

Корпоративная платформа хранения и обработки больших данных

Вызовы цифровой трансформации в ритейле



Проблемы цифровой трансформации в ритейле

- Необходим единый масштабируемый механизм для централизованного сбора, хранения и анализа информации и оперативно предоставления результатов руководству и бизнес-пользователям
- Наличие **исторических хранилищ данных и аналитических инструментов**, унаследованные от приобретённых ранее компаний
- Отсутствие консолидированного хранилища масштаба холдинга и единой точки входа к корпоративным данным для бизнес-пользователей
- Отсутствие единых регламентов хранения, проверки качества и использования данных
- Дублирование и рассогласованность данных
- Высокие затраты на обеспечение безопасности данных, поддержание и использование разнородных инструментов ВІ
- Слабая информированность бизнес-пользователей о наличии необходимых данных
- Слабое вовлечение продвинутых бизнес-пользователей в разработку аналитических сценариев (отсутствие "песочницы")
- Длительный промежуток времени между заявкой от бизнеса и разработкой отчета (набор разнообразных отчетов еще не означает BI)





С такими задачами столкнулись



































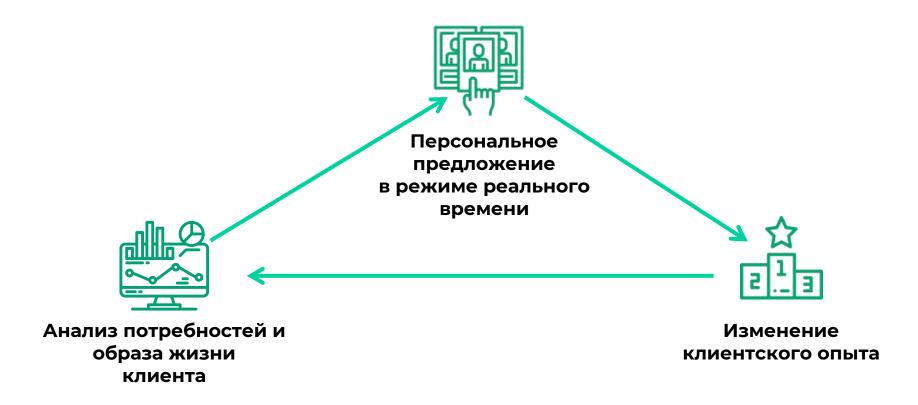






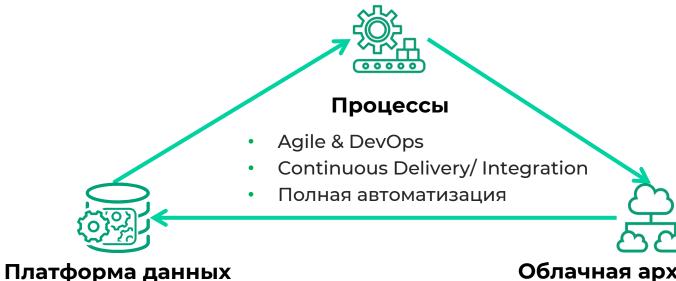


Цифровая компания - это





Основные ИТ-компоненты «Цифровой компании»



- Обработка различных типов данных
- Горизонтальное масштабирование

Облачная архитектура и приложения

- Распределенность и микросервисы
- Time-to-market



Этапы становления цифровой компании

- Построение аналитической платформы данных
- 2. Персональная и интерактивная коммуникация с клиентом
- Формирование нового клиентского опыта
- 4. Развитие новых направлений и сервисов





Характеристики современной платформы данных



Эластичность и горизонтальная масштабируемость



Поддержка любых регламентов: or real-time до mega-batch



Поддержка любых типов и форматов данных



ANSI-SQL-совместимость



Open Source как залог прозрачности



Параллельный обмен и pushdownоптимизация между компонентами платформы



Интегрируемость и модульность компонентов платформы



Единая система управления и мониторинга для всех компонент



Кто мы

ARENADATA — ведущий российский разработчик платформ сбора и хранения данных для компаний, которые хотят изменить бизнес с помощью цифровизации процессов и сложной аналитики, основанной на использовании Big Data.

