

TData

Управление данными

RT.DataGovernance

Содержание

01

О компании TData

02

Управление данными
с помощью
RT.DataGovernance

03

Реализованные
проекты

TData. Надежные решения для работы с данными

2

ключевой игрок на рынке
платформы Hadoop в России

30+

ведущих интеграторов
России — наши партнеры

8

продуктов, охватывающих
полный цикл работы с данными

5+

лет опыта
в отрасли

ТОП-5

общего рынка СУБД и средств
управления данными в России

ТОП-3

российского рынка
аналитических СУБД

Наши клиенты



ОДК

ЦИФРОВЫЕ
ТЕХНОЛОГИИ

AstraZeneca



ЕВРАЗ



TATNEFT



ШЕРЕМЕТЬЕВО
МЕЖДУНАРОДНЫЙ АЭРОПОРТ



Миранда



Ростелеком

Платформа управления данными TData

Источники данных

JDBC / ODBC

Events

API

Log

IoT

Бизнес-системы

Информационные продукты и сервисы

Bi, аналитика и отчетность

RT.DataVision
Superset + Dremio

Транспорт и преобразование данных

Набор продуктов для потоковой и пакетной обработки данных

RT.Streaming
Kafka NiFi Airflow

Хранение данных

Линейка продуктов хранения данных и аналитики

RT.DataLake
Hadoop

RT.KeyValue
Cassandra

RT.Warehouse
Greenplum

RT.WideStore
ClickHouse

Управление данными и HСИ

Бизнес-гlossарий, единая точка доступа к данным. Управление HСИ

RT.DataGovernance
Data Governance and Data Catalog

RT.MDM
Master data management

Управление и обслуживание

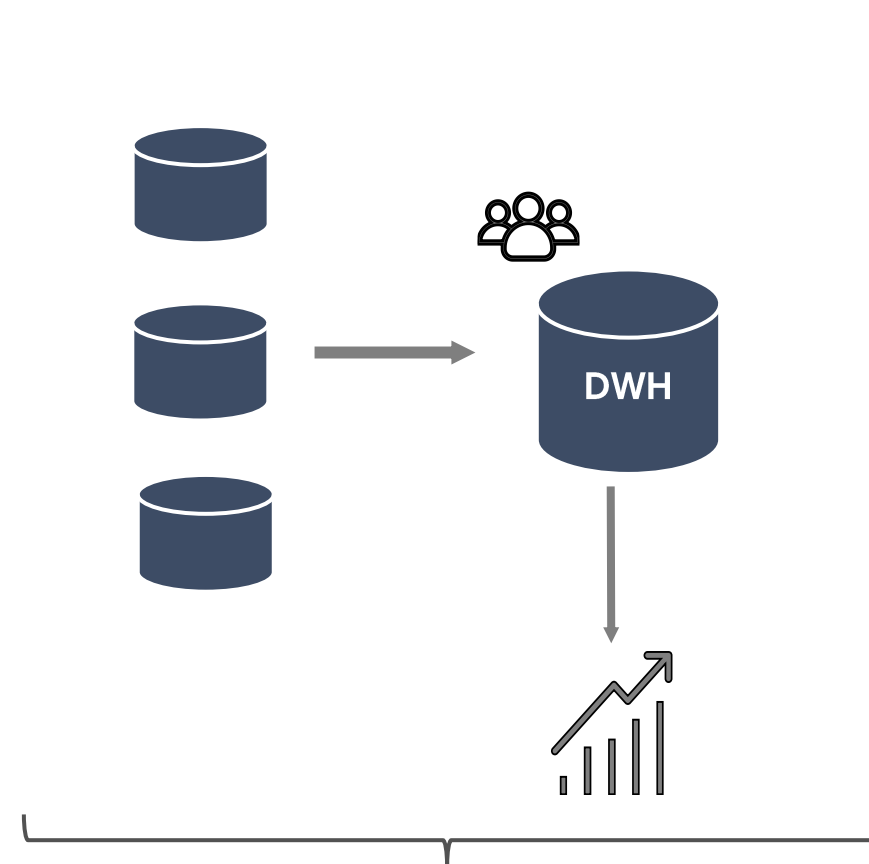
Развертывание, конфигурация и мониторинг компонентов

RT.ClusterManager

Почему необходимо управление данными?

Чем сложнее устроен аналитический ландшафт организации, тем сложнее его контролировать. Неструктурированные/слабоструктурированные данные, недостаточность консистентности, отсутствие процессов взаимодействия и доверия – основные причины появления проблем с данными, что в дальнейшем сказывается на эффективности работы самой организации.

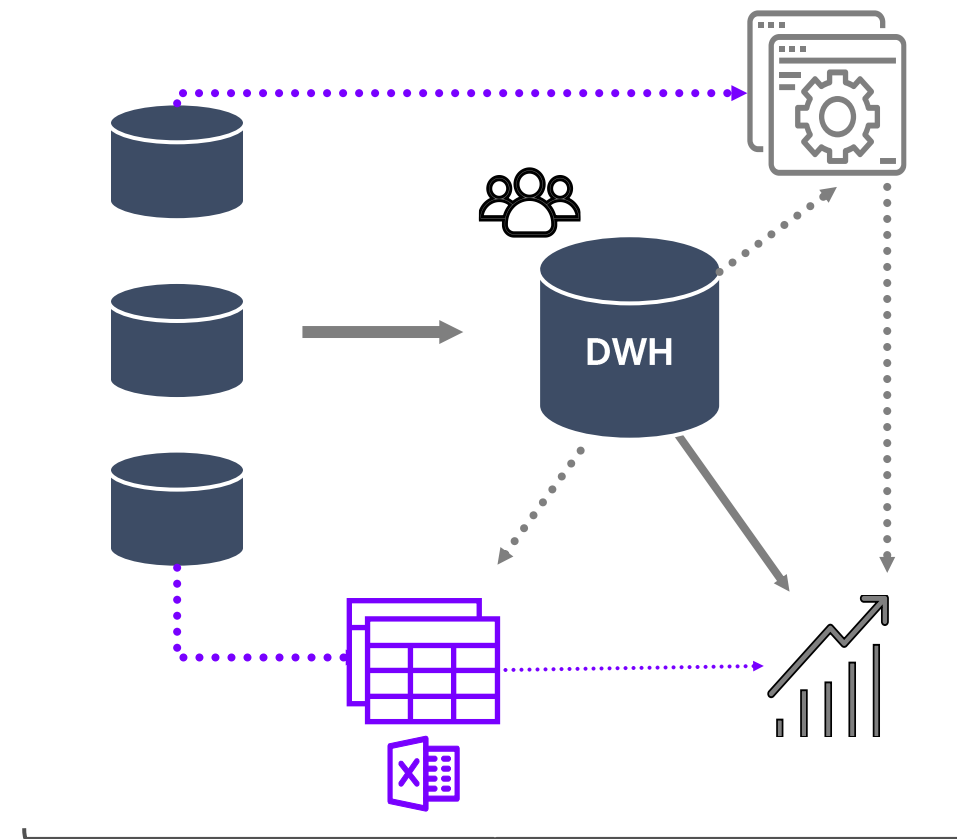
Простой аналитический ландшафт



- Единый источник или платформа данных для всех потребителей
- Данные ведутся одной командой, которая напрямую взаимодействует с владельцами источников данных

↓ Потребность в специализированных процессах управления небольшая

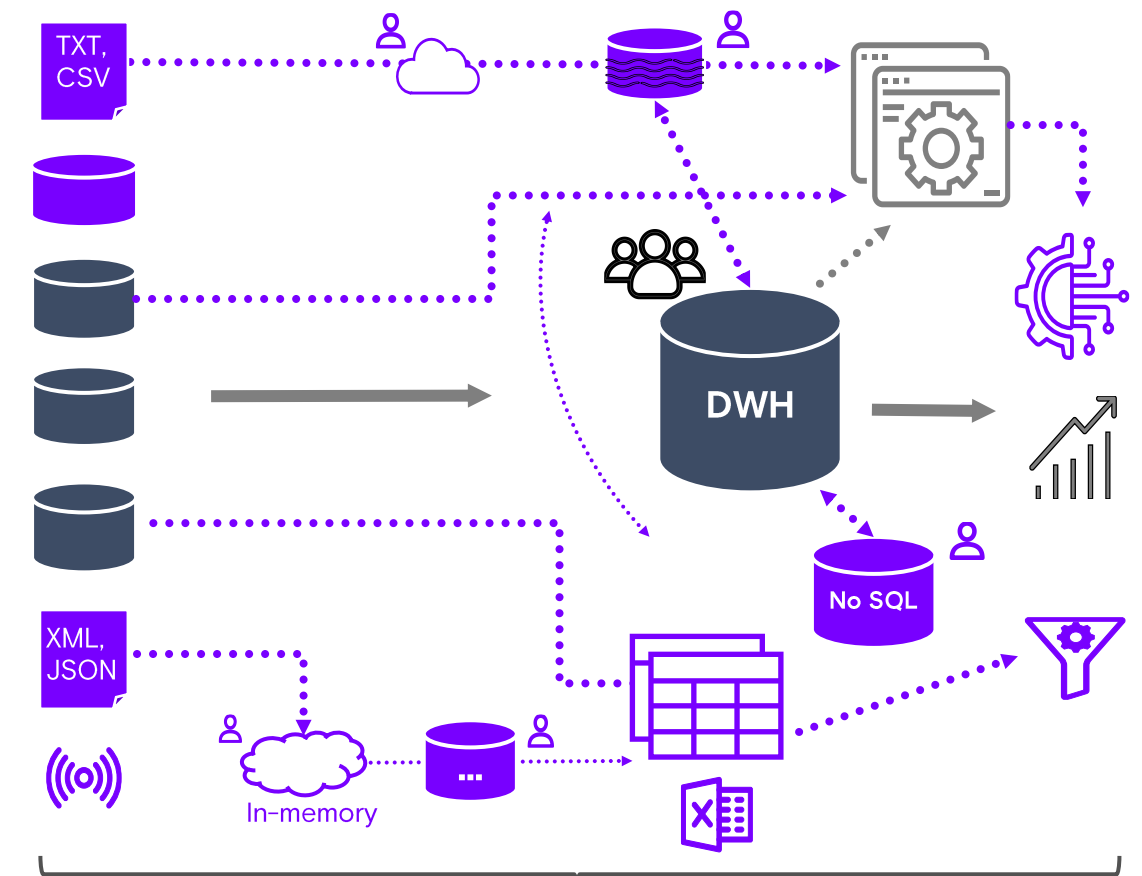
Развитие ландшафта



- Выделенные аналитические решения приводят к появлению промежуточных, часто дублирующих информацию хранилищ
- Пост-обработка и неявное использование данных в обход основных источников (дополнительные выгрузки в Excel, отчеты, csv файлы)

