

A2DATA



Цифровой Холдинг: Качественные данные как инструмент управления

DS & AI

Инжиниринг данных



BIG DATA APPLICATIONS



Визуализация



Data Science & Artificial Intelligence

ЦИФРОВАЯ ИНФРАСТРУКТУРА ДЛЯ DATA-DRIVEN COMPANY



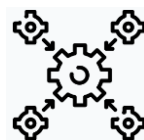
Очистка
данных



DWH & Data
Lake



Интеграции и ETL



MDM

- **Принятие решений на основе данных**
 - **Аналитические инсайты**
 - **Обновляемые показатели**
 - **Команда дата-стюардов** (управление данными и качеством) **и дата-аналитиков** (*data science, BI, ML*)
-
- **Регламентированный доступ к данным**
 - **Разработка витрин данных**
 - **Трансформация, извлечение, очистка и обогащение данных**
 - **Требования к качеству данных**
 - **Согласованная методика расчетов показателей**
 - **Определение основных источников данных и их владельцев**
 - **Описание структур базы данных**

Дэшборды

- Разработано **55** информационных панелей в дэшбордах
- Автоматизированы расчеты **150** показателей



Аналитические кейсы

Группировка заемщиков

Анализ заемщиков на основе связей



Эффективность господдержки

Оценка влияния на рабочие места и налоги



Анализ сельхозтехники

Оценка потребностей в технике по регионам



Оценка заявок

Анализ процесса и времени обработки заявок



Портрет клиента 360

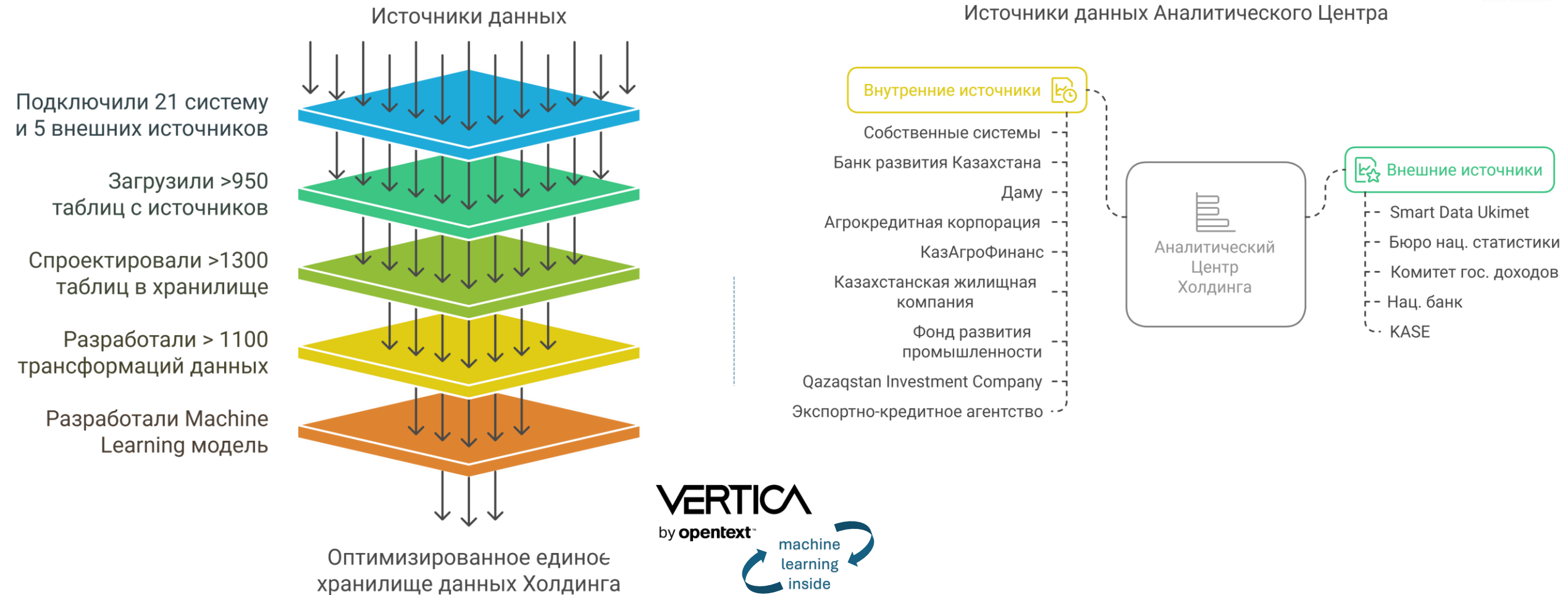
Комплексный обзор профилей клиентов



Кросс-аналитика заемщиков

Анализ заемщиков по всей группе





Data Vault — архитектура, которая масштабируется вместе с Холдингом и выдерживает изменения.