Наши решения строятся на технологиях с открытым кодом и помогают работать с



Большие данные



Интернет вещей



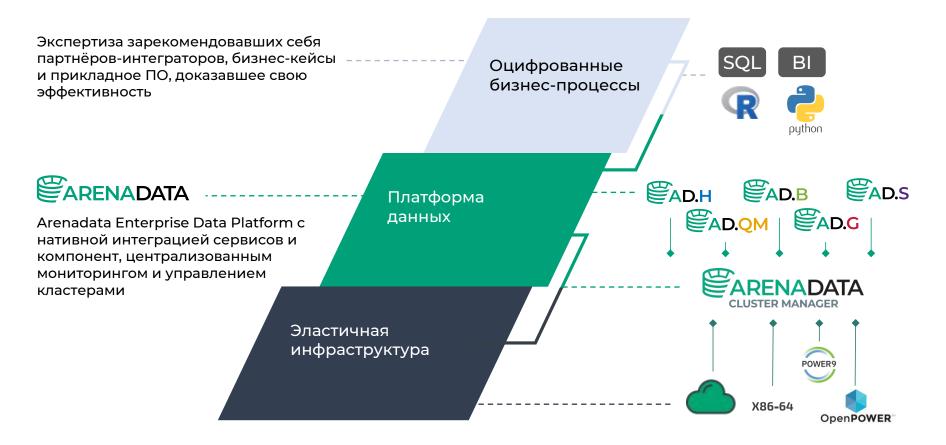
Машинное обучение



Связанные технологии



Платформа для цифровой компании



Множество инструментов для решения прикладных задач

Любое ПО на любой инфраструктуре: эластично, отказоустойчиво, под любые реалии и задачи



Решаемые задачи



Корпоративные хранилища больших данных

- Классические КХД и хранилища больших данных с обновлением в реальном и околореальном времени.
- Виртуализация и федерация данных, доступ ко всем данным через единое окно с помощью SQL-запросов.



Разгрузка дорогостоящих инженерных систем (Teradata, Exadata)

- Импортозамещающие ПАК на базе массово-параллельных СУБД с открытым исходным кодом Arenadata DB и Arenadata QuickMarts.
- Прозрачная методика перехода, минимум рисков и сохранение всех преимуществ.



Машинное обучение и искусственный интеллект

- Поддержка R, Python и Madlib.
- Distributed ML на базе Spark.
- Возможность эффективной эксплуатации моделей в SQLсреде с помощью встроенных функций Madlib.

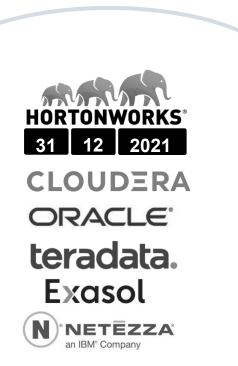


Озёра и фабрики данных

- Единый хаб всех данных компании.
- Возможность быстрого развёртывания и сворачивания зон «песочниц» для пилотных проектов и проверки статистических гипотез.
- Работа со всеми аналитическими инструментами в единой среде.



Скажите «ДА» современной платформе!



Скажите «ДО СВИДАНИЯ»

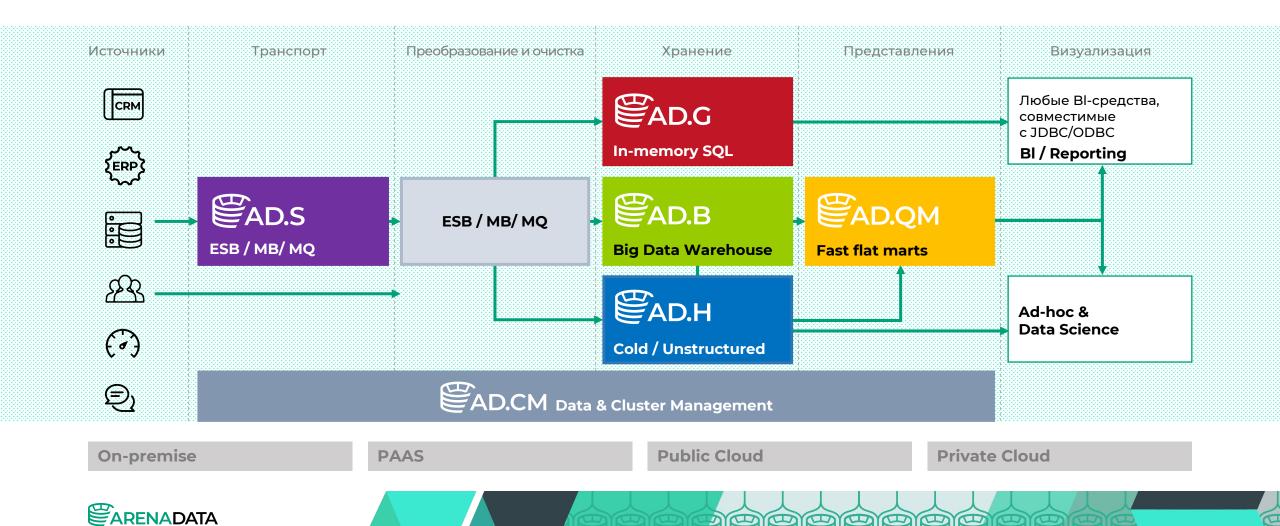
- закрытой архитектуре
- монолитной архитектуре
- vendor lock-in
- высокой ТСО



