

ARENADATA

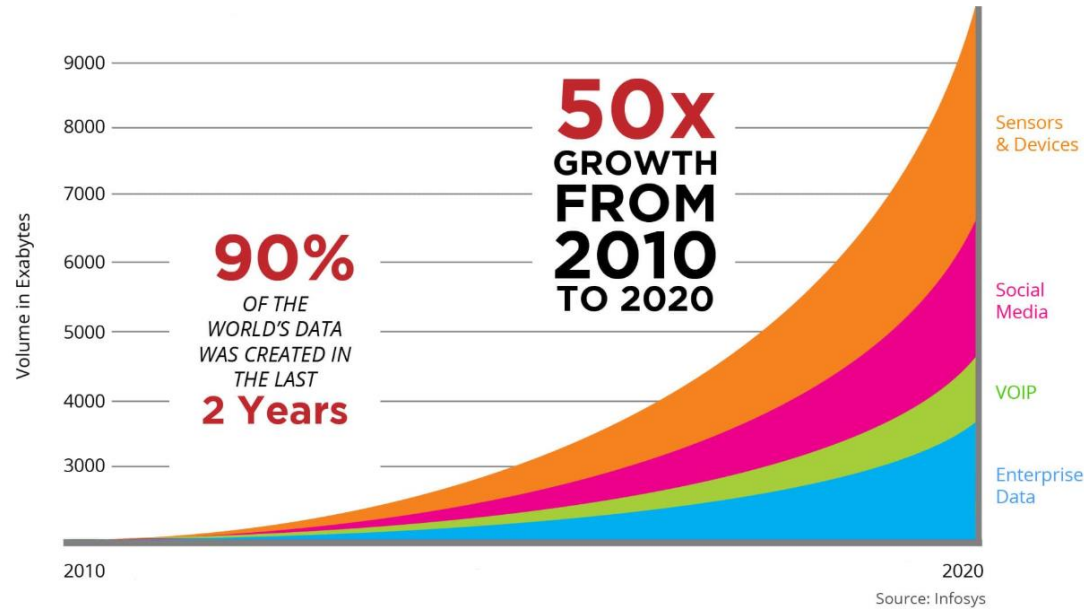
Многофункциональная платформа
данных. От стартапа до стандарта

