



Корпоративная платформа хранения
и обработки больших данных

Импортозамещение иностраннных СУБД и систем хранения данных

Импортозамещение иностранного ПО

В России активно развивается политика импортозамещения иностранного программного обеспечения на отечественные решения.

29 июня 2017 года Приказом Минкомсвязи России №334 были утверждены методические рекомендации по переходу ФОИВ и государственных внебюджетных фондов на отечественное ПО.

В дальнейшем были выпущены другие законодательные акты, регулирующие переход на российское ПО региональных органов власти и компаний с государственным участием в капитале.

Импортозамещение с Arenadata

ARENADATA — ведущий российский разработчик платформ сбора и хранения данных предлагает вариант импортозамещения программного обеспечения на собственные решения, превосходящие по многим параметрам лучшие иностранные предложения.

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



Большие
данные



Интернет
вещей



Машинное
обучение



Связанные
технологии

Факты об ARENADATA

Специалисты из топовых data-driven компаний:

Pivotal™

Yandex



Microsoft



Tinkoff



@ mail.ru
group

20+

Лидеров РФ выбрали нас
в качестве центральной
платформы данных

25 ПБ

Суммарный объем
данных, который наши
клиенты хранят в
платформе Arenadata



Развитая сеть партнёров



Партнёрство
с ведущими вендорами
data-решений

Российское программное обеспечение

Регистрация в
Реестре
Министерства
Связи

Отсутствие
валютных рисков -
цены в рублях

Центры разработки
в РФ - Москва,
Санкт-Петербург,
Хабаровск

Поддержка,
документация и
учебные курсы на
русском языке





Наше программное обеспечение

- имеет свидетельства о государственной регистрации программ для ЭВМ,
- включено в Единый реестр российских программ для электронных вычислительных машин и баз данных,
- и в начале 2021 года мы ожидаем получение сертификата ФСТЭК на Arenadata DB, и далее на остальные продукты.



Корпоративная платформа хранения
и обработки больших данных

Открытый исходный код

- Вносим вклад в проекты Apache Software Foundation:
 - Apache Ambari
 - Apache Bigtop
 - Apache PXF
- 2^{ой} в мире контрибьютор Greenplum Database
- Активный контрибьютор ClickHouse и Tarantool
- Являемся членами ODPi (Linux Foundation) с 2015 года наряду с другими крупными компаниями

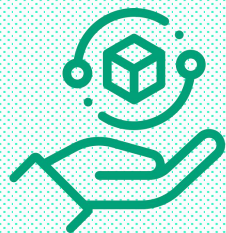
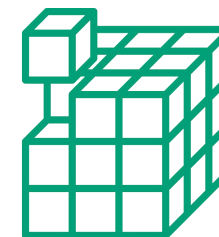


Система лицензирования

Каждый продукт Arenadata представляет собой бандл, включающий собственные разработки и основной OpenSource проект.

Программное обеспечение доступно в двух версиях:

- бесплатной (Community)
- платной (Enterprise)



Версии отличаются друг от друга комплектацией бандлов. Пользователям бесплатного пакета доступно ядро проекта и небольшая часть собственных доработок Arenadata. Enterprise-версия включает максимально полное решение, созданное нашими специалистами.

Система лицензирования достаточно гибкая, и даёт возможность найти индивидуальный подход к каждому клиенту.