Мы обладаем опытом миграции клиентов с других систем



Современная платформа данных



Компоненты платформы EDP

₩ AD.B	₩ AD.H	₿ AD.QM	₩ AD.G	₩ AD.S
Хранилище данных (МРР-СУБД)	Хранение холодных и слабо структурированных данных	Витрины данных	In-Memory SQL	Streaming (ESB / MB / MQ)
Агепаdata DB (ADB) — аналитическая, распределённая СУБД с открытым исходным кодом, использующая концепцию MPP (massively parallel processing), построенная на базе СУБД Greenplum. Она реализована на кластере из множества серверов и предназначена для хранения и обработки больших объёмов данных — до десятков петабайт.	Arenadata Hadoop (ADH) — это корпоративный дистрибутив на базе Apache Hadoop, предназначенный для хранения и обработки слабоструктурированных и неструктурированных данных.	Arenadata QuickMarts (ADQM) — кластерная колоночная система управления базами данных, созданная на основе Yandex ClickHouse. С её помощью вы можете в режиме реального времени генерировать аналитические отчёты разного плана, используя большие объёмы информации, хранящейся в плоских витринах. ADQM многократно быстрее традиционных СУБД.	Arenadata Grid (ADG) — система распределённых вычислений в оперативной памяти, которая помогает реализовать real-time сценарии обработки данных в составе единой корпоративной платформы данных. ADG создана совместно с компанией Mail.Ru на базе технологии Tarantool.	Arenadata Streaming (ADS) — масштабируемая отказоустойчивая система для потоковой обработки данных в режиме реального времени, адаптированная для корпоративного использования и построенная на базе Apache Kafka и Apache Nifi.



Задачи для каждого компонента

₩AD.B	₩AD.H	₩AD.QM	₩ AD.G	₩AD.S
Хранилище данных (МРР-СУБД)	Хранение холодных и слабо структурированных данных	Витрины данных	In-Memory SQL	Streaming (ESB / MB / MQ)
Агепаdata DB лучше всего справляется с: • Сложными запросами, обрабатывающими большие объёмы, в том числе содержащими сложные аналитические функции • ETL/ELT • Эффективным соединением больших таблиц • Работой с индексами • Data Science • Аналитическими функциями на процедурных языках • Ad-hoc аналитикой	Построение распределённой платформы хранения на базе Apache Hadoop, адаптированный для корпоративного использования для холодных и слабо структурированных данных, которые хранить на МРР-СУБД или проприоритарных СУБД было бы слишком дорого.	Как использовать: В связке с Arenadata DB / Greenplum Аналитика над структурированными логами и событиями Быстрые витрины Распределённые вычисления для бизнеса и науки Анализ серверных логов	Система класса решений In- Memory Data Grid (IMDG). Она позволяет построить распределённый кэш в памяти и добавить туда функционал базы данных. При этом пользователю предоставляется возможность крайне быстро получать ответы на запросы по точечным фильтрам, буквально в течение миллисекунды.	 Используется как: Корпоративная шина обмена данными для всех приложений Среда консолидации больших потоков данных и управления ими в режиме реального времени Инструмент разграничения прав доступа к потокам данных Фреймворк для разработки потоковых аналитических приложений за считанные минуты в режиме реального времени и без единой строчки кода.