↑ Возрастают требования к процессам управления данными, в первую очередь к качеству данных

Ландшафт крупной компании



- Отдельные платформы для обработки и хранения неструктурированных и слабоструктурированных данных, real-time и облачных платформ, озера данных
- Self-service аналитика
- Данные находятся под контролем разных команд

↑↑↑ Выстроенные процессы управления данными становятся ключевым фактором для эффективного использования данных

Сложности в управлении данными

- Низкое качество данных не позволяет создавать достоверные отчеты
- На поиск необходимых данных в компании уходит много времени
- Отсутствие информации о владельцах объектов данных или ответственных за них
- Сложность обеспечения актуальности и доступности данных
- Необходимость замены решения иностранного вендора на отечественное ПО
- Могут быть не определены политики и процедуры управления данными



Количественные предпосылки внедрения Data Governance

Исследование Gartner

70–80%

Всего времени аналитиков может тратиться **на приведение данных в порядок**

От 15 до 35%

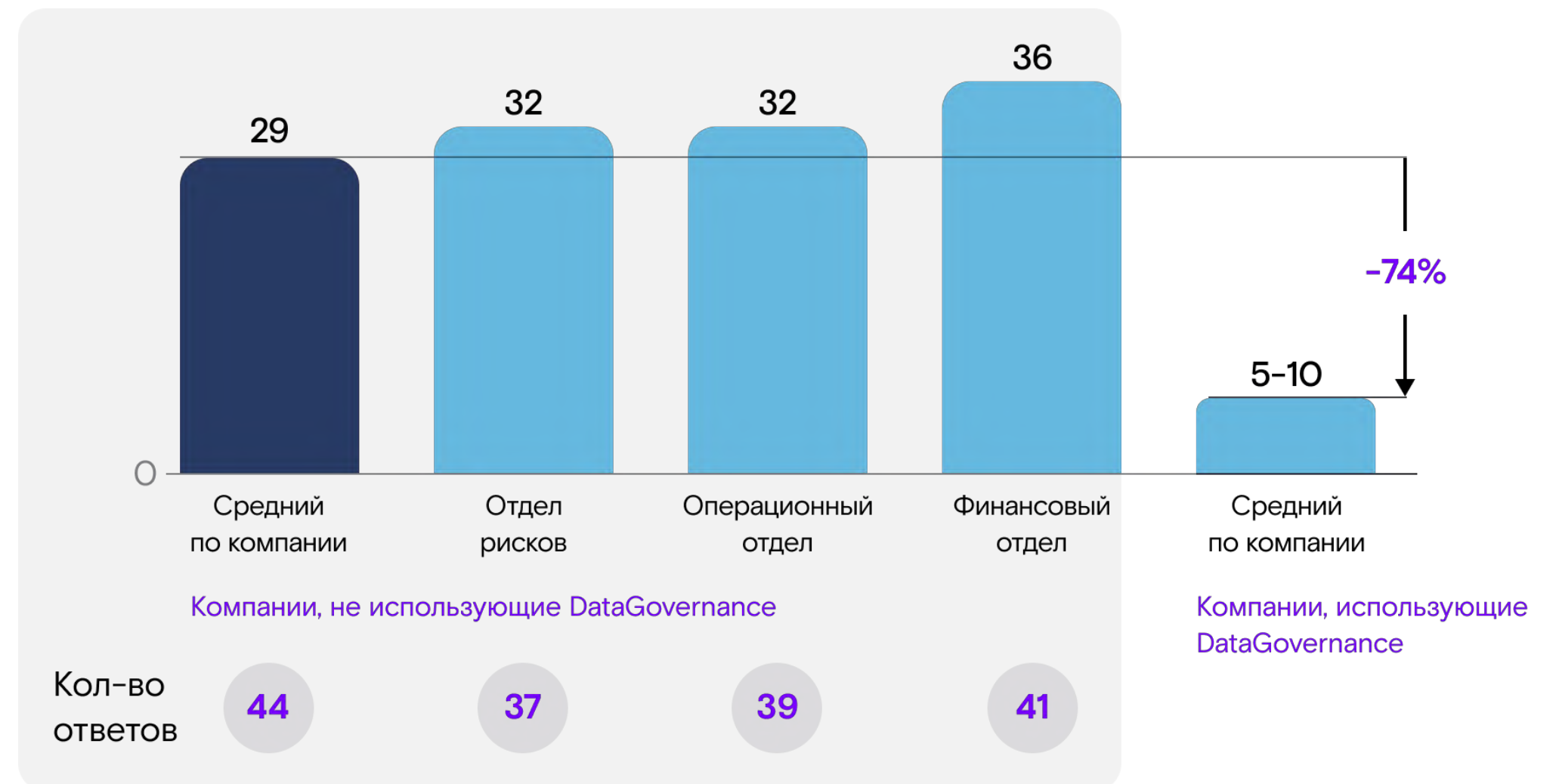
Годового бюджета организации тратится **неэффективно из-за низкого качества** данных. При этом в организациях, ориентированных на предоставление услуг (например, банки, страховые компании), **объем потерь доходит до 40%**

Исследование McKinsey¹

Недостаток качества и доступности данных может привести к тому, что сотрудники будут тратить **значительное количество** времени на задачи, которые **не добавляют** экономической ценности компании

Рис. 1: Приблизительный % от общего времени сотрудников

Время, потраченное на задачи, не добавляющие экономической ценности по причине низкого качества и доступности данных



¹. Data sourcing, data aggregation, data reconciliation, data cleansing, manual reporting, etc. Source: McKinsey Global Data Transformation Survey, 2019

Развитие системы управления данными

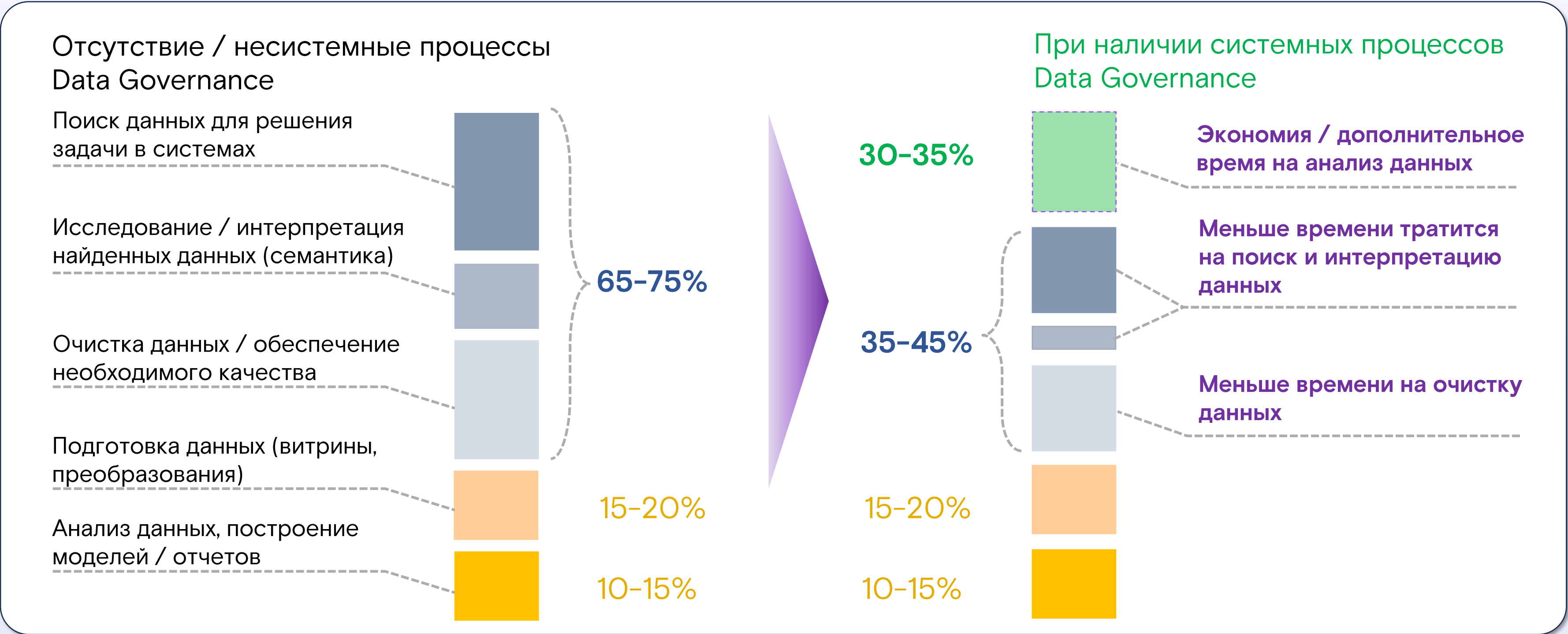
Для успешного функционирования система управления данными должна включать не только ИТ решение, но и соответствующие бизнес-процессы, определенные роли и функции участников этих процессов



