

# Комплексная интеграция: от разрозненных данных к прозрачности бизнеса



**Сергей Парахин**, руководитель отдела бизнес анализа и операционной эффективности

Опыт работы в аналитике: 10+ лет

Достижения:

- Собрал BI-систему для визуализации данных на базе Visiology (Positive Technologies)
- Внедрил систему управления маржинальностью клиентов на базе SAP BDMA (Tele2)



**Игорь Колупатин**, руководитель отдела развития автоматизированных систем отображения данных

Опыт работы в аналитике : 10+ лет

Достижения:

- Внедрил ETL+DWH для анализа и визуализации данных (Positive Technologies)
- Внедрил систему поддержки принятия решений для кредитного конвейера (Автокредитование) на системе Loginom

# Positive Technologies – разработчик продуктов в области кибербезопасности и единственная компания отрасли на Московской бирже

## 22

года экспертизы

## 400+

партнеров

## +76%

рост продаж за 2023 год

## 3000+

человек в команде

## 200+

тысяч акционеров

## x4

рост котировок акций  
с момента размещения

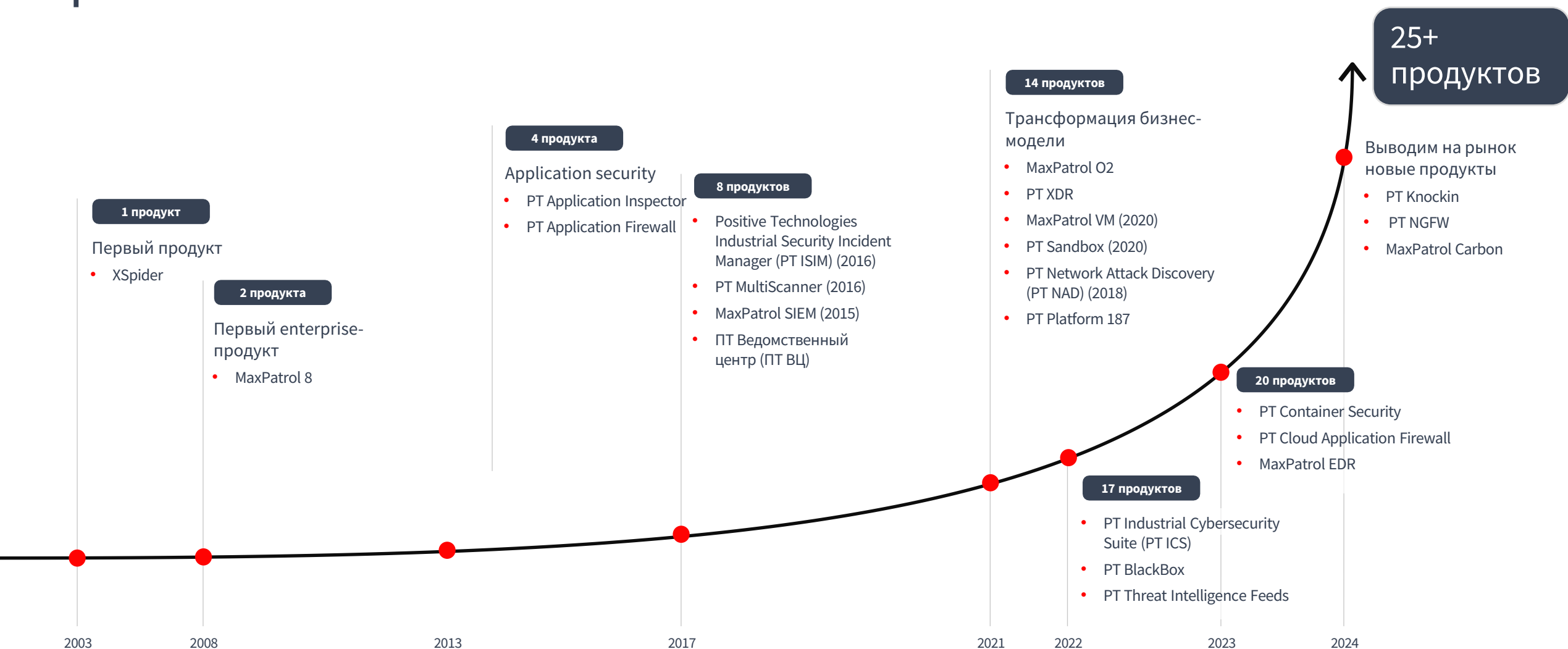
## 200

млрд рублей —  
капитализация

## ТОП-50

компаний  
на Московской бирже

# Самый широкий продуктовый портфель среди российских ИБ-компаний



# Предпосылки для создания единого хранилища и BI-системы



- Расширение клиентской базы, запрос на детальный анализ воронки продаж
- Кратный рост портфеля продуктов
- Развитие внутренних систем и сервисов
- Увеличение штата сотрудников и усложнение структуры организации
- Разрозненность источников и низкое качество данных в них