Сложная структура и частые изменения

- Разнообразие систем и юридических лиц
- Многочисленные стейкхолдеры
- Глубокая иерархия данных
- Частые изменения в системах и бизнес-процессах

Требования Холдинга



Возможности Data Vault



Обеспечение гибкости и масштабируемости

- ✓ Гибкая интеграция разнородных источников
- ✓ Хабы, линксы и сателлиты = разные витрины для разных ролей
- ✓ Масштабируемая структура без потери связности и отслеживаемости
- ✓ Устойчивость к изменениям: можно добавлять новые данные без переделки модели



Зоны отсутствия автоматизации
Некоторые данные ведутся в Excel

Автоматизация ключевых процессов

- Покрытие автоматизацией всех бизнес-процессов Холдинга и 11 дочерних компаний
- Внедрение excel-free практики в отчетности



Низкое качество данных на источниках
Отсутствует контроль вводимых данных в системах и ответственность за ввод некорректных данных

Работа с источниками данных

- Повышение качества данных в ИС за счет доработки интерфейсов ИС на форматно-логический контроль Холдинга и 11 дочерних компаний
- Внедрение модуля Observability по контролю изменения структуры данных источника
- Внедрение модуля скоринга качества данных на источниках



Ручная корректировка данных
Корректировка отчетных данных, экспертный «ручной» метод расчетов показателей

Изменение стиля работы

- Переход с экспертного метода расчета на data driven
- Изменение процесса со «сдачи отчетности» на регулярное автоматизированное обновление данных Аналитическим центром
- Дата стюарды Холдинга на регулярной основе выверяют несоответствия и отрабатывают с дочерними организациями и кураторами Холдинга по исключению ручной отчетности и логики расчетов в Excel

Data Driven

AI Driven

Выстраивание фундамента Data Governance:

- Единые справочники по группе
- Контроль качества данных на источниках
- Самообслуживание в структурных подразделениях
- Внедрение института CDO
- Адаптация KPI под Data driven подход
- Автоматизация процессов Data Quality

Расширение фундамента под AI:

- Полное покрытие всей группы Data Driven KPI
- Действующие процессы, регламенты и ответственность по работе с данными

Расширение аналитики и data science:

- Оперативная отчетность в Аналитическом центре
- Отчетность в виде дэшбордов
- Мониторинг показателей, ad-hoc анализ
- Оптимизация бизнес процессов

Machine learning & AI:

- Принятие решения на основе скоринга
- Прогнозирование и моделирование
- Подключение единого хранилища к ИИ платформе через RAG
- Таргетированные меры поддержки