Эффект от внедрения процессов Управления данными

Опыт рынка: внедрение процессов управления данными позволяет ускорить реализацию и сократить затраты на инициативы, связанные с обработкой данных, реализацией аналитических решений, отчётности и ML-моделирования

Структура затрат в инициативах, связанных с аналитикой



* На основе исследований Gartner в части затрат на Data Science

RT.DataGovernance

Инструмент упорядочивания и документирования данных

Позволяет бизнесу понять:



Какими активами
данных он владеет



Как эти данные связаны
друг с другом



Кто и как
ими пользуется

Основные функции



Сбор, хранение, обработка,
удаление метаданных



Обеспечение доступности
и высокого качества данных



Оптимизация процессов
по работе с данными



Координация работы
технических
и бизнес-пользователей

Модули RT.DataGovernance



Дата Каталог



Бизнес-гlossарий



Data Quality



Поиск



Бизнес-портал



Data Lineage
Business Lineage



Детальный Реестр
отчетов



Каталоги любых объектов

3 уровня пользователей

1 уровень

Бизнес-портал

- Максимально простой интерфейс
- Новости и полезная информация
- Чат-бот для поиска нужной информации

2 уровень

Бизнес-гlossарий

- Подробная информация по каждому термину
- Логическая и концептуальная модель организации/департамента
- Связь логической и физической модели данных

3 уровень

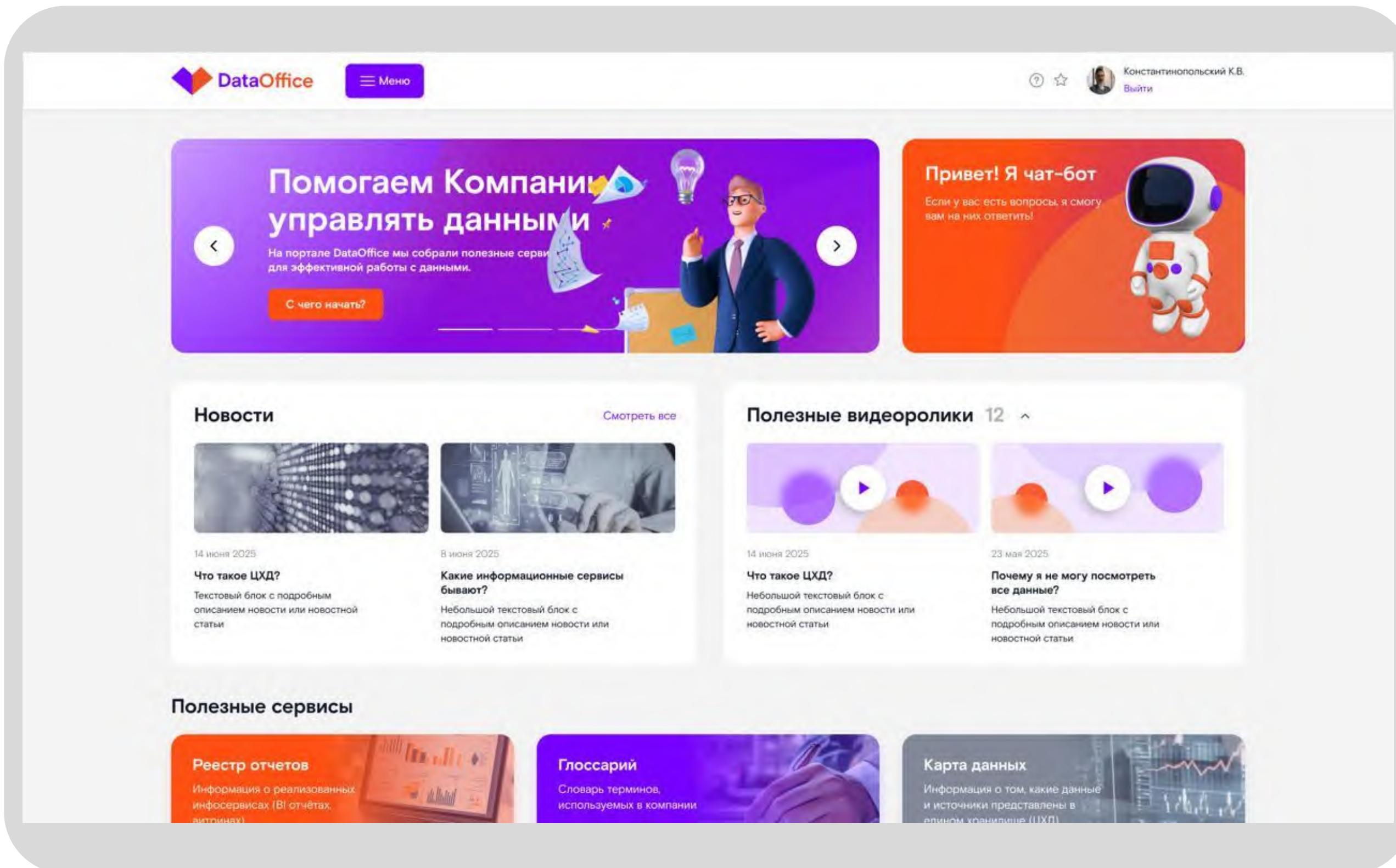
Каталог данных

- Максимально подробная информация по потокам данных
- Физическая модель данных
- Потоки и зависимости данных



Бизнес-портал

Портал для бизнес-пользователей, который позволяет оперативно искать и анализировать данные компании



⚙ Раздел управления заявками и доступами к системам

📊 Аналитические сервисы для работы с данными

🔗 Ссылки на внешние данные и ИС

📈 Ключевые показатели компании с детальным описанием

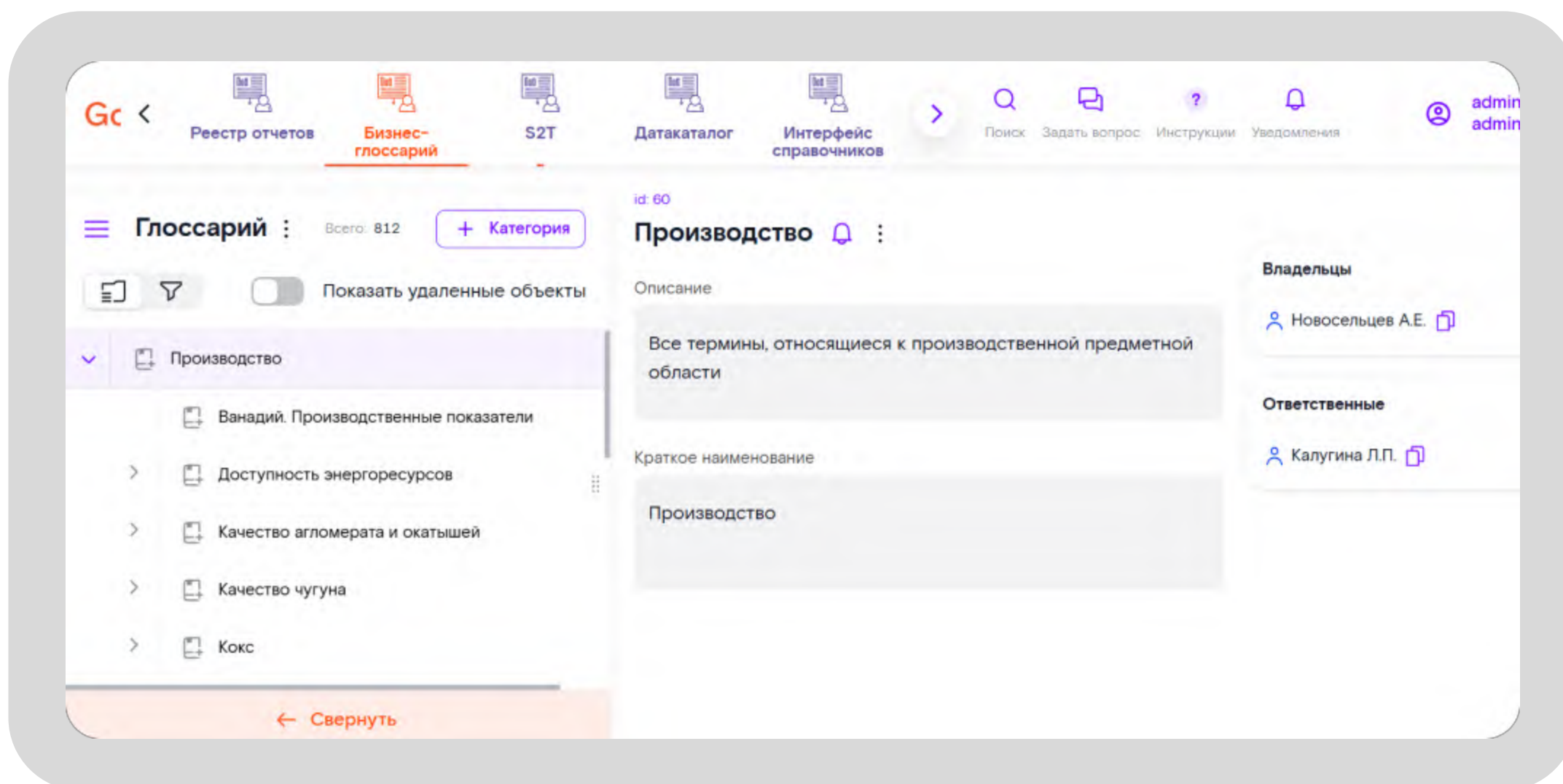
🔒 Доступ к Реестру отчетов и Бизнес-глоссарию

