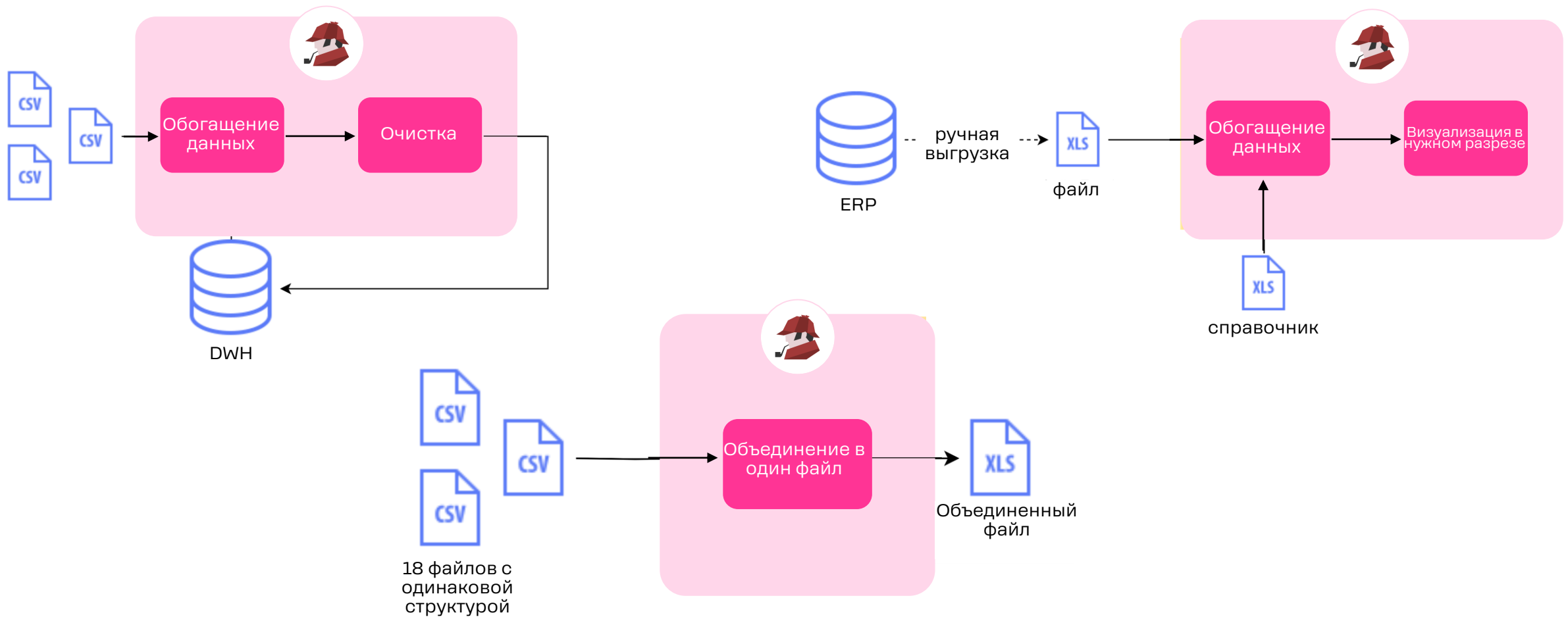
On Demand
Custom
Yearly
Monthly
Weekly
Daily
Hourly

6PM

The event will run:
Daily at 6:00pm

Job ID	Event Name	Category	Plugin	Hostname	Result	Start Date/Time	Elapsed Time
ItstZach00	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Sep 6, 2023 7:00 PM	0 sec
ItstZach01	О Дневная отчетная Листинг	Lognom_packages	Shell Script	Chu-igpnom-01	Success	Aug 14, 2023 4:15 PM	30 sec
ItstZach02	О Дневная отчетная Листинг	Lognom_packages	Shell Script	Chu-igpnom-01	Error	Aug 14, 2023 10:20 AM	0 sec
ItstZach03	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 14, 2023 10:17 AM	0 sec
ItstZach07	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 12:04 PM	0 sec
ItstZach04	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 7:14 PM	0 sec
ItstZach05	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 7:11 PM	0 sec
ItstZach01	О Дневная отчетная Листинг	Lognom_packages	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 7:10 PM	0 sec
ItstZach03	О Дневная отчетная Листинг	Lognom_packages	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 7:09 PM	0 sec
ItstZach02	О Дневная отчетная Листинг	Lognom_packages	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 7:08 PM	0 sec
ItstZach01	О Дневная отчетная Листинг	Lognom_packages	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:21 PM	0 sec
ItstZach04	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:00 PM	0 sec
ItstZach05	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:00 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 3:58 PM	0 sec
ItstZach03	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:58 PM	0 sec
ItstZach02	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:57 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:54 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 4:51 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 3:25 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 3:20 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 3:43 PM	0 sec
ItstZach02	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 3:41 PM	0 sec
ItstZach01	О SMTP test	General	Shell Script	Chu-igpnom-01	Success	Aug 15, 2023 3:38 PM	0 sec