Arenadata EDP — платформа для цифровизации

Экспертиза зарекомендовавших себя партнёров-интеграторов, бизнес-кейсы и прикладное ПО, доказавшее свою эффективность

Оцифрованные бизнес-процессы

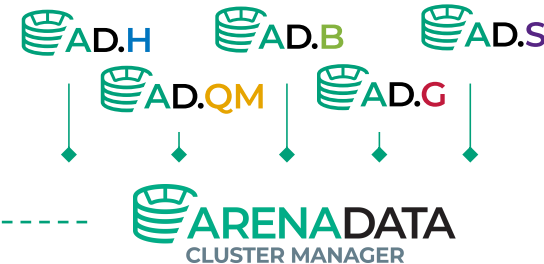


Множество инструментов для решения прикладных задач



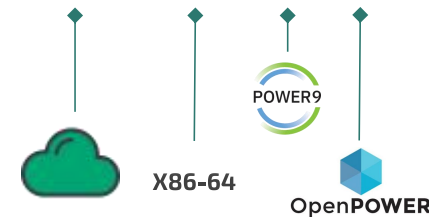
Arenadata Enterprise Data Platform (EDP) с нативной интеграцией сервисов и компонент, централизованным мониторингом и управлением кластерами

Платформа данных



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

Эластичная инфраструктура



Enterprise Data Platform — решаемые задачи



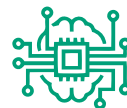
Корпоративные хранилища больших данных

- 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)

- Импортзамещающие 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

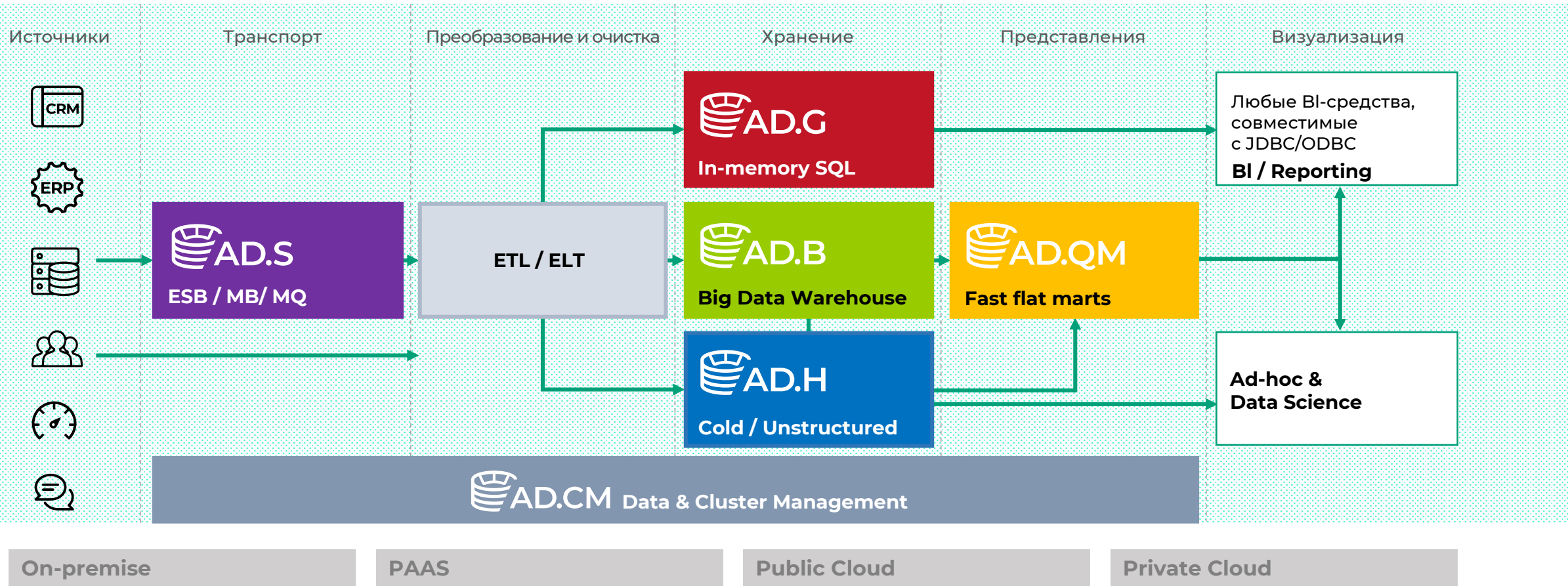
- Поддержка R, Python и Madlib.
- Distributed ML на базе Spark.
- Возможность эффективной эксплуатации моделей в SQL-среде с помощью встроенных функций Madlib.