Бизнес-гlossарий

Инструмент для согласования и утверждения терминологии компании — единый «источник правды» для бизнес-терминов

Возможности

- Описание терминов и методологий их расчета
- Повышение качества отчетности с использованием расчетных показателей
- Удобное отображение взаимосвязей терминов с другими объектами данных
- Визуализация иерархии терминов



Преимущества

- Снижение трудозатрат на поиск терминов
- Снижение трудозатрат на создание, изменение и согласование терминов
- Отсутствие хаотичности в части источников и авторов терминов

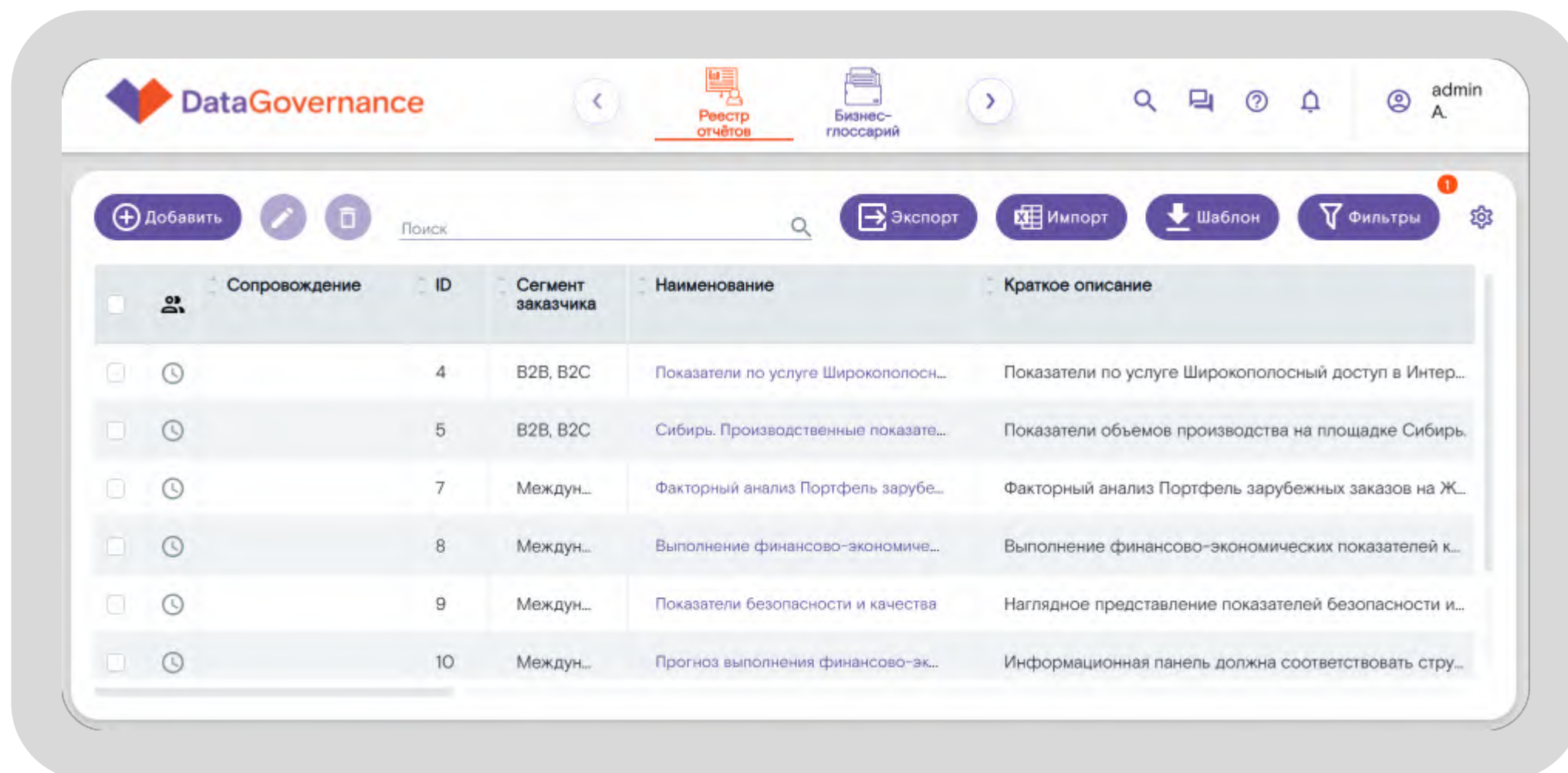


Реестр отчетов

Модуль для описания отчетов (BI-отчеты, OLAP-кубы), которые содержатся в корпоративных хранилищах данных и используются в компании

Возможности

- Предоставляет детальную информацию об отчете (владелец, тип, категория, статус)
- Исключает факты создания отчетов-дублей
- Отображает перечень атрибутов отчета
- Визуализирует путь отчета от источника данных



Преимущества

- Снижает трудозатраты на поиск и преданализ отчетов
- Снижает трудозатраты на привязку терминов из Бизнес-гlossария к отчетам
- Дает возможность оперативного экспорта данных о существующих отчетах



Каталог данных

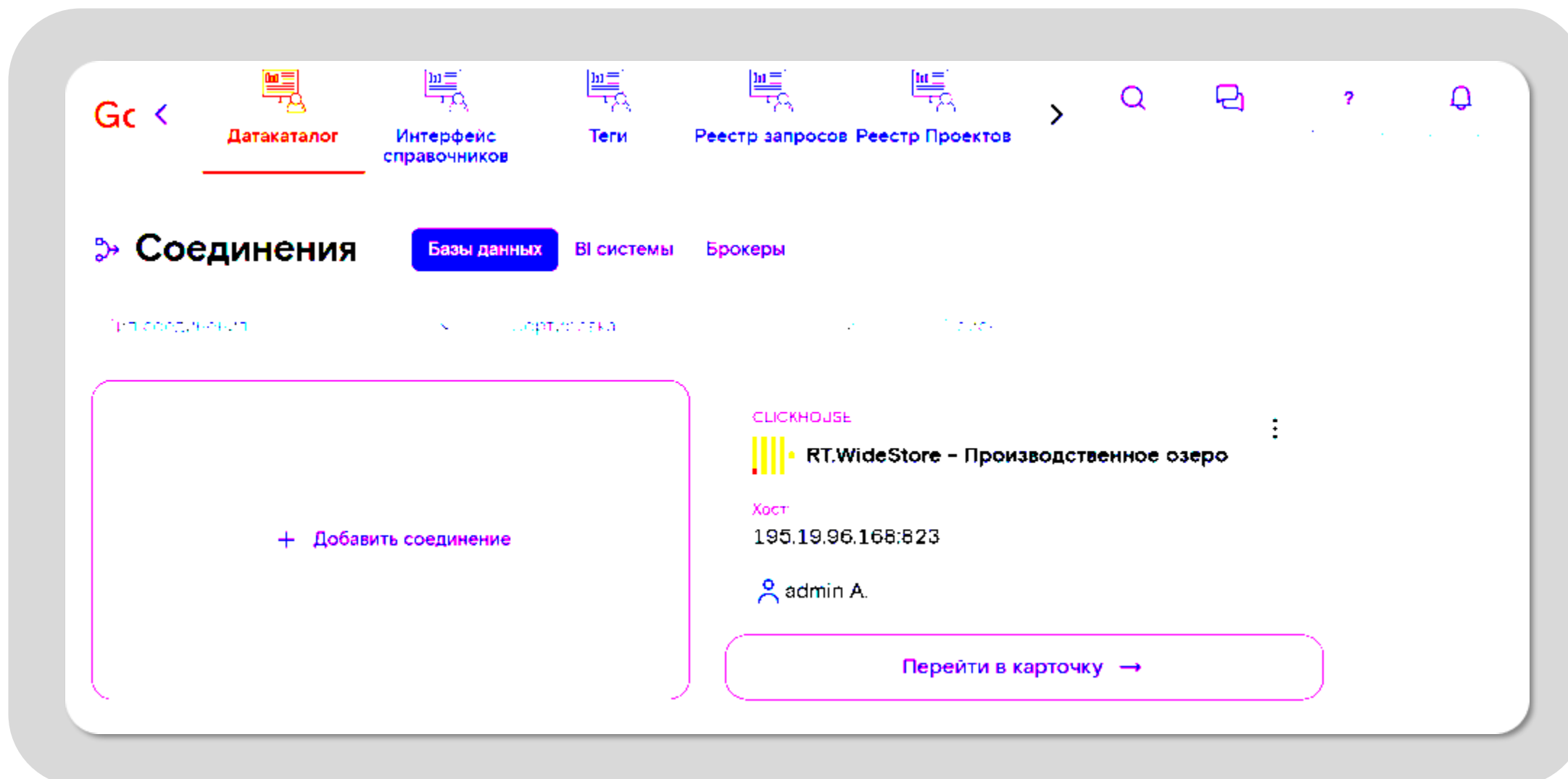
Хранилище метаданных, которое предоставляет информацию о доступности и актуальности существующих данных

Возможности

- Снижение трудозатрат на поиск и предоставление метаданных
- Оптимизация и упорядочивание процессов управления метаданными
- Увеличение возможностей для использования Self-Service аналитики
- Увеличение возможностей для глубокого анализа данных

Преимущества

- Сбор метаданных из различных источников (хранилище данных, БД, ИС и т.д.)
- Поиск и обнаружение метаданных
- Формирование иерархии данных и метаданных
- Визуализация пути данных с использованием S2T
- Экспорт метаданных в различных форматах
- TData 1C коннектор – выгрузка данных из 1С в BI и аналитическую базу



Модуль трансформации данных. Data Lineage

Модуль трансформаций — инструмент хранения и описания преобразования данных на физическом уровне по слоям хранилищ.