Преимущества Arenadata EDP



Хранение любых типов данных:

- структурированных
- неструктурированных



Обработка любого типа нагрузки на чтение и запись:

- OLAP
- OLTP
- Streaming



Поддержка гибридной ИТинфраструктуры или Multi-clouds



Единая точка входа в платформу для бОльшей части пользователей



Модульность



Горизонтальная масштабируемость всех компонент



Open Source





© ARENADATA

ВОЗМОЖНОСТИ
ПО ИНТЕГРАЦИИ









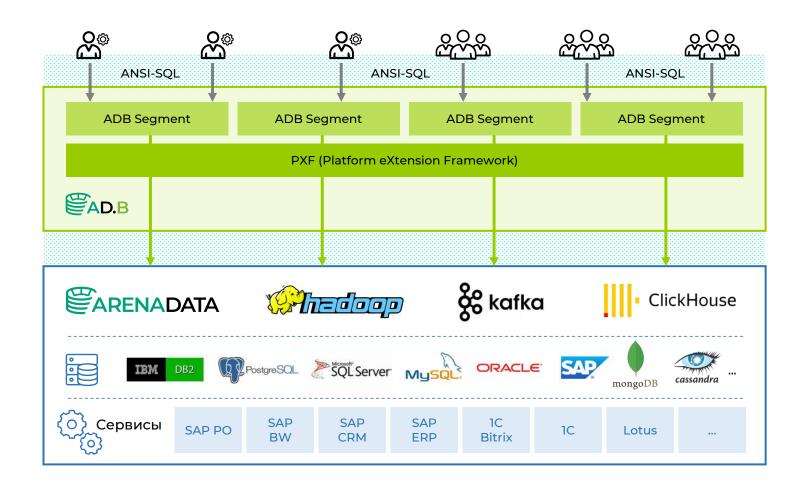


Platform Extension Framework (PXF)

Виртуальное хранилище

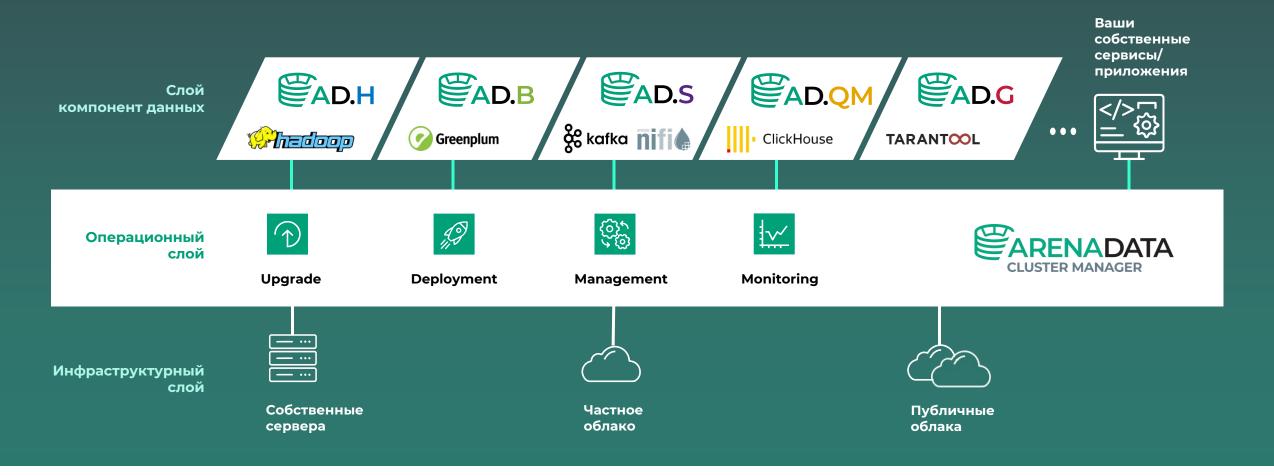
Возможности Виртуального хранилища на базе Arenadata Database (ADB):