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.

Современная платформа данных



Компоненты платформы EDP

 AD.B	 AD.H	 AD.QM	 AD.G	 AD.S
Хранилище данных (MPP-СУБД)	Хранение холодных и слабо структурированных данных	Витрины данных	In-Memory SQL	Streaming (ESB / MB / MQ)
Arenadata 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 (MPP-СУБД)	Хранение холодных и слабо структурированных данных	Витрины данных	In-Memory SQL	Streaming (ESB / MB / MQ)
Arenadata 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 MPP-СУБД или проприетарных 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.

Что внутри?

 AD.B	 Greenplum
 AD.H	
 AD.QM	 ClickHouse
 AD.G	
 AD.S	
 AD.CM	 ANSIBLE

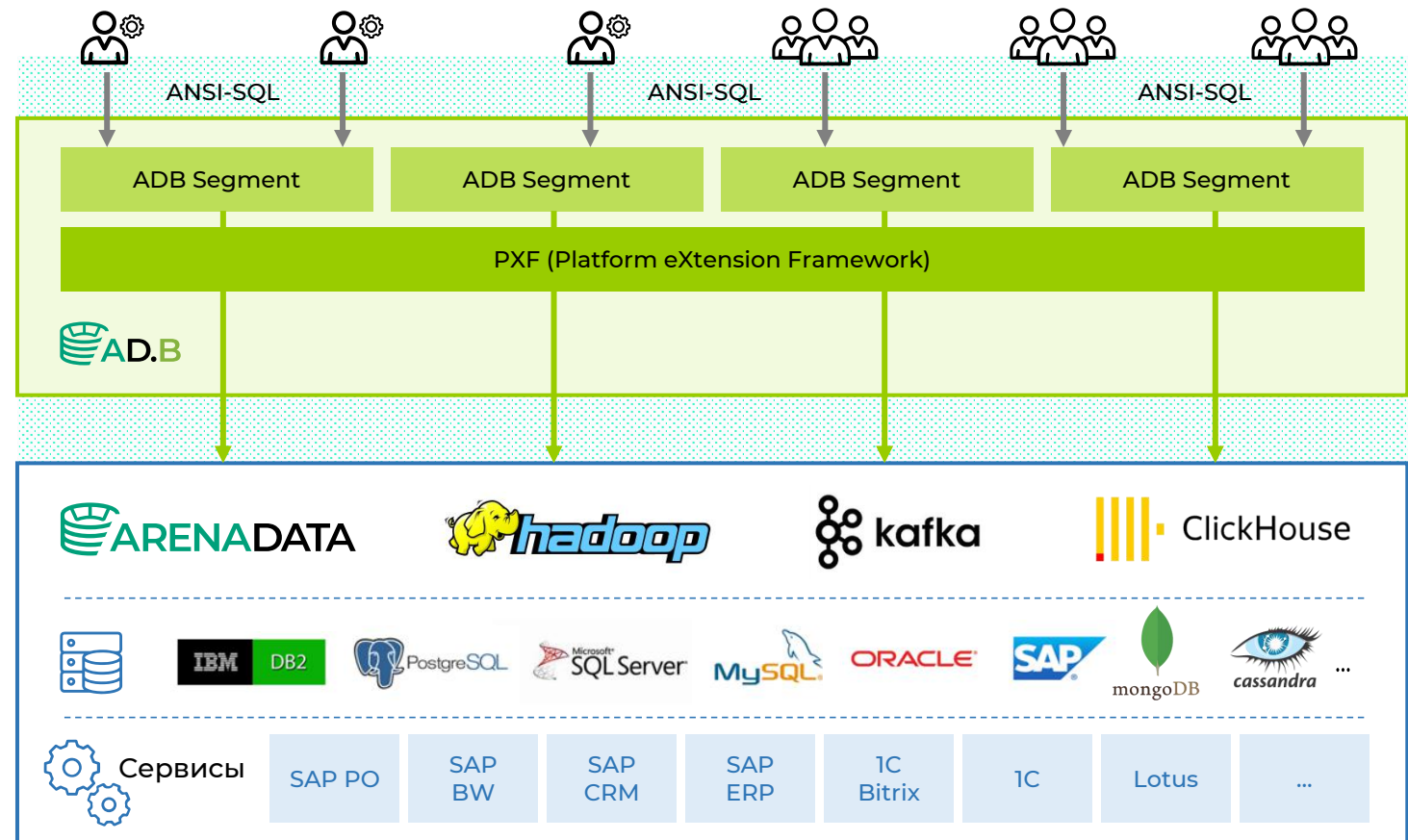
Каждый добавляемый компонент платформы доработан, чтобы стать её частью.