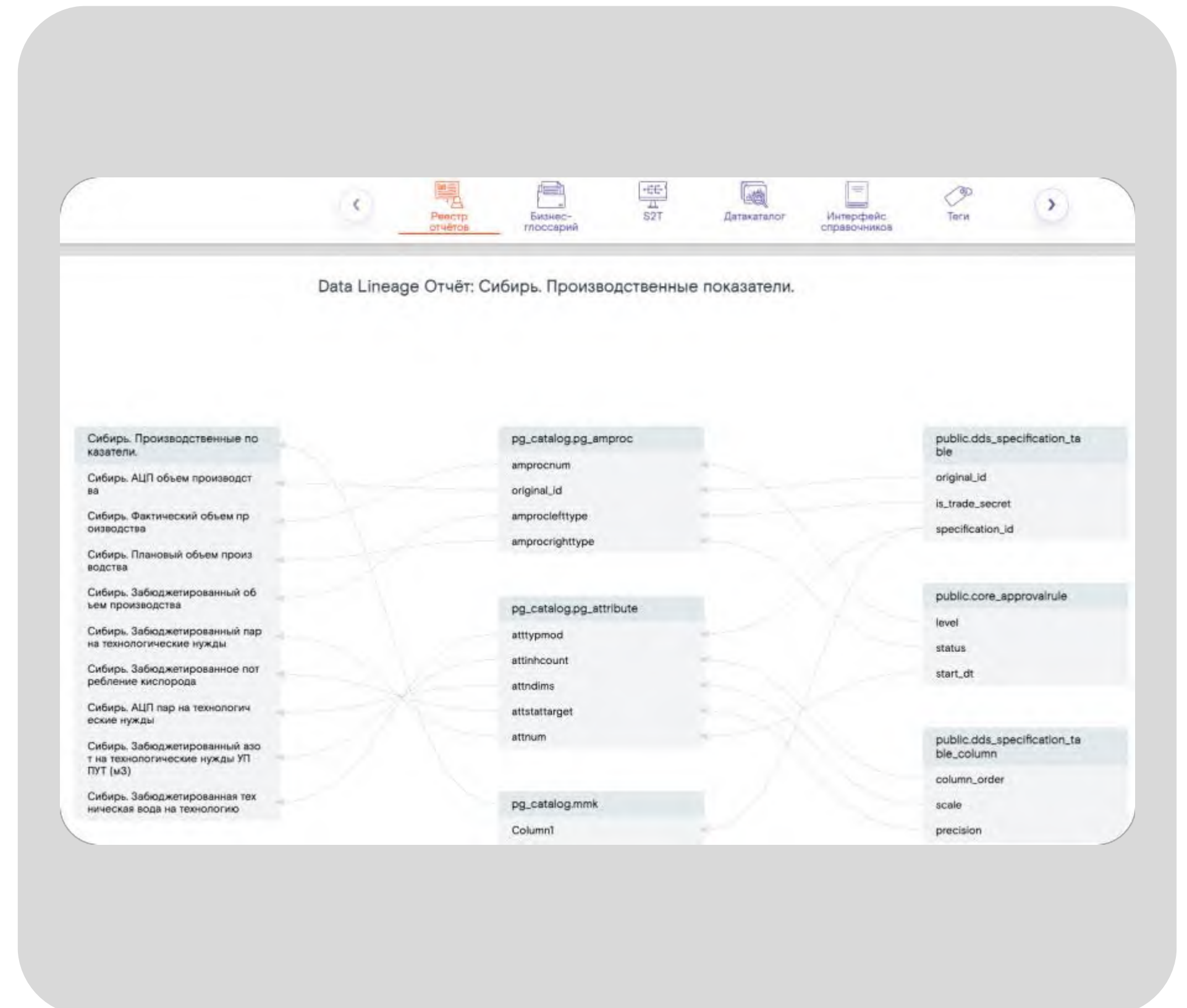
Возможности

- Позволяет отслеживать описания зависимостей для атрибутов данных
- Позволяет проводить мониторинг историчности версий объектов-источников

Data Lineage — инструмент визуализации жизненного цикла данных, построенного на основании документации Source-to-Target (S2T)

Возможности

- Позволяет отслеживать описания зависимостей для атрибутов данных
- Позволяет проводить мониторинг историчности версий объектов-источников



Data Quality

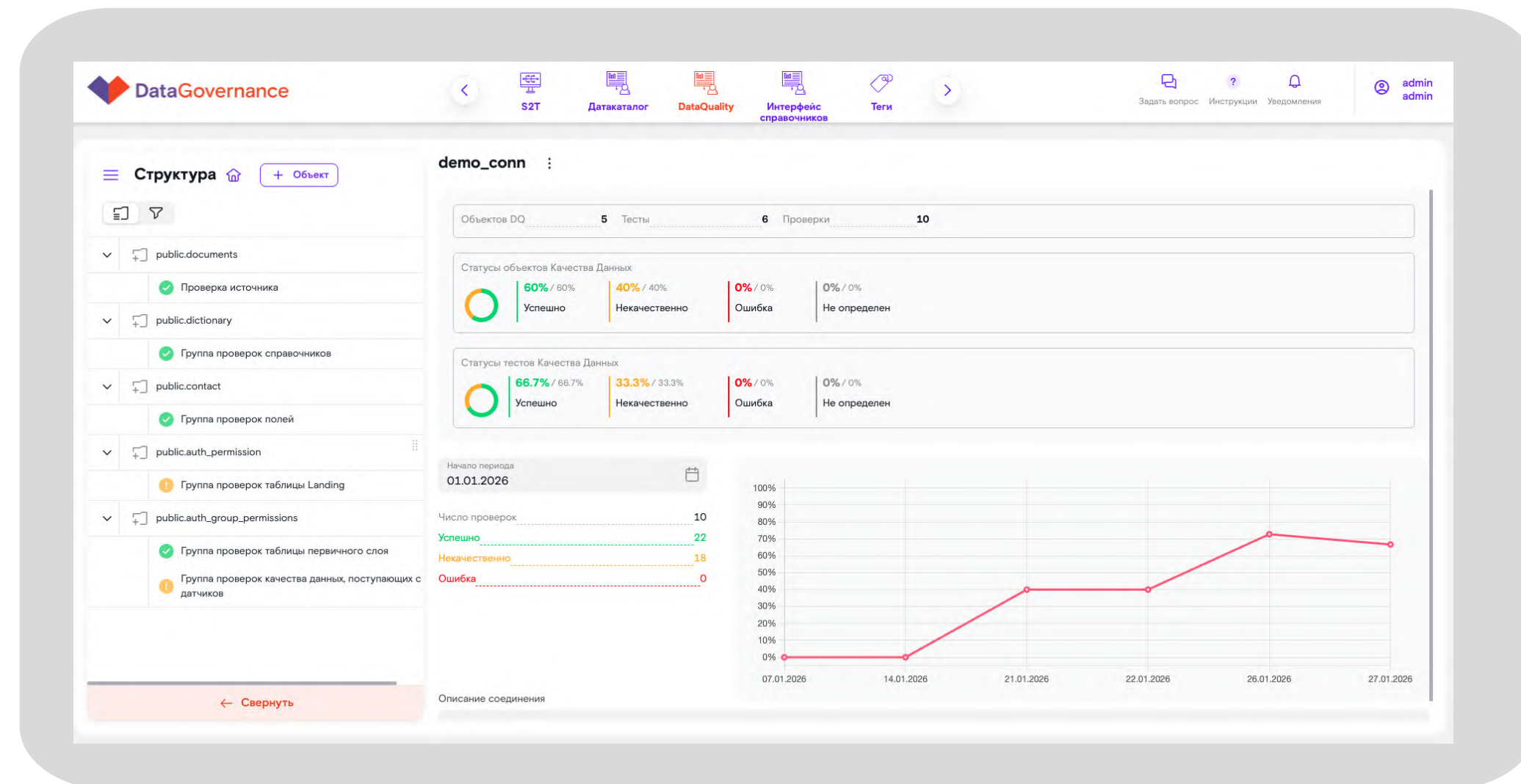
Решение Data Quality RT.DataGovernance позволяет на уровне настраиваемых бизнес-правил проводить анализ качества данных

Возможности

- Проверка для различных источников данных (EDW, DWH и т.д.)
- Реализация типовых и сложных проверок
- Формирование и рассылка отчетов об ошибках в данных
- Визуализация результатов проверок на дашбордах