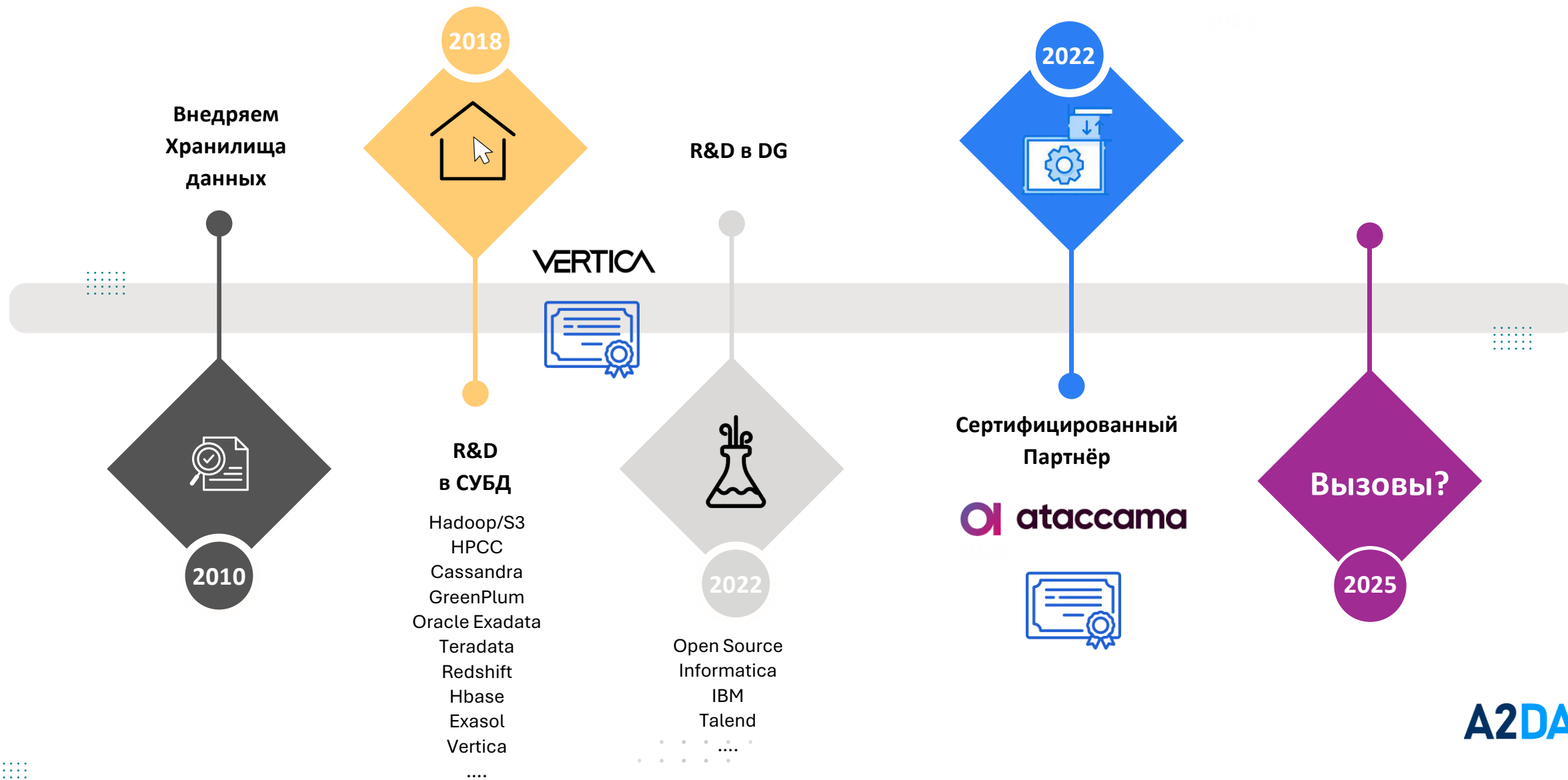
Расширение платформы:

- Подключение новых источников
- Подключение новых модулей по Data Governance

Расширение платформы:

- Автоматизированная отчетность в ГО
- Расширенный обмен данными с рынком
- Разработка скоринговой модели

Роль R&D для производства качественных данных



Что нужно бизнесу от данных

Time-to-market изменений, реализуемых на основе данных, возможности self-service



Доступность данных



Скорость поиска

- › Корпоративный Google-поиск терминов, метаданных в каталоге данных, владельцев, источников и пр.
- › Интерпретация данных – четкое понимание одинаковых и разных по смыслу терминов, атрибутов, расчетных показателей



Анализ зависимостей

- › Исследование связей между различными объектами метаданных, анализ происхождения, построение data lineage
- › Определение того, что может «сломаться» при внесении изменений, оптимизация ресурсов тестирования и поддержки



Доверие к данным



Достоверность данных

- › Определение качества данных на любом шаге
- › Проверка данных при заведении в учетные системы
- › Автоматическая очистка, валидация трансформация и стандартизация данных



Прозрачность качества данных

- › Профилирование данных – автоматический поиск аномалий, выбросов, ошибочных данных
- › Понимание общего состояния данных во всех доступных информационных системах