**Как мы организовали  
сбор и хранение данных?**

# Выбор решения и архитектуры



## Почему выбрали Loginom?

1. Apache AirFlow
2. Apache Nifi
3. Loginom



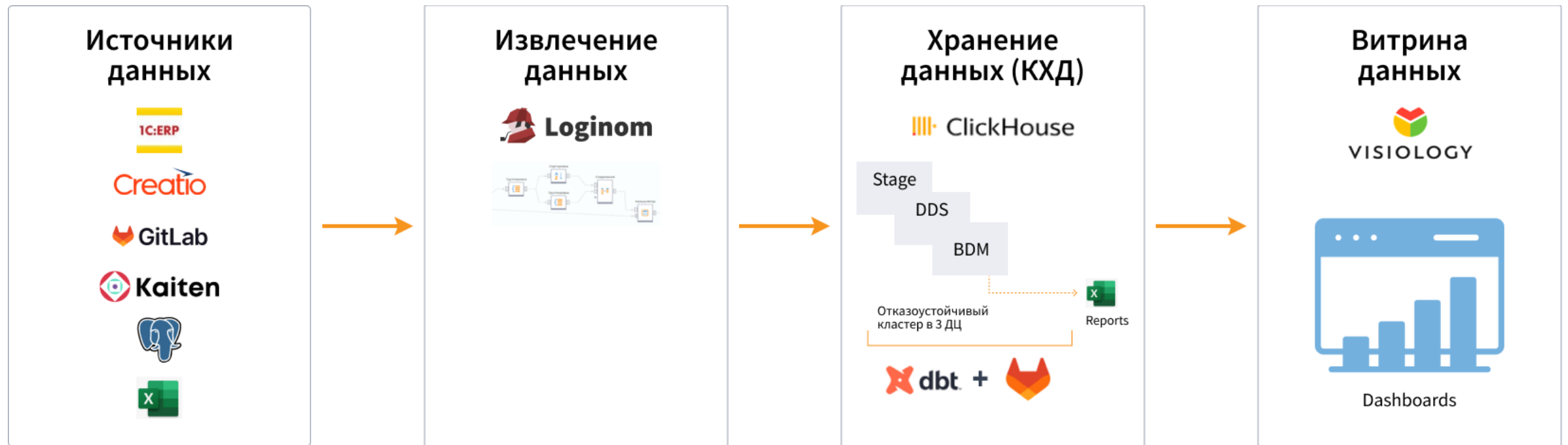
- Российский вендор – нет санкционных рисков
- Поддержка и документация на русском языке
- Продукт развивается: патч версии каждые 1.5 месяца, новые версии раз в год
- Удобный пользовательский интерфейс
- Low-code платформа

**Чего мы добились за 2 года?**



# Архитектура потока данных на 2024 год (EtLT подход)

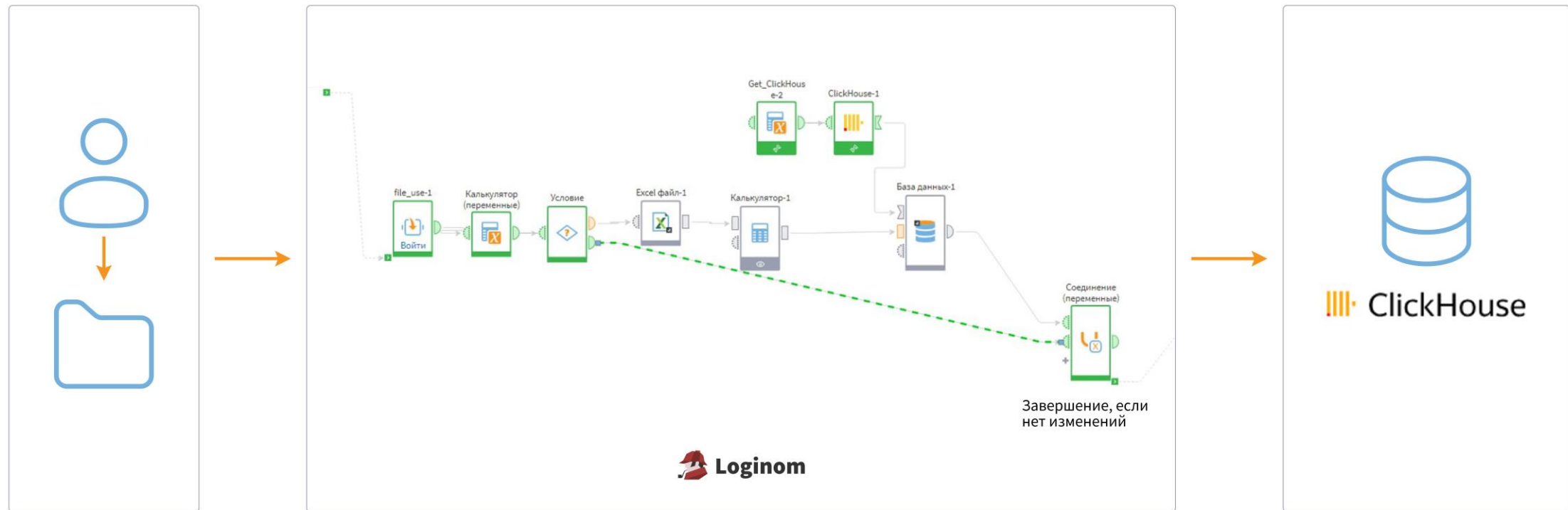
- Loginom – загрузка разрозненных источников с незначительной трансформацией (EtLT)
- DBT – трансформации данных внутри хранилища (T)