Преимущества

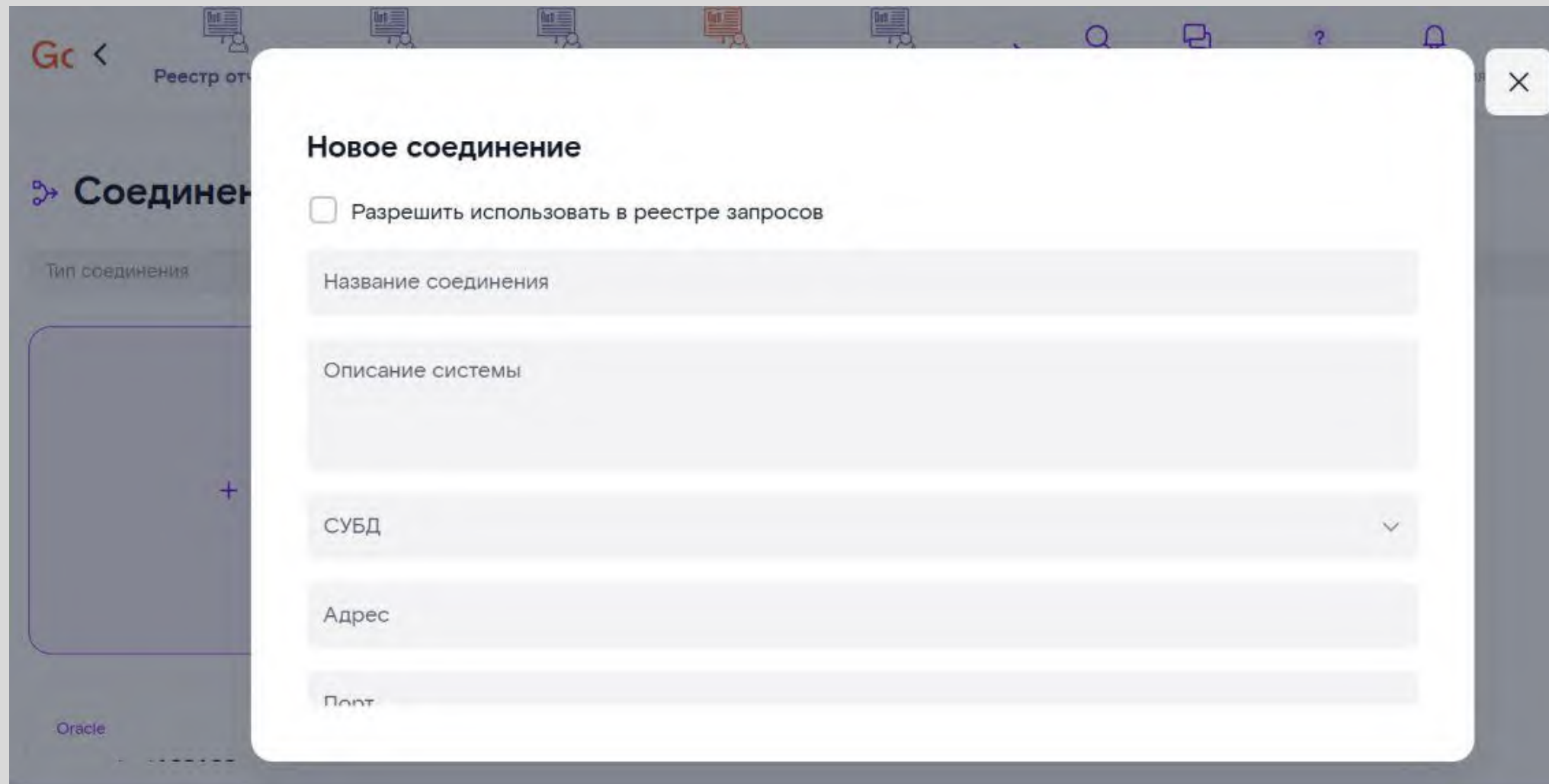
- Поддержание высокого уровня качества данных в компании
- Повышение эффективности принимаемых бизнес-решений на основе данных
- Снижение трудозатрат на перепроверку данных
- Визуализация пути данных с использованием S2T
- Оптимизация ролевой модели по проверке качества данных



Автоматический сбор метаданных

Реализованы следующие коннекторы:

- **СУБД:** Greenplum, Oracle, PostgreSQL, Clickhouse, MS SQL, MySQL, Teradata, Hive
- **BI:** Power BI, Superset, Форсайт, Fine BI, QlikSense (в разработке коннекторы к SSAS, BI Navigator, Tableau)
- **Транспорт:** Kafka, Yandex Data Streams и Airflow
- **Объектное хранилище S3**
- **Корпоративные системы:** 1С



The screenshot shows a web application interface for creating a new connection. The main window is titled "Новое соединение" (New connection). It contains a checkbox labeled "Разрешить использовать в реестре запросов" (Allow to use in the query registry). Below this are several input fields: "Название соединения" (Connection name), "Описание системы" (System description), "СУБД" (Database type) with a dropdown arrow, "Адрес" (Address), and "Порт" (Port). The background shows a sidebar with a "Соединения" (Connections) section and a "Реестр от" (Registry of) section.

Роли и выполняемый функционал

Инструмент DataGovernance	Владелец данных (Data Owner)	Куратор данных (Data Steward)	Менеджер по данным / Партнер по данным* (Data Manager / Data Partner)	Специалист по данным (Data Scientist)	Бизнес-аналитик (Business Analyst)
Дата Каталог	Определение приоритетов по управлению метаданными	Сканирование и описание метаданных	Формирование процессов по управлению метаданными	Использование метаданных в методах предиктивного анализа	Формирование бизнес-описания по метаданным
Бизнес-гlossарий	Управление терминами по своему домену	Контроль корректности заполнения карточек терминов	Формирование процессов по ведению терминологии	Поиск и использование терминов для составления проектной документации	Формирование и использование терминов при составлении БФТ
Реестр отчетов	Контроль ведения отчетов в разрезе бизнес-функции	Создание карточек отчетов	Формирование процессов по управлению отчетами	Поиск и выбор отчетов на основании описания	Формирование карточек отчетов
Модуль трансформации данных (S2T)	Мониторинг корректности Data Lineage по данным бизнес-домена	Построение и визуализация связей между объектами данных	Контроль процессов построения Data Lineage	Поиск объектов данных на основании Data Lineage	Заполнение форм S2T
Data Quality	Принятие бизнес-решений на основе качественных данных	Контроль и проведение проверок качества данных	Формирование процессов по управлению качеством данных	Анализ качества данных в БД	Мониторинг порогов качества данных
Единая точка входа	Контроль деятельности кураторов бизнес-функции	Формирование рекомендаций Офису данных по курируемым данным	Сбор требований к данным в рамках своего процесса по УпД** для формирования правил по УпД	Взаимодействие с бизнес-пользователями	Поиск и анализ данных

*Менеджер по внедрению УпД

** Управление данными

Новый функционал в RT.DataGovernance – 2025

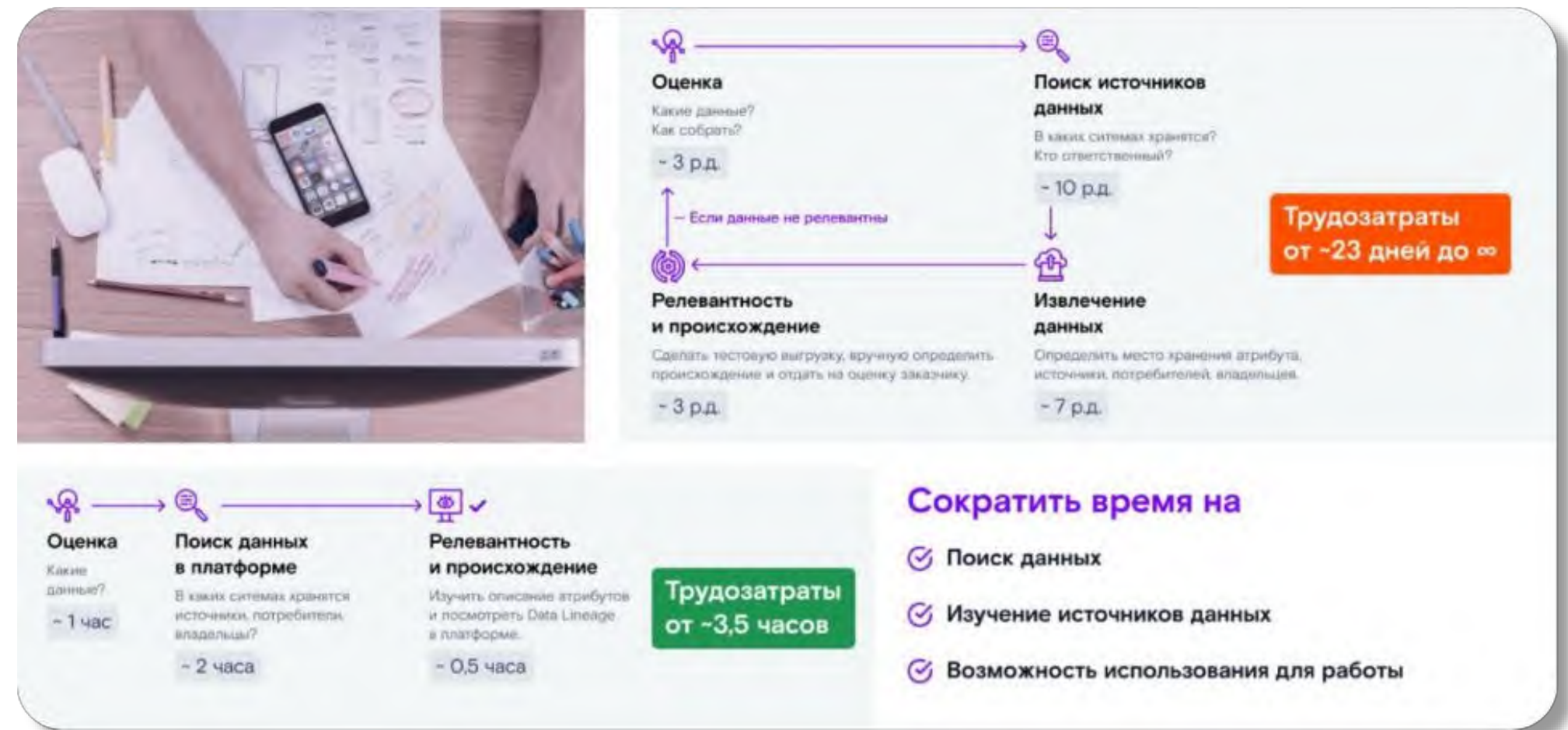
- Автоматическая разметка персональных данных на основе модели ИИ
- Встроенный модуль проверки качества данных с выводением результатов непосредственно в карточки объектов данных
- Возможность самостоятельно без привлечения вендора создавать каталоги/реестры и настраивать связи
- Возможность произвольной настройки контактов в карточках объектов
- Генерация описаний объектов
- Business Lineage
- Новая удобная главная страница с расширенными возможностями поиска, закладками, историей посещений и возможностью делиться фильтрами
- Значительные доработки интерфейса для удобства пользователя
- Новый дизайн Бизнес-портала



Проект внедрения ПАО «Ростелеком»

Задача

Обеспечить прозрачность данных отчетности, формируемой на базе Централизованного хранилища данных и более чем 10 локальных хранилищ, объединяющих данные более тысячи систем-источников в едином аналитическом контуре объемом более 750 ТБ.