Производится:

- Анализ кода
- Интеграция с модулем мониторинга
- Проверки отказоустойчивости
- Подбор компетентного персонала поддержки
- Доработка возможностей параллельной интеграции
- Множественные прочие проверки и доработки

Управление федерацией данных с PXF

- Единая точка входа для всех SQL-запросов
- Специализированный движок виртуализации данных
- Межкомпонентный массово-параллельный проброс предикатов запроса
- В том числе во внешние системы (SAP HANA, Oracle DB, MS SQL, PostgreSQL, etc)
- Кластеризация, шардирование и локальность данных в каждой компоненте хранения.



Преимущества Arenadata EDP



Хранение любых типов данных:

- структурированных
- неструктурированных



Обработка любого типа нагрузки на чтение и запись:

- OLAP
- OLTP
- Streaming



Поддержка гибридной ИТ-инфраструктуры или Multi-clouds



Единая точка входа в платформу для большей части пользователей



Модульность



Горизонтальная масштабируемость всех компонент



Open Source

Отличия Arenadata EDP от Open Source сборок

Выстроенная экосистема данных
с единым мониторингом и управлением всеми кластерами платформы

Модульность и заменяемость компонент платформы
для максимальной простоты встраивания в существующий ландшафт

Наличие компонент под любые задачи

Единая точка входа для SQL-запросов
как следствие нативной виртуализации данных и параллельной pushdown-оптимизации между всеми компонентами хранения

ГАРАНТИЯ ЦЕЛОСТНОГО РЕШЕНИЯ:

Параллельная интеграция компонент

Нативная отказоустойчивость всех компонент

Удобная в использовании и эффективная параллельная интеграция

Линейная масштабируемость как следствие кластерности решения, а также шардирования и локальности данных

Покрытие любых регламентов нагрузки и любых типов данных

ARENADATA

Скажите «ДА» современной российской платформе!



31 12 2021

CLUDERA

ORACLE®

teradata.

Exasol



Скажите «ДО СВИДАНИЯ»