- Поддержка работы с любыми JDBCсовместимыми СУБД
- Параллельная запись во внешние СУБД (Postgres, Oracle, MS SQL, SAP HANA и пр.)
- Передача предикатов (Predicate Pushdown)
- Батчинг данных при передаче
- Соединение (JOIN) данных из источников в real-time ресурсами ADB
- Ролевая модель, разграничение прав доступа
- Запуск Python-кода поверх всех данных всех источников
- Запуск предварительного кода во внешнем источнике (pre-SQL) скоро



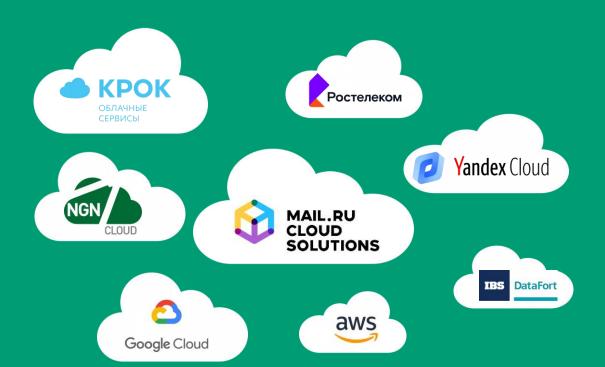


Гибридное корпоративное облако



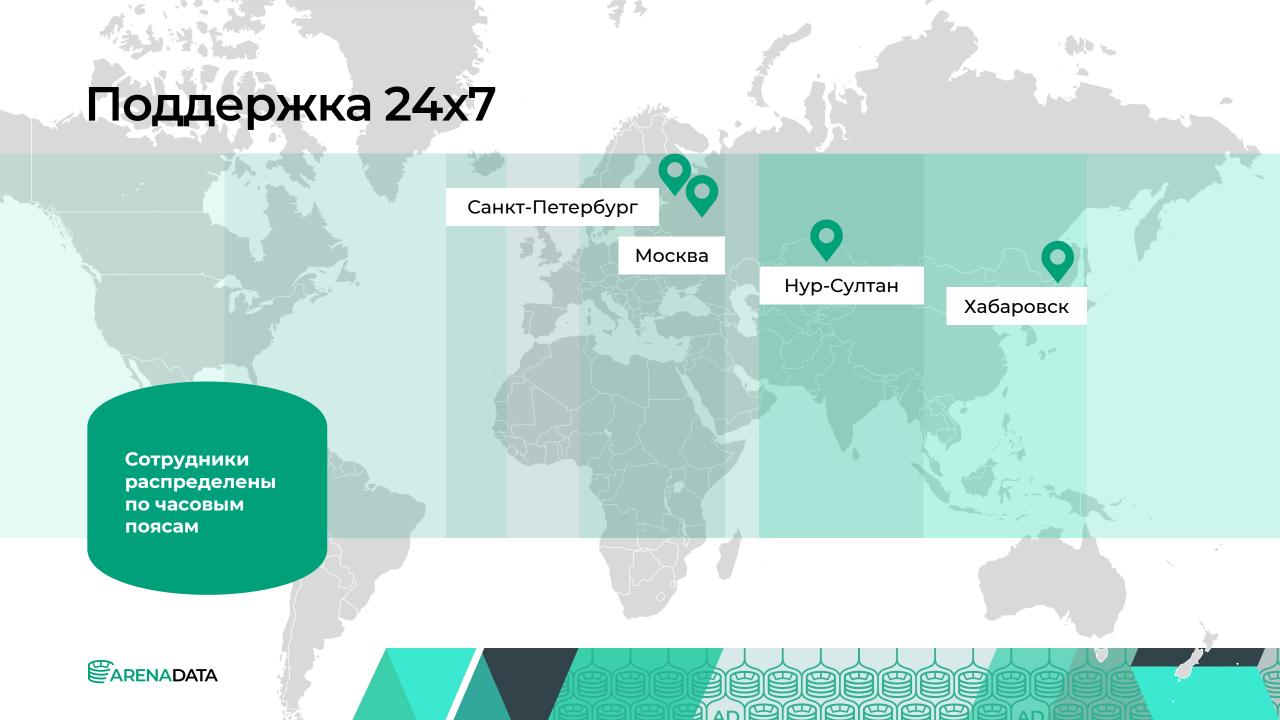


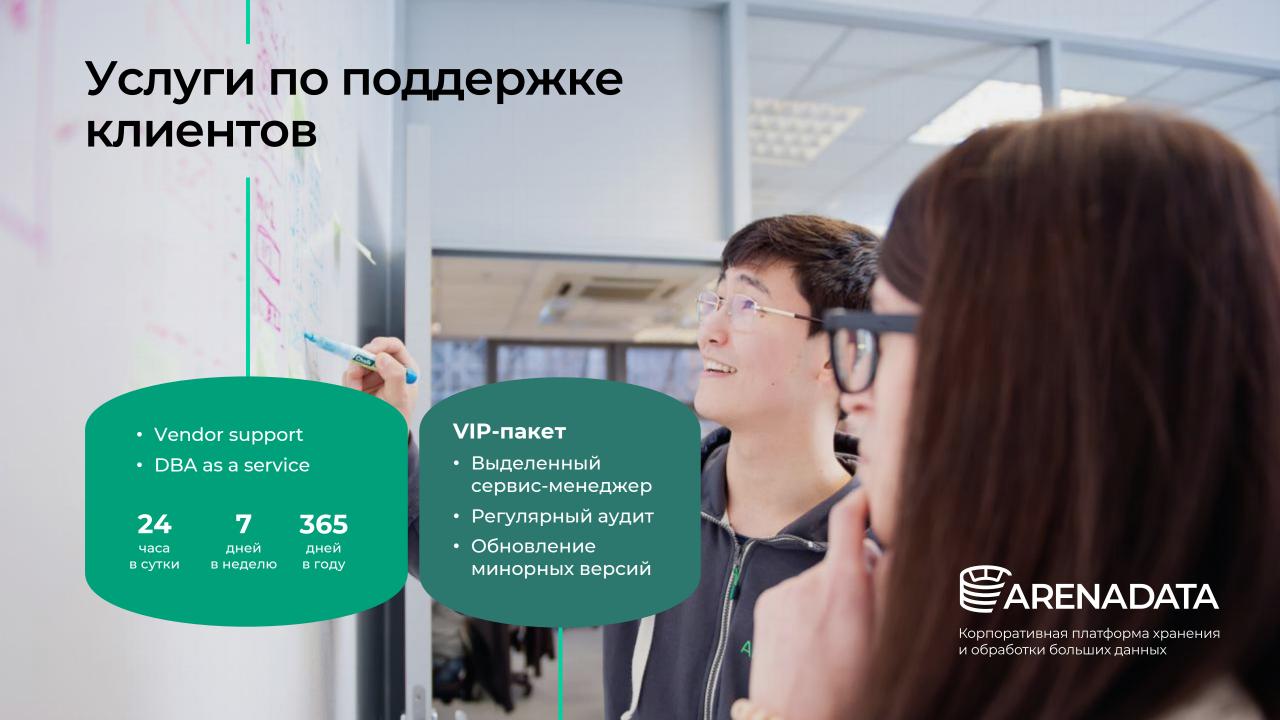
Enterprise Multi-cloud платформа данных



- Вы можете развернуть любой компонент платформы Arendata EDP как на bare-metal оборудовании, так и в облаке, или же воспользоваться востребованным сегодня сценарием гибридной ИТ-инфраструктурой или Multi-clouds.
- Установите EDP распределённо на разные инфраструктуры, обеспечив между ними сетевую доступность.







Некоторые проекты в ритейле





В ноябре 2017 года в X5 Retail Group были заложены основы для будущего технологического рывка.

«Главной целью для нас стало построение платформы, которая позволила бы собрать в одном месте все данные X5, структурировать и описать их, привязать к ним систему Data Governance, и с помощью этого инструмента начать менять культуру работы с данными внутри компании», — рассказал Андрей Молчанский. директор департамента разработки и сопровождения продуктов больших данных

Задача:

- построить платформу, которая позволила бы собрать в одном месте все данные компании, структурировать и описать их, привязать к ним систему Data Governance;
- добиться соответствия показателей и возможностей высоким требованиям, чтобы МРР база данных выдерживала высокую нагрузку от пользователей (2000 аналитиков и более 800 конкурентных запросов);
- с помощью полученного инструмента начать менять культуру работы с данными внутри компании.