Проблемы:

- Не было понимания, где и какие данные лежат и какая отчетность и где уже реализована
- Документация была разрознена, единые стандарты документирования отсутствовали
- Возникало дублирование функционала, бюджетов и времени на получение результата

Эффект от внедрения

Внедрение процессов управления данными на основе продукта RT.DataGovernance позволило:

- ✓ Более, чем в 20 раз сократить время на поиск и анализ данных сотрудниками
- ✓ Сформировать self-service контур, что значительно снизило количество консультаций и обращений по данным
- ✓ Оптимизировать процессы работы с метаданными
- ✓ Снизить затраты на формирование отчетности
- ✓ Исключить дублирование данных даже в географически-распределенных филиалах
- ✓ Утвердить единый модуль, содержащий бизнес-термины компании

Проект внедрения «Т2»



Задача

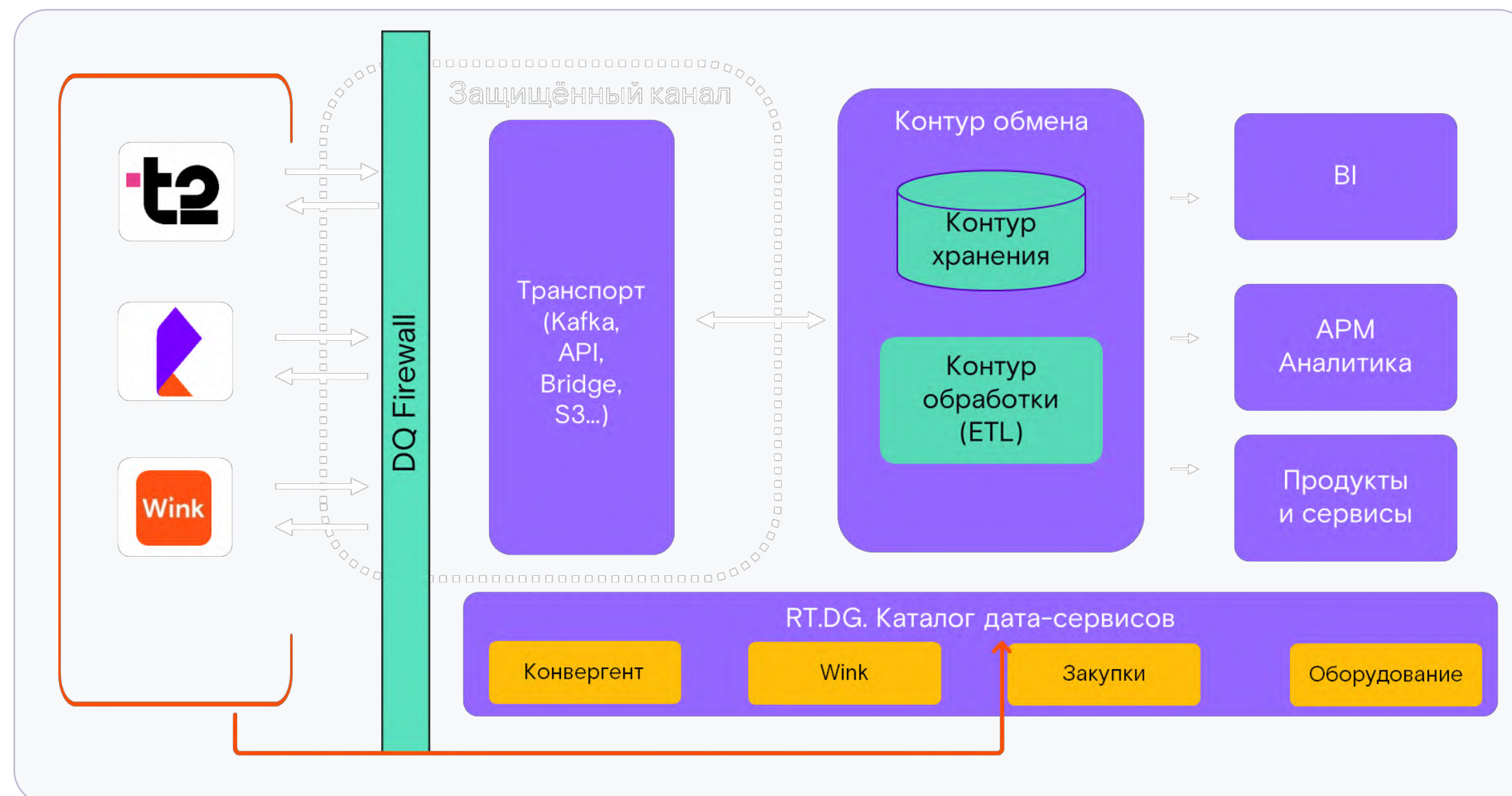
Построение универсальной масштабируемой платформы по управлению данными внутри экосистемы компаний с выстраиванием единых, объединяющих политик в области работы с данными

Ключевые компоненты

RT.DataGovernance
RT.Streaming

Результат

- ✓ К платформе подключены 3 юридических лица
- 👥 500 рабочих мест автоматизировано
- 📈 Обеспечен унифицированный управляемый механизм обмена любыми данными по ~270 атрибутам
- 💰 1,5 Тб (~16,5 млн строк) ежедневно передается внутри контура обмена
- 🛡️ Реализованы все требования информационной безопасности и безопасного (зашифрованного) канала обмена данными между разными юридическими лицами



Проект внедрения ЕВРАЗ



9000+

показателей в глоссарии

30+

СИСТЕМ ИСТОЧНИКОВ

1000+

разработано с использованием DG

3000+

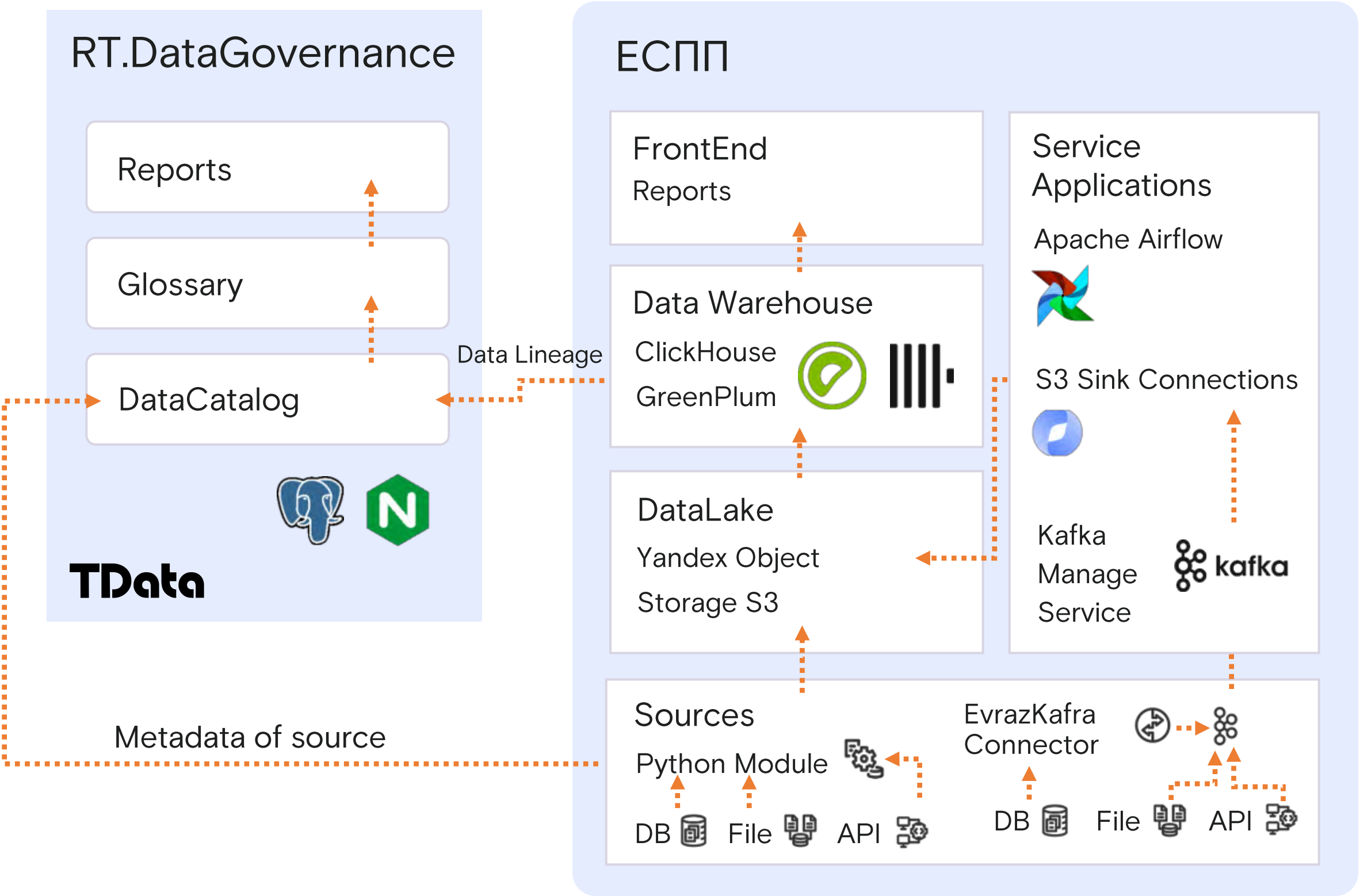
в работе

400+

Пользователей (из них 90% бизнес-пользователей)

X2

Ускорилаь разработка показателей в DWH



Проект внедрения AstraZeneca



Описание проекта

Внедрение решения для управления корпоративными данными в рамках цифровой трансформации международной биофармацевтической компании AstraZeneca. Для создания единой экосистемы, которая обеспечит эффективное управление информацией, повысит контроль качества и ускорит доступ к аналитике.