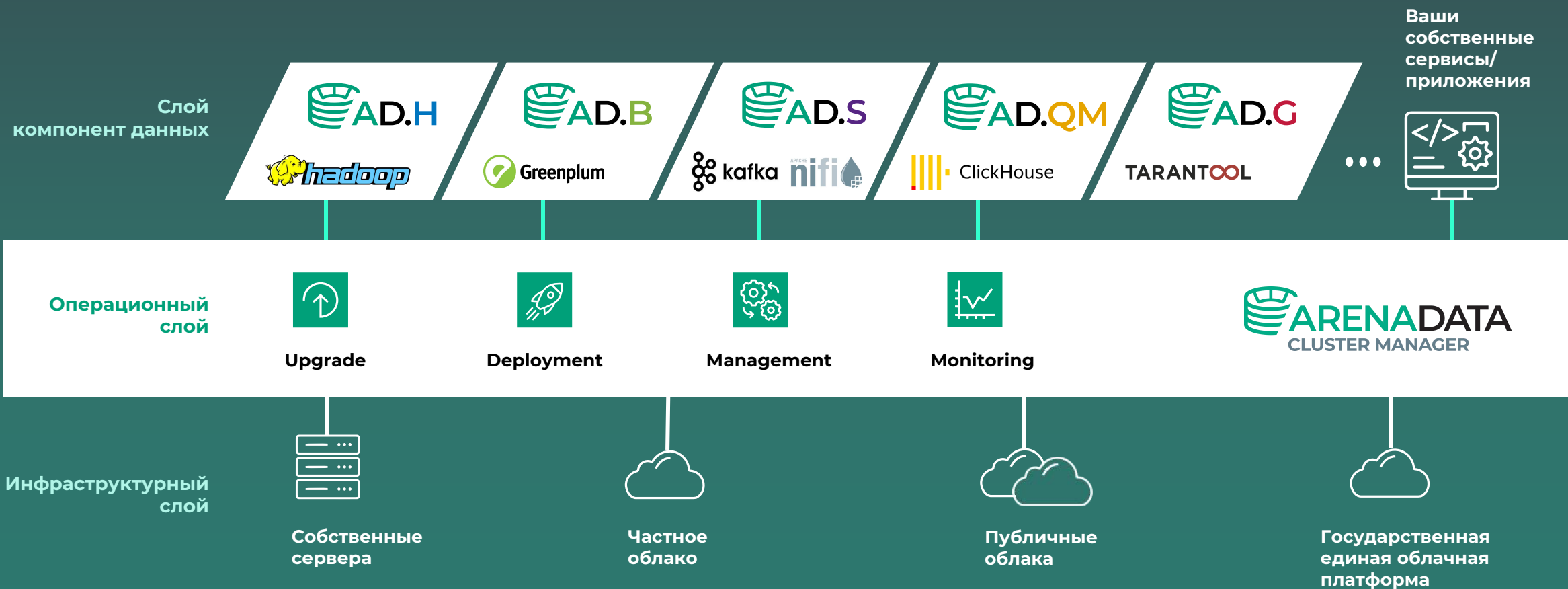
- западным закрытым системам
- монолитной архитектуре
- vendor lock-in
- высокой TCO