Сергей Золотарев
Управляющий партнер

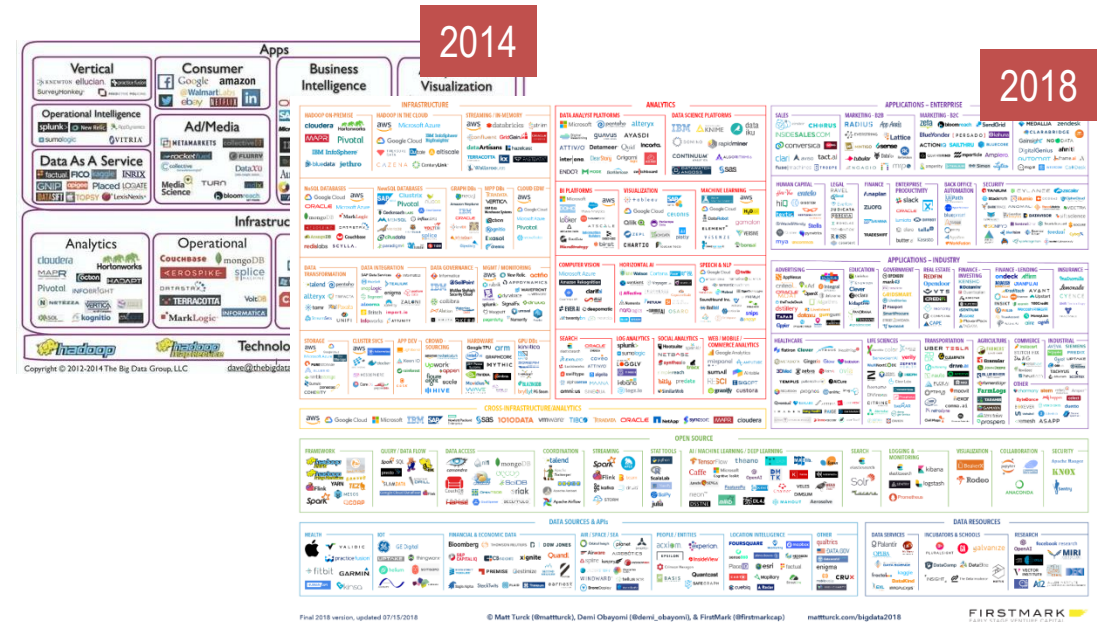
Москва 12 марта

Новый рынок платформ данных

Революция данных

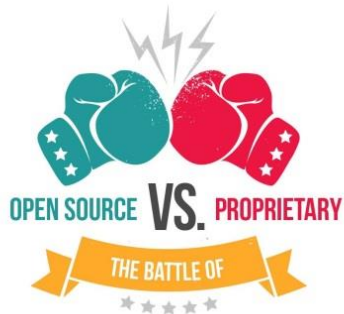


Революция платформ



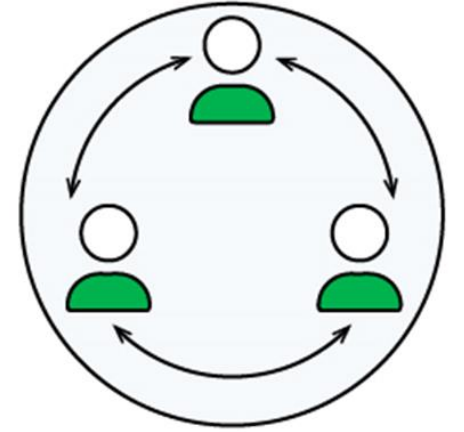
Изменение правил

- Вместо монолитных систем – интегрированные программные платформы
- Вместо нескольких приложений с закрытой архитектурой – сотни open source проектов для решения конкретных задач
- Вместо закрытых и «неповоротливых» монстров-производителей – сотни мобильных и динамичных команд разработчиков, открыто обменивающихся идеями, наработками, кодом



Представление команды

- Arenadata основана в конце 2015 года;
- Простая и амбициозная идея – создать корпоративную платформу данных на базе Open Source проектов
- Ядро команды **Pivotal / EMC Greenplum** ;
- Многолетний опыт в классических КХД;
- Крупнейшие на рынке России и СНГ проекты по проектированию/построению/аудиту платформ данных на технологиях MPP и Hadoop;
- Разработчики из самых инновационных ИТ компаний:



Pivotal™

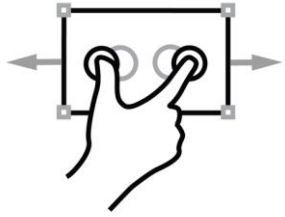


Yandex



Рамблер/

Основные характеристики современной платформы данных



Эластичная и горизонтально-масштабируемая архитектура



Сниженная стоимость обработки и хранения без изменения архитектуры при многократном росте объемов



Различные типы и форматы данных



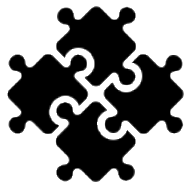
Сейчас нельзя предсказать, какие типы данных понадобятся в перспективе, надо уметь работать со всеми



Решение с открытым кодом



Снижение vendor lock-in, гибкость и кастомизация под специфику заказчика



Интегрируемость и модульность компонентов платформы



Эффективное взаимодействие компонентов платформы /
Прозрачный обмен данными между компонентами