Свежие овощи и фр







На базе Arenadata DB команда Big Data X5 начала строить аналитический слой детальных данных с описанной моделью, бизнес-владельцами доменов, правильно привязанной Data Governance и Classic Data-Dictionary. В результате у ритейлера получился надёжный сервис, доступный любому аналитику компании.

Реализация:

- X5 Retail Group использует Arenadata DB как приоритетный движок массовопараллельных расчётов и аналитики в рамках КХД.
- Также в X5 сейчас идёт перевод аналитических витрин данных на Clickhouse, лежащий в основе **Arenadata QuickMarts**.
- В качестве ETL/ELT-средства используется Apache NiFi, один из компонентов платформы транспорта данных **Arenadata Streaming**.





«Нам важно, чтобы используемые нами ВІинструменты давали возможность людям принимать быстрые решения на основе Big Data. И Arenadata DB — та самая база, с помощью которой мы могли недорого и эффективно реализовать первые кейсы», — считает Андрей Молчанский.

Результаты:

- X5 Retail Group получила надёжный сервис, доступный любому из 2000 аналитиков компании, для принятия быстрых решений на основе Big Data;
- С помощью Arenadata DB X5 планирует анализировать все основные домены данных в ритейле: товары, чеки, клиентов, контрагентов;
- Компании не отдаёт сторонним подрядчикам данные, поскольку вопрос сохранности и защищённости Від Data стоит для неё на первом месте.



Выбор экспертного совета GlobalCIO





=======



ПРОЕКТ



Сеть АЗС - розничные продажи

Розничные продажи в «Газпромнефть»:

- 1809 A3C;
- Выручка от реализации сопутствующих товаров и услуг сети АЗС «Газпромнефть» по итогам 2018 года составила более 17 млрд руб.;
- Более 29 млн чашек кофе.

Фудсервис и ритейл для заправок не профильный бизнес, но помогают увеличивать общую доходность топливных компаний.

Задача:

Дирекция региональных продаж ПАО «Газпром нефть» в конце 2017 года инициировала проект «Умное озеро данных» по внедрению комплексной платформы обработки и хранения данных с интегрированными компонентами Data Governance.

Одной из предпосылок проекта выступила потребность в надёжном хранилище «сырых» и первично интегрированных данных.







Сеть АЗС - розничные продажи

Первые задачи «умного озера данных» включили в себя обработку транзакций сети АЗС «Газпром нефть», расчёт сегментов для клиентской аналитики, анализ обратной связи от клиентов. В платформу были интегрированы данные значительного числа внешних источников - товарно-сырьевой биржи, Центробанка, географические и метеорологические ресурсы, метрики и отзывы Google, App Store, «Яндекс», открытые данные социальных сетей, различные данные партнёров и сведения о конкурентах, данные мобильных приложений.

Реализация:

«Умное озеро данных» включает в себя

- подсистемы Data Governance (каталог данных, управление качеством данных, бизнесглоссарий, пользовательский портал обзора данных),
- компоненты интеграции данных,
- а также интегрированные с ними компоненты обработки «больших данных» (Data Lake на основе **Arenadata Hadoop**),
- аналитическое хранилище данных, BI,
- аналитическую лабораторию с набором инструментов Data Science.







Сеть АЗС - розничные продажи

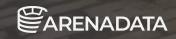
Проект создания платформы («Умное озеро данных») реализовал уникальный функционал и получил награду «Проект года» Global CIO в номинации «Аналитические решения и Big Data» за 2019 год и отдельного внимания Gartner.

Результаты:

- На текущий момент около всех 50% аналитических проектов и инициатив Дирекции региональных продаж ПАО «Газпром нефть» уже реализуется в контуре «умного озера данных», план на 2020 год охватить не менее 75%.
- Пользователями «озера данных» являются аналитики бизнес-подразделений в 29 регионах России и четырёх странах СНГ.
- Кроме этого, созданное решение является поставщиком данных для различных систем в контуре Компании и в целевом виде для внешних партнеров.



Лучший проект в области Big Data







CEANLANDA MA

ПРОЕКТ



В поисках подходящего инструмента, новая команда определила для себя следующие критерии:

- МРР-база с возможностями горизонтального масштабирования и полноценного распараллеливания;
- Разумная стоимость владения в пересчёте на терабайт;
- Классический SQL-синтаксис;
- Приемлемое время отклика.