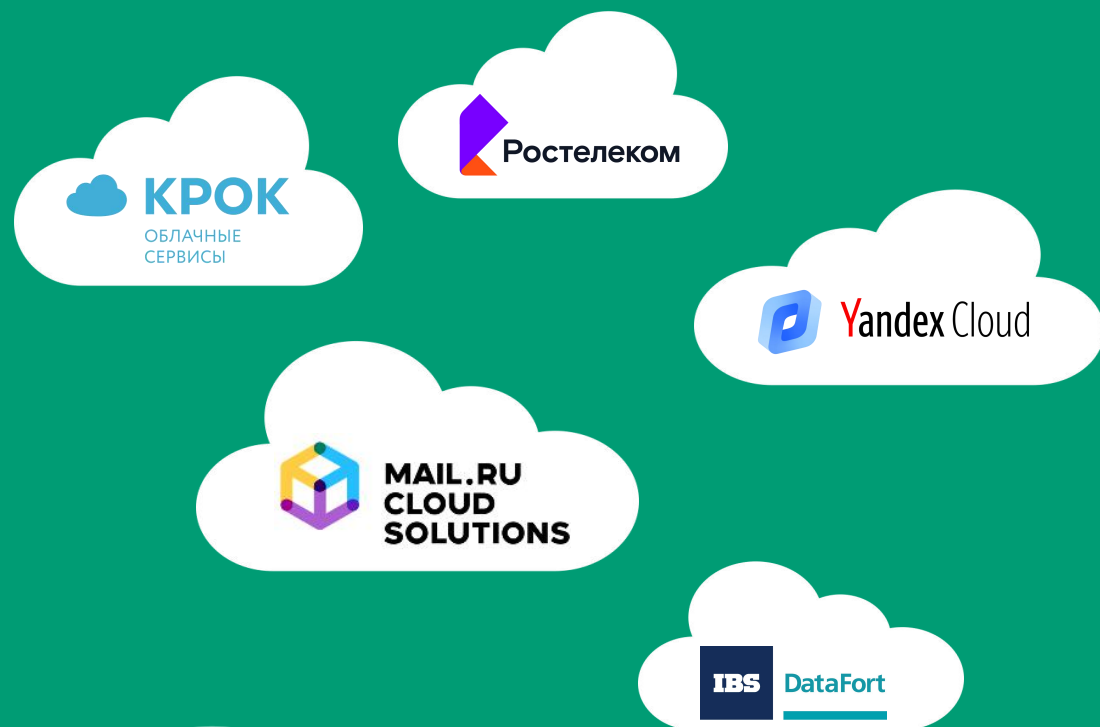
ARENADATA

Мы обладаем опытом
миграции клиентов с
других систем

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 — гетерогенной ИТ-инфраструктурой или Multi-clouds.
- Установите EDP распределённо на разные инфраструктуры, обеспечив между ними сетевую доступность.

Вендорский консалтинг и профессиональные сервисы



Архитектурный контроль

Мы предоставляем проверенную вендорскую точку зрения на проект внедрения, позволяющую принять правильное архитектурное решение в непростых ситуациях.



Аудит цифрового ландшафта

Позволяет ИТ и менеджменту научиться отвечать на вопрос «Как должен измениться ландшафт данных и цифровых сервисов, чтобы обеспечить мои бизнес-потребности сейчас и в будущем?».



DBA as a service

Каждой СУБД необходим сведущий DBA. Достаточно обратиться к нам как к вендору СУБД, готовому предоставить услугу DBAaaS.



Smart Start

Это комплекс услуг по установке и настройке платформы Arenadata и интеграции наших продуктов в инфраструктуру данных заказчика. Мы также проводим инструктаж для администраторов, и выполняем ряд не столь значительных, но полезных вспомогательных действий.



Разгрузка дорогостоящих СУБД

Как вендор, мы поможем вам снизить ТСО владения проприетарными СУБД за счёт внедрения и интеграции аналитических MPP-систем под ключ.

Поддержка 24x7

Санкт-Петербург

Москва

Алма-Аты

Хабаровск

Сотрудники
распределены
по часовым
поясам

Услуги по поддержке клиентов

- Vendor support
- DBA as a service

24

часа
в сутки

7

дней
в неделю

365

дней
в году

VIP-пакет

- Выделенный сервис-менеджер
- Регулярный аудит
- Обновление минорных версий



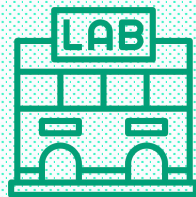
Корпоративная платформа хранения
и обработки больших данных

Курсы по продуктам Arenadata



Знания из первых рук

Наши преподаватели — практикующие специалисты, имеющие за плечами богатую экспертизу и значительное количество реализованных бизнес-кейсов.



Только 40% теории

Все курсы построены на реальных кейсах и практических примерах, а лабораторные работы проходят на стендах, симулирующих рабочую среду.



Сертификация

Пройдя обучение на курсах Arenadata и успешно сдав экзамен, слушатель станет обладателем именного сертификата, подтверждающего полученные знания.

Нам доверяют



Истории успеха



Организация «Умного озера данных» для комплексной системы анализа и управления данными Дирекции региональных продаж на базе Arenadata Hadoop.



Создание единой технологической платформы, способной объединить хранилища данных, создать систему отчётности и аналитики по миллионам клиентов, сокращение TCO на базе массово-параллельной СУБД Arenadata DB и Arenadata Streaming.