Единая система управления и мониторинга



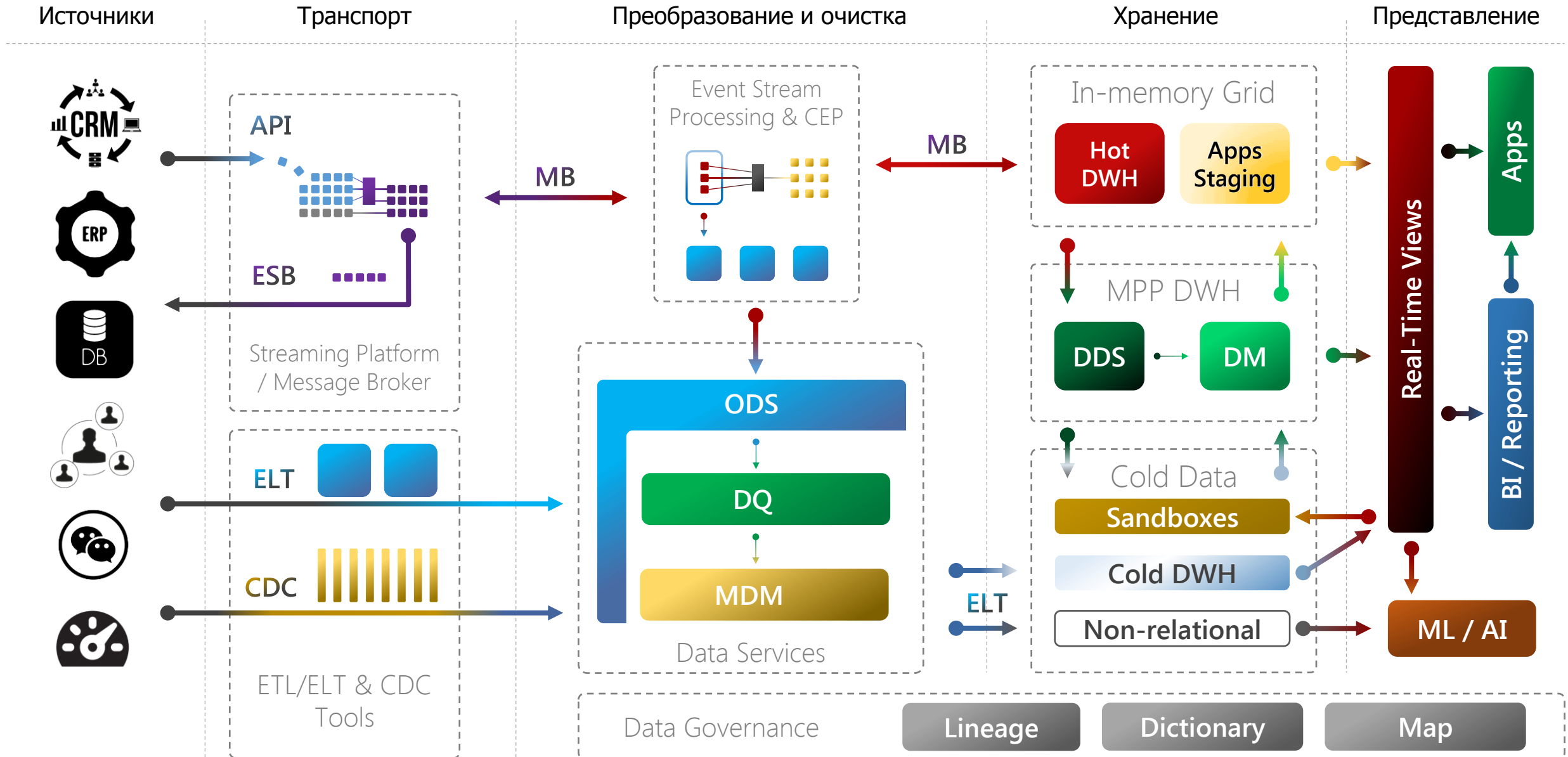
Снижение трудозатрат по мониторингу и эксплуатации платформы