ПРОЦЕССЫ DATA GOVERNANCE

КАТАЛОГ ДАННЫХ

УПРАВЛЕНИЕ КАЧЕСТВОМ ДАННЫХ

Особенности холдинговых структур, групп компаний и эко-систем

ОСОБЕННОСТИ ХОЛДИНГОВ

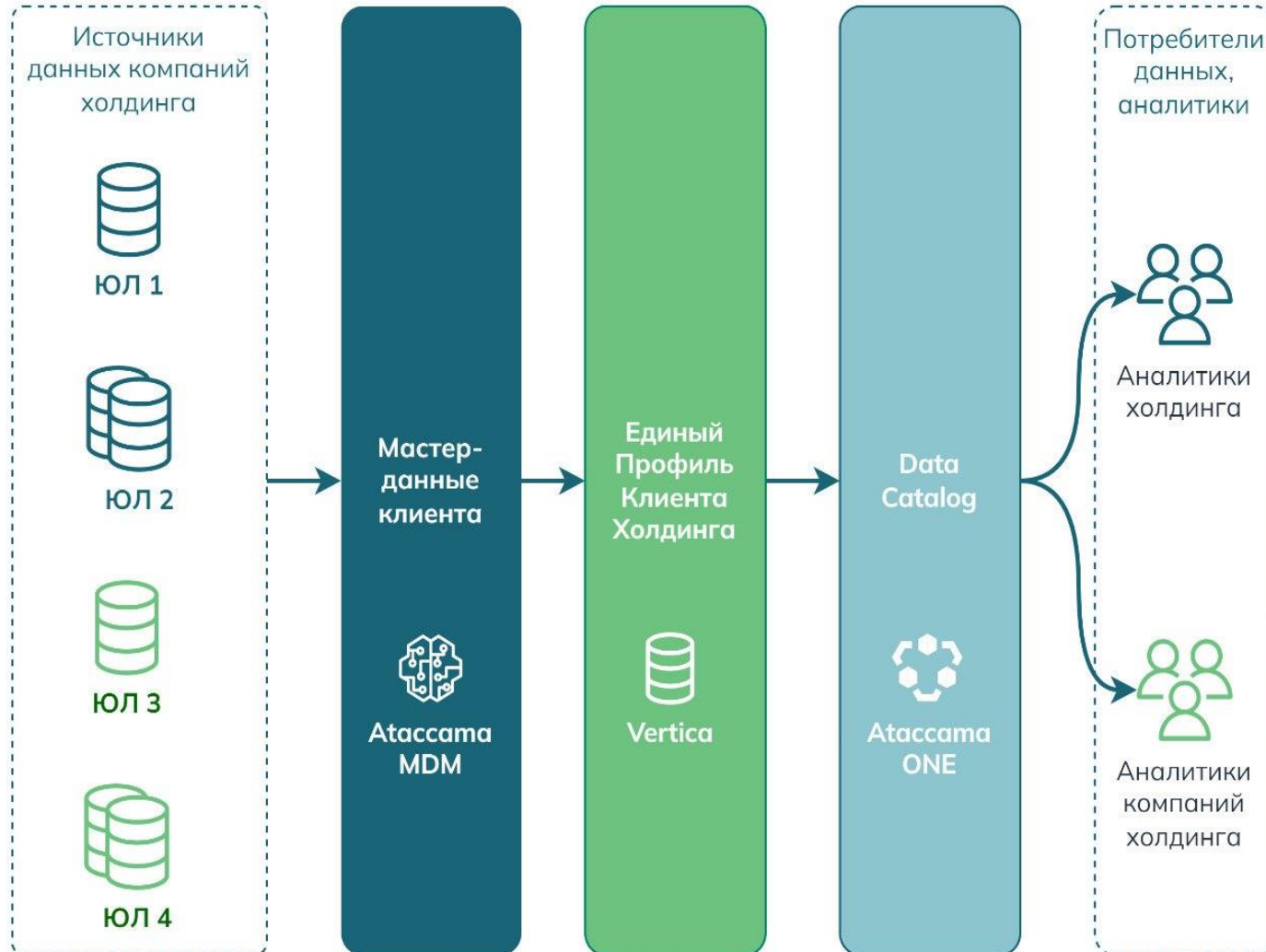
- › Важность контроля качества данных, собираемых от дочерних структур, важность эффективности (скорость и стоимость) процессов контроля
- › Огромные возможности по клиентской аналитике и кросс-холдинговой монетизации данных
- › Синергия внутри холдинга для развития практики Data Governance/Data Quality