Примеры пользовательских процессов



Эффекты от использования

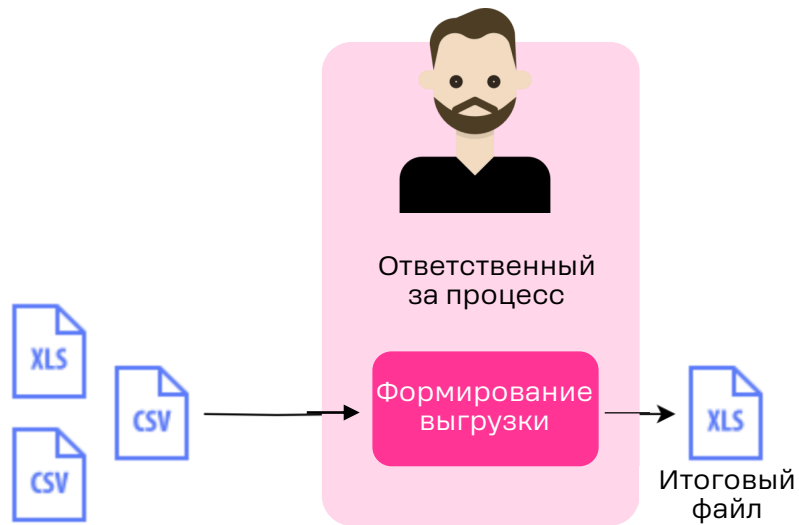
Бизнес-кейс

Из нескольких файловых источников на регулярной основе формируются данные для крупного бизнес-подразделения компании, которое занимается сопровождением

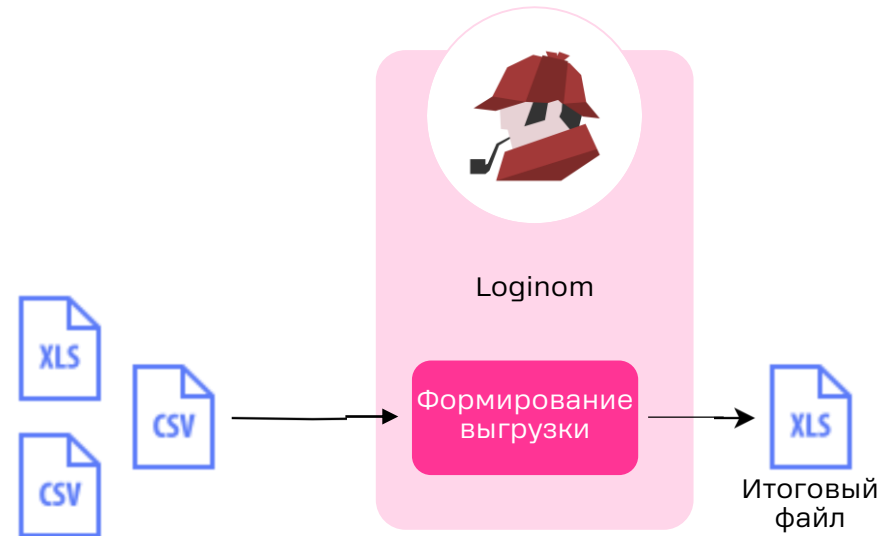
Времязатраты на реализацию процесса сократились:

1 раб.день → **6** секунд

БЫЛО



СТАЛО



Вам нужен self-service data preparation



Есть регулярный, постоянно меняющийся **процесс по работе с данными в нескольких источниках**



Есть потребность в проверке **гипотез на данных**



Есть необходимость оперативно **подготовить данные по специфичной бизнес-логике**



Спасибо за внимание!



Станислав Уколкин
stanislav.ukolkin@t2.ru



24.10.24