Реализация:

- Команда «М.Видео-Эльдорадо» остановилась на технологии Greenplum. Ритейлер сразу отказался от Community Edition, поскольку посчитал такой подход рискованным, да и не соответствующим специфике компании, которая не склонна расширять свою деятельность для дополнительной загрузки имеющихся мощностей или активов.
- Выбирая конкретный продукт, ритейлер рассмотрел ПО от двух разработчиков Pivotal и Arenadata. В результате проведённого конкурса группа «М.Видео-Эльдорадо» предпочла продукт от Arenadata Arenadata DB.







В перспективе «М.Видео-Эльдорадо» планирует провести перевнедрение управления маркетинговыми кампаниями, используя Arenadata DB как бэкэнд. Раньше для этого служил BW on HANA. В новой версии Campaign Management затронет как кампании, которые запускаются в Ad hoc режиме раз в сутки руками бизнеспользователей, так и автоматические, более частые рассылки, которые запускаются раз в 10 минут.

Результат:

- По сравнению со старой «песочницей» ритейлера, **Arenadata DB** работает в разы быстрее;
- Команда заказчика может оперативно выполнять запросы, работать с большим количеством источников данных, а сами SQLзапросы составлять без значительного overhead'a;
- Группа «М.Видео-Эльдорадо» теперь способна производить вычисления и предобработку с такой глубиной исторических данных, которые было невозможно выполнять на старой базе. Например, на сбор витрины Customer 360 теперь уходит не 12 часов, а всего лишь 2 часа.



Мнение клиентов



Дмитрий Алексеев заместитель директора департамента технологического развития общебанковских систем ВТБ



Андрей Молчанский директор департамента разработки и сопровождения продуктов больших данных X5 Retail Group



Сергей Носов директор по управлению данными ПАО Ростелеком



Иван
Черницын
руководитель
Центра
аналитических
решений
Дирекции
региональных
продаж ПАО
«Газпром нефть»



Все текущие хранилища данных банка будут поэтапно мигрировать на новое целевое единое хранилище, построенное на базе ядра розничного хранилища. При этом текущую технологическую платформу заменит отечественное решение от Arenadata.





Нам нравится то, что сотрудничество с Arenadata лежит в плоскости совместного развития. Если у нашего бизнеса появляется необходимость в новом инструменте, например, мониторинге запросов и отслеживании их выполнения, то мы просто обсуждаем с коллегами из Arenadata, когда он появится в road map.



~

Посредством перевода централизованного хранилища на стек Open Source удалось сэкономить средства, сравнимые с годовым бюджетом построения хранилищ данных в крупных компаниях. С учетом требований, предписывающих компаниям с государственным участием переходить на отечественное ПО, это достижение особенно ценно.



((

Взаимодействие с командой Arenadata было весьма эффективным. Специалисты компании помогали с вопросами внедрения и безопасности, предлагали собственные решения и оперативно выпускали релизы с учётом специфических требований заказчика.





Наши преимущества

Продукты Arenadata помогают ритейлу построить цифровую платформу, предоставляющую возможность обрабатывать поступающие данные, анализировать их, строить разнообразные аналитические модели, снижать трудовые и временные затраты ресурсов на выполнение рутинных операций.

Опыт реализации проектов миграции с западных СУБД и legacy-систем для крупнейших российских компаний и госорганов

Arenadata выбирают лидеры различных индустриальных сегментов российской экономики, являющихся флагманами цифровизации.

Снижение до 50% затрат на лицензирование и поддержку ПО

Открытый исходный код, возможность доработки ПО

Соблюдение требований о хранении и обработке персональных данных (ФЗ -152)

Среди наших партнеров — ведущие системные интеграторы, облачные провайдеры, разработчики ПО и оборудования

3 шага к цифровизации с Arenadata

Расскажите нам о задачах, ИТ-инфраструктуре и получите рекомендации по компонентам платформы и выбору партнёра.

Загрузите бесплатно необходимые пакеты для платформы данных с нашего сайта.

Разверните нашу платформу в своей инфраструктуре и отправьте нам технические и функциональные требования, чтобы получить индивидуальное предложение.