# Примеры использования Loginom

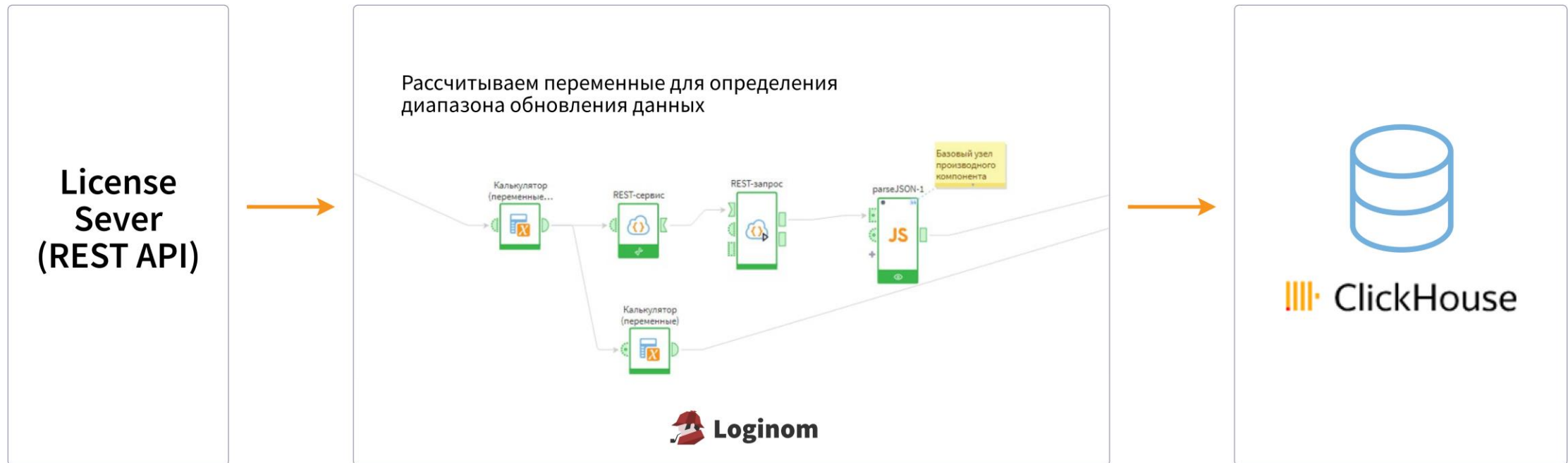
# Оркестрация загрузки ручных справочников и таблиц