Создание платформы данных по всем данным компании с привязкой к системе Data Governance, способной выдерживать высокую нагрузку от пользователей (2000 аналитиков и более 800 конкурентных запросов) с помощью Arenadata DB, Arenadata QuickMarts и Arenadata Streaming.



Построение «Озера данных» для осуществления внешнего государственного аудита на новом качественном уровне. Используются данные всех возможных типов из совершенно разных источников, перечень которых заранее не определён. «Озеро данных» построено с помощью Arenadata Hadoop и Arenadata DB.



Разработка цифровых сервисов и моделей машинного обучения, создание удобного инструмента для вычислений и анализа, а также оптимизация производственных процессов и снижение себестоимости готовой продукции на базе Arenadata Hadoop.



Создание аналитического хранилища, отдельного от системы поддержки управленческой отчётности, объединение трёх «песочниц» и их миграция с SAP HANA на Arenadata DB.

ПРОЕКТ



Счётная палата РФ с 2018 года реализует приоритетные направления развития, заложенные в её Стратегии на 2018-2024 годы, на новом качественном уровне на основе цифровизации информационного обеспечения своей деятельности по осуществлению внешнего государственного аудита (контроля). Особое внимание Счётная палата уделяет задачам создания цифровой инфраструктуры для поддержки аудита и аналитической деятельности. Решать их Счётной палате помогает Arenadata.

Задача:

Создать программно-аппаратный комплекс, предоставляющий возможности дата-аналитикам иметь в своём распоряжении необходимые данные и на их основании с помощью инструментария Цифровой платформы обрабатывать их, получая

- риск-ориентированные модели,
- аналитические модели,
- визуальные инструменты.

ПРОЕКТ



Счётная палата не планировала строить громоздкую реляционную супер-базу, поскольку ей приходится оперировать данными из совершенно разных источников, не совместимых между собой. Этому способствовало и отсутствие требований к очень большой оперативности результатов. В результате построено хранилище «сырых» данных на Arenadata Hadoop, в которое загружаются копии источников, преобразованные в формат JSON.

Реализация:

В связи с лицензионными рисками использование западного проприетарного ПО было не желательно, а ПО Open Source не всегда полностью удовлетворяет требованиям.

В результате Счётная Палата РФ выбрала зарекомендовавшие себя **Arenadata DB** для

- массово-параллельных расчётов,
- хранения витрин данных
- и промежуточных таблиц,
- а также **Arenadata Hadoop** для создания «Озера данных».

ПРОЕКТ



Масштаб проекта Счётной палаты занял 36 000 человеко-часов, и в его рамках удалось сократить трудозатраты на создание отчётов по некоторым объектам контроля примерно на 70%. Благодаря «озеру данных» Счётная палата ведёт пилотный проект «Цифровой департамент» и создаёт различные аналитические модели.

Проект Счётной палаты стал победителем в 2019 году в номинации «Лучшее цифровое решение для трансформации в госсекторе» от GlobalCIO.

Результаты:

С помощью Цифровой платформы

- расширили список используемых для аудита источников данных и повысили их качество,
- применили современные методы управления информацией,
- снизили трудоемкость традиционных видов аудита
- и обеспечили развитие стратегического аудита за счет создания инструментов риск-ориентированных и аналитических моделей и применения современных методов предиктивной аналитики.



ПРОЕКТ



Каждая из заявленных целей в рамках такой крупной компании, как Ростелеком, уже претендует на статус полномасштабного отдельного проекта. Объединение этих целей в один проект со сжатыми сроками делает задачу поистине амбициозной и трудновыполнимой.

Клиентами являются представители всех блоков компании. На платформе представлены данные из более, чем 200 СИ. Общий объем данных достигает 2ПБ, количество объектов исчисляется сотнями тысяч.