СЛОЖНОСТИ И ВЫЗОВЫ

- › Экспоненциальный рост сложности и разнообразия данных, проблемы с time-2-market
- › Распределенная ответственность за данные
- › Сложности управления персональными данными и консентами
- › Сложность решения задач информационной безопасности



ОПЦИЯ 1: ЦЕНТРАЛИЗОВАННАЯ АРХИТЕКТУРА ДАННЫХ



1

Вызовы и видение

- ✓ Холдингу требуется **эффективная аналитическая платформа** для быстрого запуска клиентских кампаний **cross-sell** и **up-sell**
- ✓ Аналитика должна строиться на данных, которым можно доверять, то есть на основе **очищенных, унифицированных и консолидированных** данных
- ✓ Платформа должна базироваться на **зрелых** решениях и обеспечивать потребности **аналитики** при кратном росте **объемов данных**

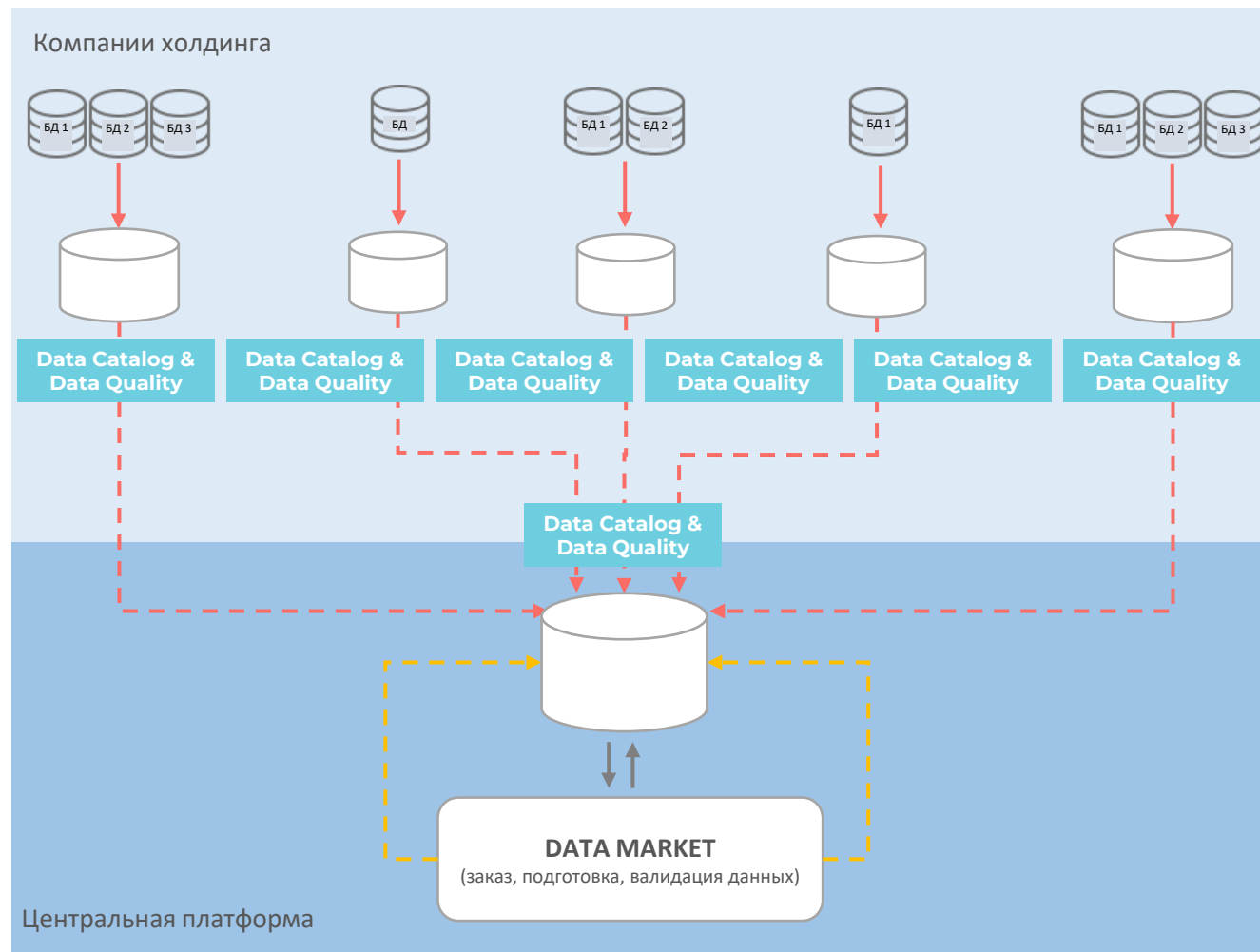
2

Результаты проекта

- ✓ В рамках холдинга спроектирована и внедрена **аналитическая платформа**, состоящая из следующих основных компонентов
 - › **Клиентский хаб, Master Data Management**, выполняющий функции **очистки, унификации и консолидации** основных данных существующих и потенциальных клиентов холдинга
 - › **Единый Профиль Клиента**, в котором основные данные клиента **объединяются** с продуктовыми, финансовыми, транзакционными и прочими данными клиентов