- Проверка формата таблиц
- Проверка изменений
- Добавление метаданных



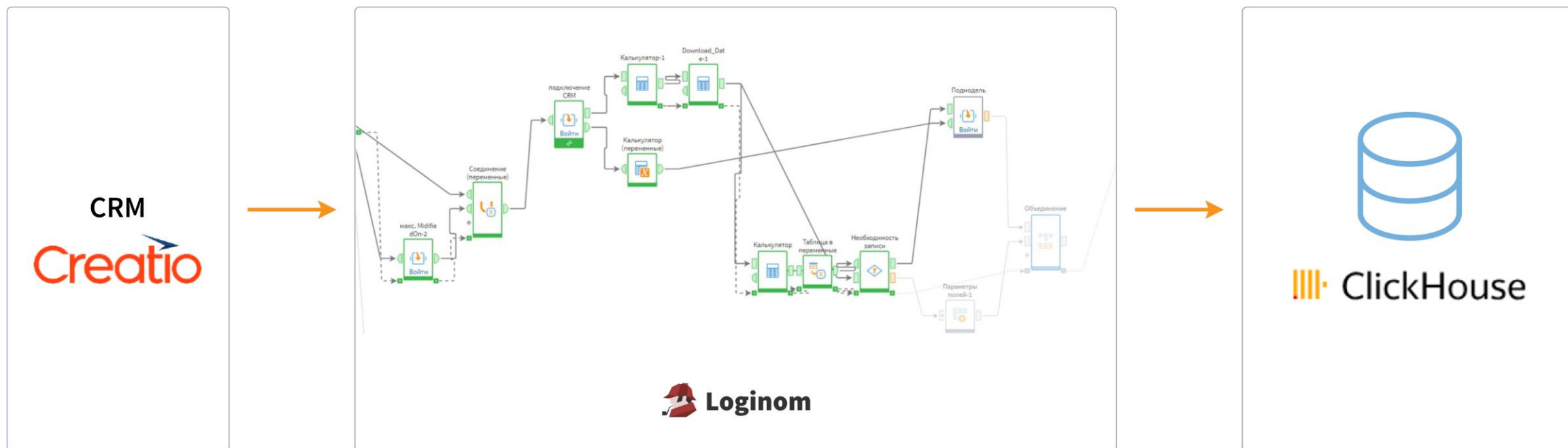
# Загрузка данных из сервера лицензирования (REST API)

- REST – запрос
- Парсинг JSON в плоскую таблицу
- Инкрементальная загрузка данных



# Загрузка данных из CRM (база данных MS SQL)

- Инкрементальная загрузка
- Добавление метаданных о загрузке
- Очистка ошибочных дат (DQ)



# Итоги проекта в цифрах



**~20**

источников с использованием разных коннекторов

**1 день**

подключение источника (ТТМ)

**DBT**

сборка потока данных внутри ХД на базе DBT

**30+**

дашбордов для функций

**150**

активных пользователей BI-системы в месяц

**2**

области на КХД для поставки данных в другие ИС

# Качественные результаты



Низкое качество данных в системах



- Подключили внешние справочники DaData и тд.
- Доработали процессы внутри систем - источников
- Назначили ответственных за данные

Много ручных источников (Excel)



- Автоматизировали загрузку из файлов по расписанию
- Создали инструмент валидации файлов
- Назначили ответственных за загрузку

Единая методология расчета метрик



- Согласовали методику расчета с ответственными и опубликовали во внутренней Wiki
- Разрабатываем переход на DataHub + Business Glossary

Стабилизация работы Pipeline поставки данных до BI



- Поставили подключение источников на поток
- Подключили DBT для трансформации данных внутри хранилища
- Подключили GIT для контроля версий

Удобство работы BI для пользователя



- Подключили доменную авторизацию с разграничением доступа
- Создали портал навигации по дашбордам
- Провели обучение для сотрудников

# Планы на будущее



## **Data Quality – продолжаем развивать:**

- Проверка работы «шедулера» с уведомлением data-инженеров (Telegram)
- Работа с библиотеками качества DBT (Great Expectations)
- Делегирование ответственности за качество на владельцев данных в источниках

## **Внедрение CI процесса для EtLT:**

- Определиться с инструментом для контроля версий для Loginom

## **Расширение пула сотрудников – пользователей аналитического хранилища:**

- Настройка областей в Хранилище для гибкого доступа сотрудников
- Описание таблиц и витрин в хранилище (DataHub)
- Проведение обучения аналитиков из функций
- Описание Data Lineage для удобства пользователей



# Команда проекта



## Бизнес-анализ + BI

---

- 7 сотрудников
- Сбор потребностей бизнеса
- Поиск и исследование источников и описание методики расчета метрик
- Построение модели данных ViQube
- Определение ответственных
- Построение пилотных визуализаций и self-service отчетов
- Продвижение data-driven культуры

## Дата-инженеры

---

- 4 сотрудника
- Работа с БФТ и поиск вариантов передачи данных
- Построение архитектуры КХД
- Настройка ETL-процессов, их поддержка и оркестрация
- Развитие модели данных внутри слоев КХД на базе DBT

## ИТ-администратор

---

- 1 сотрудник
- Установка КХД и BI-системы
- Поддержка работы и связанности систем
- Безопасность внутри ИБ-компании
- Обновление систем до актуальных версий

**Спасибо за внимание**