Решение

Внедрение системы RT.DataGovernance

Интеграция источников данных

- CRM-системы (система управления взаимоотношениями с клиентами)
- Системы DWH (корпоративное хранилище данных)
- Цифровые активы компании

Создание каталога данных и подготовка метаданных

- Один домен данных
- Пять ключевых систем-источников метаданных
- Более 4 тысяч таблиц
- Более 300 отчётов

Совместно с партнёром Ahenix разработана специальная методология управления данными, учитывающая специфику процессов и требований к безопасности и конфиденциальности данных в фармацевтической компании

Ключевые компоненты

RT.DataGovernance

Результат

-  Повышение эффективности внутренних процессов
-  Улучшение качества принимаемых управленческих решений
-  Обеспечение доступности актуальной информации
-  Существенное сокращение временных затрат на получение необходимых аналитических данных

Результаты внедрения RT.DataGovernance



Снижение затрат
на поиск и изучение данных



Обеспечение согласованности
данных для всех доменов данных



Развитие культуры работы
с данными в компании



Увеличение качества данных
для принятия более
эффективных бизнес-решений



Увеличение эффективности
взаимодействия технических
и бизнес-пользователей



Оптимизация бизнес-
планирования и рост финансовых
показателей компании



Оптимизация процессов
работы и получения доступов
к аналитическим системам



Увеличение эффективности
анализа данных
и предиктивной аналитики

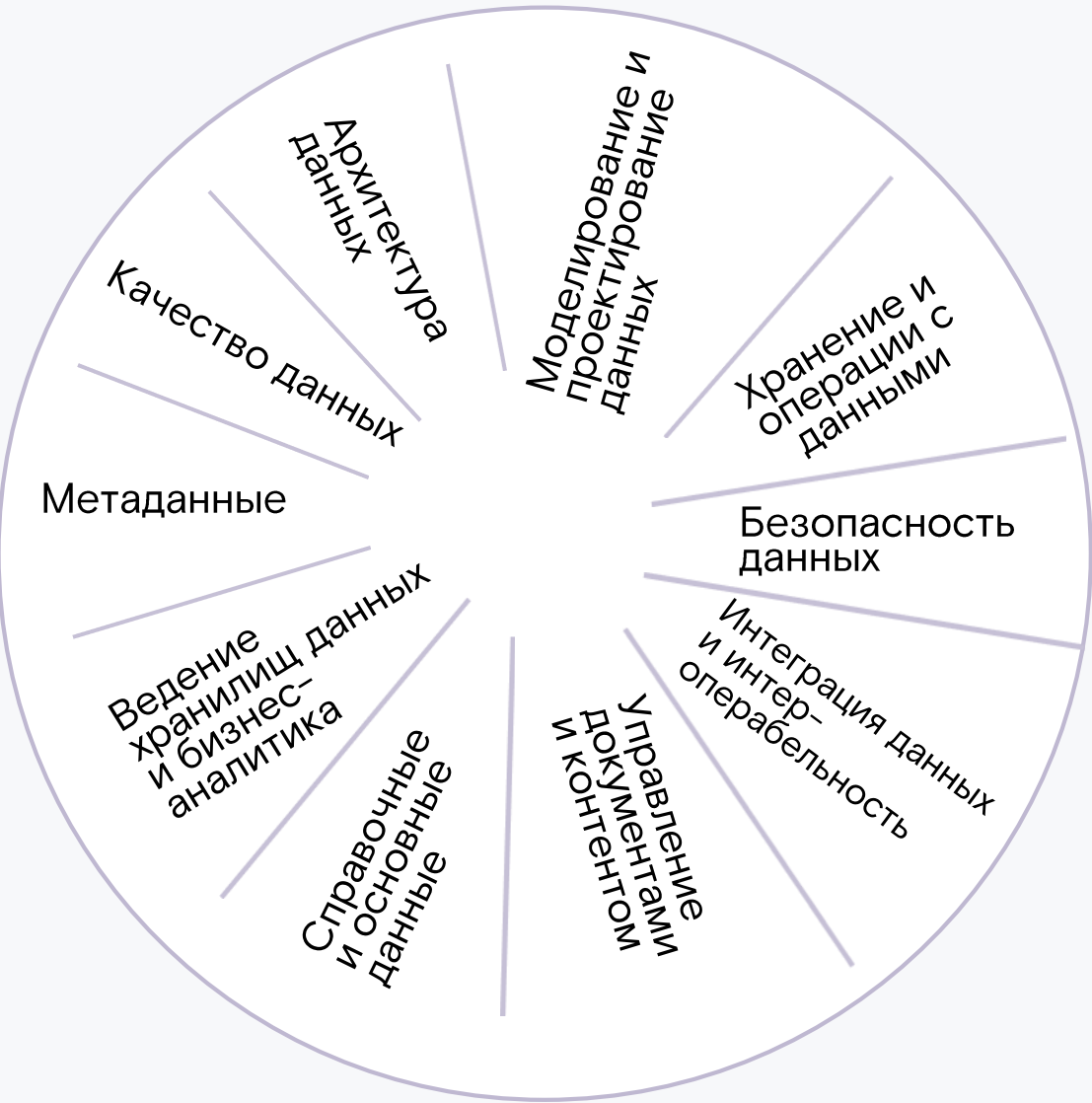


Снижение длительности
согласований по данным
в периметре компании

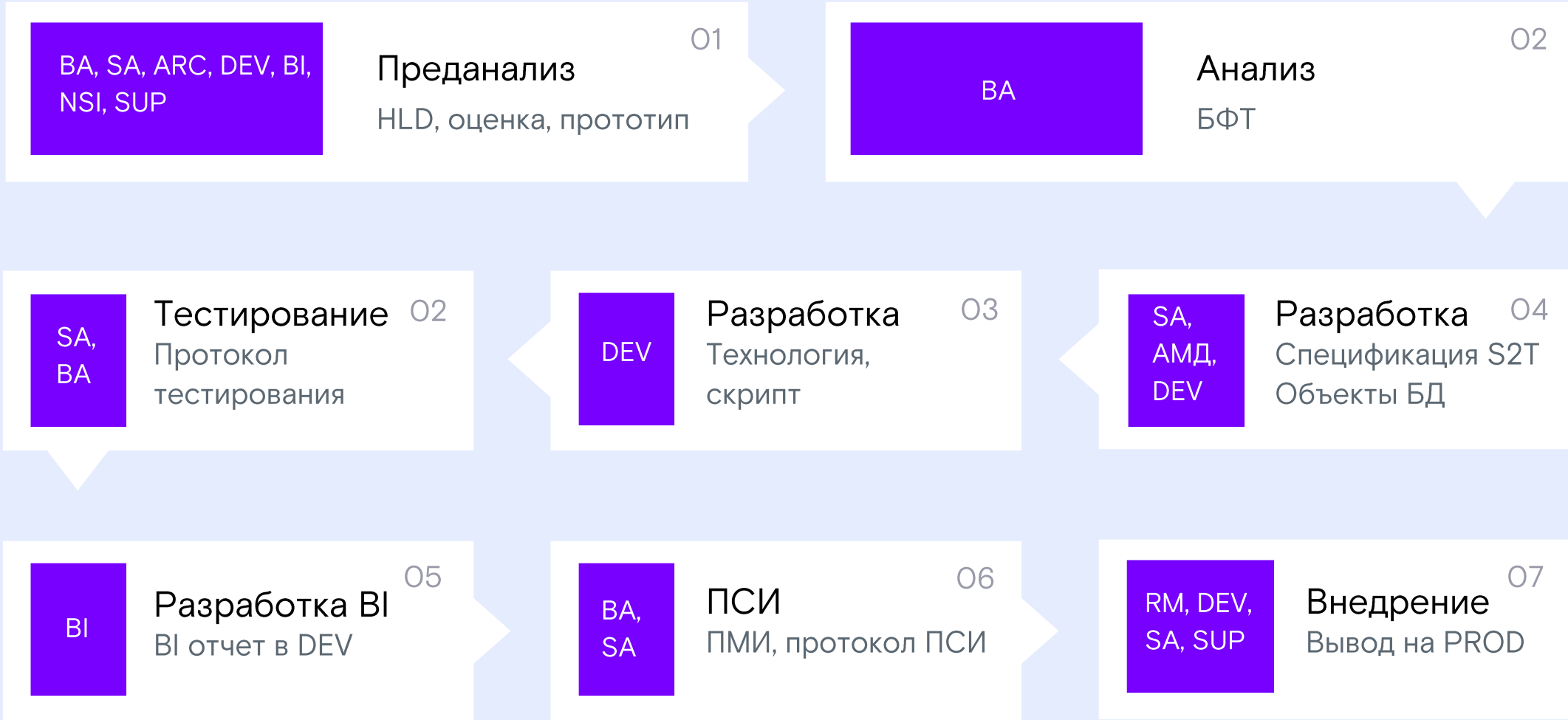
Платформа полного цикла

RT.DataGovernance реализует поддержку ключевых функций управления корпоративными данными, в т.ч. архитектура, моделирование и качество данных, фокусируясь на управлении метаданными

Рамочная структура управления данными DAMA-DMBOK2 (колесо DAMA)



Типовой процесс разработки/изменения дата-актива в Enterprise



TData

Спасибо за внимание

TData

Будем рады
ОТВЕТИТЬ НА ВОПРОСЫ

sales@tdata.tech