Задача:

- Объединение федеральных ХД организацией информационно-аналитического пространства на базе единой отказоустойчивой масштабируемой платформы
- Исполнение директивы Правительства по импортозамещению смещением баланса в сторону отечественного ПО, собственных разработок и opensource
- Сокращение TCO/TCC систем отчетности и аналитики

ПРОЕКТ



На начало проекта существовал целый ряд разрозненных ХД и систем отчетности федерального и регионального уровней развиваемых и сопровождаемых хаотичным образом отдельными внутренними и внешними командами. ХД отличаются стеком технологий, подходами работы с СИ, нет согласованной методологии расчета показателей, единого глоссария. Некоторые ХД при этом представляют сложные программные комплексы, работая с большим количеством источников (> 100).

Реализация:

Хранилище данных Ростелеком построено на базе **Arenadata DB** и **Arenadata Hadoop**.

Отчётность и аналитика по миллионам клиентов крупнейшего в России телекоммуникационного оператора строится на единственной отечественной массово-параллельной СУБД.

Как и многие другие наши клиенты, в качестве ETL/ELT-средства коллеги используют Apache NiFi, включаемый нами в платформу транспорта данных **Arenadata Streaming**.

ПРОЕКТ



Единая технологическая платформа «Ростелекома» официально введена в промышленную эксплуатацию в 2019 году. В рамках реализации проекта было автоматизировано 3000 рабочих мест, задействовано 500 абонентов и потрачено 115000 человеко-часов. Сегодня платформа используется сотрудниками без исключения во всех регионах России.

Результаты:

Платформа подготовлена к росту объема данных, числа пользователей BI, витрин, песочниц. Развернуты дополнительно несколько контуров (DEV/TEST), что с внедрением релизного процесса, практик CI/CD позволяет повысить качество конечных продуктов.

Переход на российское ПО и OpenSource сократили требуемый объем закупки лицензий проприетарных продуктов в несколько раз, что уменьшило бюджет проекта на миллионы долларов, позволило нивелировать корпоративные и санкционные риски.

Мнение клиентов



**Дмитрий
Алексеев**
заместитель
директора
департамента
технологического
развития
общественных
систем ВТБ



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.



**Михаил
Петров**
директор
департамента
Цифровой
трансформации
Счётной палаты РФ



С Arenadata мы построили продуктивное сотрудничество. Нам крайне радует, что рядом всегда есть специалисты, обладающие обширной экспертизой и высокой квалификацией.



**Сергей
Носов**
директор по
управлению
данными
ПАО Ростелеком



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.



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



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 государственному сектору 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 промышленных 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

Наши партнёры

ВНЕДРЕНИЕ

accenture

КРОК

IBS

Jet

DIS
GROUP

АПЛАНА
ЦЕНТР РАЗРАБОТКИ

IBS

Platformix

ИТ ФИЛОСОФИЯ

POLYMEDIA



отр



ЛАНИТ

sapiens
solutions



quadrium

Rubbles
SBDA Group

LUXMS
группа компаний

ARENADATA

ОБЛАЧНЫЕ ПРОВАЙДЕРЫ

MAIL.RU
CLOUD
SOLUTIONS

Yandex Cloud

Ростелеком

КРОК
ОБЛАКО

IBS DataFort

ТЕХНОЛОГИЧЕСКИЕ ПАРТНЁРЫ

accenture digital

@ mail.ru
group

Яндекс

YADRO

IBM

Informatica

POLY//ATICA

VISIOLOGY

LUXMS
группа компаний

Форсайт
Аналитическая платформа

3 шага к импортозамещению с Arenadata

1

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

2

Загрузите бесплатно необходимые пакеты для платформы данных с нашего сайта.

3

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





Корпоративная платформа хранения
и обработки больших данных

Оцените потенциал Big Data и Open Source вместе с Arenadata Enterprise Data Platform

- Узнайте больше на arenadata.tech
- Скачайте бесплатно на store.arenadata.io
- Получите консультацию по почте info@arenadata.io