 - › **Data Catalog**, служащий для описания **бизнес-терминов**, привязки их к физическим данным, и обеспечивающий **легкий поиск** нужных данных для аналитиков
- ✓ Аналитическая платформа обеспечивает **эффективную** работу **аналитиков** как самого **Холдинга**, так и дочерних компаний **холдинга** - в рамках **совместных** маркетинговых и продуктовых инициатив

ОПЦИЯ 2: РАСПРЕДЕЛЕННАЯ АРХИТЕКТУРА

Распределенный подход к процессам Data Governance и Data Quality требует больше усилий, но позволяет реализовать сценарии аналитики между структурами группы и компаний и кросс-продажи даже в условиях невозможности обмена персональными данными



ВЫЗОВЫ И ВИДЕНИЕ

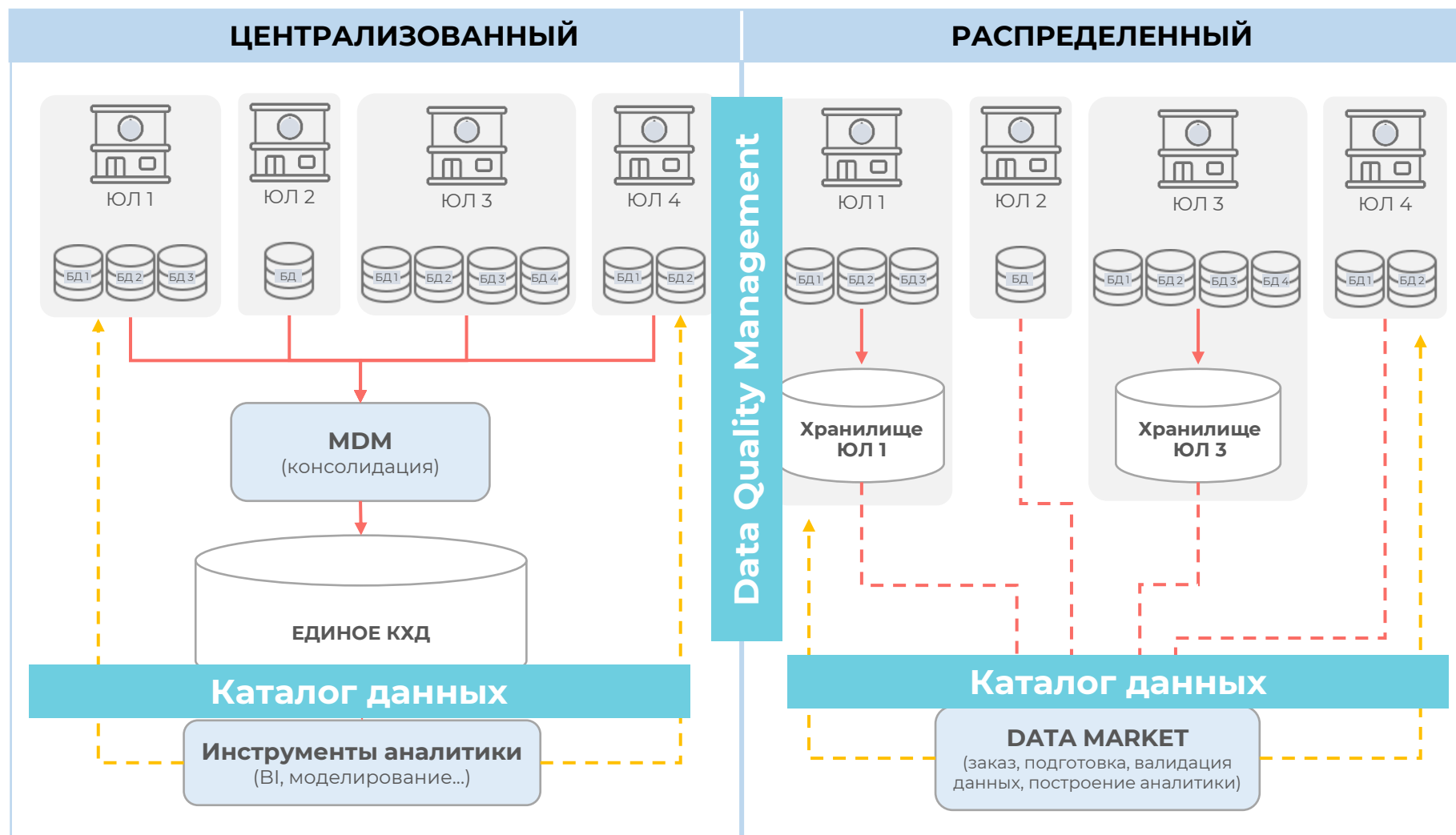
- Холдингу требовалась **эффективная платформа** для быстрого и удобного обмена данными для запуска клиентских кампаний **cross-sell** и **up-sell**
- Данные** любой компании Холдинга **могут быть запрошены любым сотрудником**, имеющим доступ к центральному КATALOGу данных
- Создание нового запроса должно происходить в **удобном интерфейсе** путем выбора интересующих бизнес-терминов конкретных организаций
- Это должен быть **полностью автоматизированный процесс** от момента создания запроса до получения нотификации о готовности данных

РЕЗУЛЬТАТЫ ПРОЕКТА

В рамках Холдинга была спроектирован и внедрен **self-service по заказу данных**, который обеспечил:

- Единую точку входа** - Любой сотрудник, имеющий доступ в центральный Каталог, может запросить данные из других организаций
- Гибкость процесса** – запрос может быть создан как на новый датасет, так и на изменение существующего
- Процесс согласования** - на каждого Владельца данных создается отдельная задача для получения разрешения, при этом предусмотрена автоматическая проверка по списку ранее выданных разрешений
- Прозрачность процесса** - у инициатора запроса всегда есть возможность отслеживать детальный статус по своему запросу
- Автоматизированную выдачу данных** - процесс запроса данных полностью автоматизирован: от момента создания запроса, до получения готовых данных

ВОЗМОЖНАЯ АРХИТЕКТУРА ДАННЫХ ДЛЯ ГРУППЫ КОМПАНИЙ: ЦЕНТРАЛИЗОВАННЫЙ vs РАСПРЕДЕЛЕННЫЙ ПОДХОД



- **Централизованный подход** подразумевает централизацию всех усилий по консолидации данных, их очистке и управлению данными. Данный подход требует построения «золотой записи о клиенте»
- **Распределенный подход** подразумевает выстраивание процессов управления данными и качеством данных в каждой из структур холдинга, создание **единого каталога данных и внедрение процессов data market** – возможность обмена обезличенными набором данных для аналитиков и data scientist-ов
- Чаще всего для больших организаций используется **комбинированный подход**
- **Ключевым фактором успеха является** внедрение процессов и платформы управления **качеством данных**

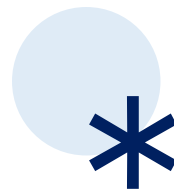
Summary



Независимо от выбора подхода – централизованного vs распределенного – невозможно построить data driven решения без инвестиций в data governance и data quality



Для холдингов и экосистем – это сделать на порядок сложнее чем для отдельной компании, тем не менее инвестиции в DG/DQ – обязательное условие успешности любых дата-инициатив в группах компаний



Важны и процессы DG/DQ и инструмент – покупка инструмента без инвестиций в процессы не даст эффекта, с другой стороны, внедрение процессов DG/DQ без правильного инструмента практически никогда не срабатывает, эффективный современный инструмент – помогает сделать проекты DG/DQ успешными