Основные подходы к построению платформ данных

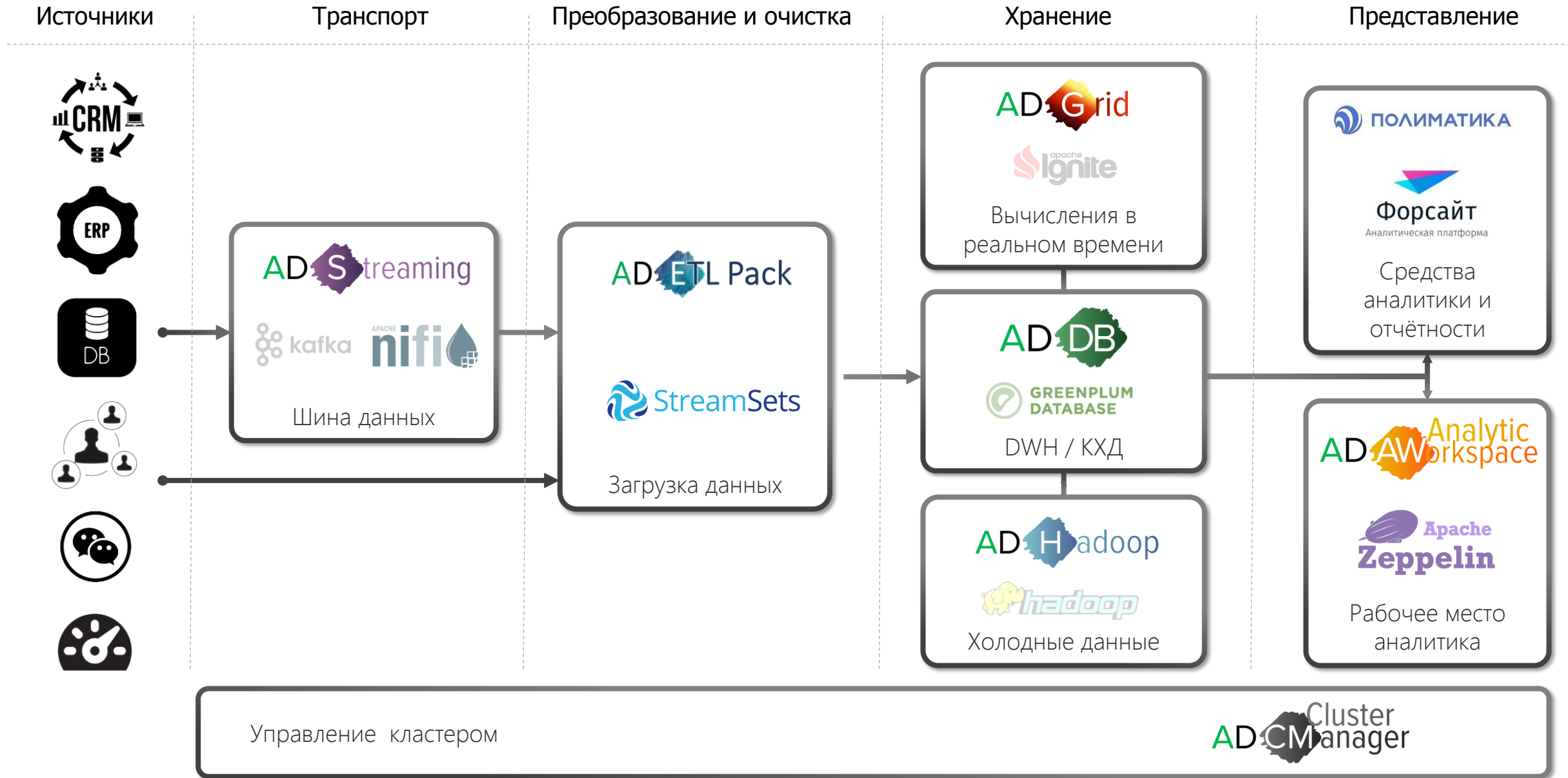
Платформы данных эволюционировали из классических хранилищ данных (DWH), «новые» архитектуры создавались, чтобы решать нетипичные для DWH задачи в нетипичных условиях. Ключевые характеристики больших данных были сформулированы в виде «Четырёх V»: качественных метрик, различие в которых в основном определяет подход к построению архитектуры Big Data.

4 V's / Концепт	Хранилище данных →	«Озеро Данных» →	«Лямбда Архитектура» ←	«Каппа Архитектура»
Тип нагрузки Velocity	<ul style="list-style-type: none"> • Batch • Micro-Batch 	<ul style="list-style-type: none"> • Batch • Micro-Batch • Mega-Batch 	<ul style="list-style-type: none"> • Batch • Micro-Batch • Near-Real Time 	<ul style="list-style-type: none"> • Near-Real Time • Streaming
Типы данных Variety	Структурированные	<ul style="list-style-type: none"> • Структурированные • Квази-структурированные • Неструктурированные 	<ul style="list-style-type: none"> • Структурированные • Квази-структурированные • Неструктурированные 	<ul style="list-style-type: none"> • Структурированные • Квази-структурированные
Качество данных Veracity	Очищенные данные	<ul style="list-style-type: none"> • Сырые данные • Очищенные данные 	<ul style="list-style-type: none"> • Сырые данные • Очищенные данные 	Сырые данные
Глубина истории Volume	2 дня – 5 лет	> 5 лет	От ~0 сек до ∞	< 24 часов
Прикладные задачи	<ul style="list-style-type: none"> • Единая отчетность • Базовая аналитика 	<ul style="list-style-type: none"> • Холодное хранение • Аналитические песочницы 	<ul style="list-style-type: none"> • Машинное обучение • Принятие решений 	<ul style="list-style-type: none"> • Complex event processing • Internet of Things

Типичная современная платформа данных

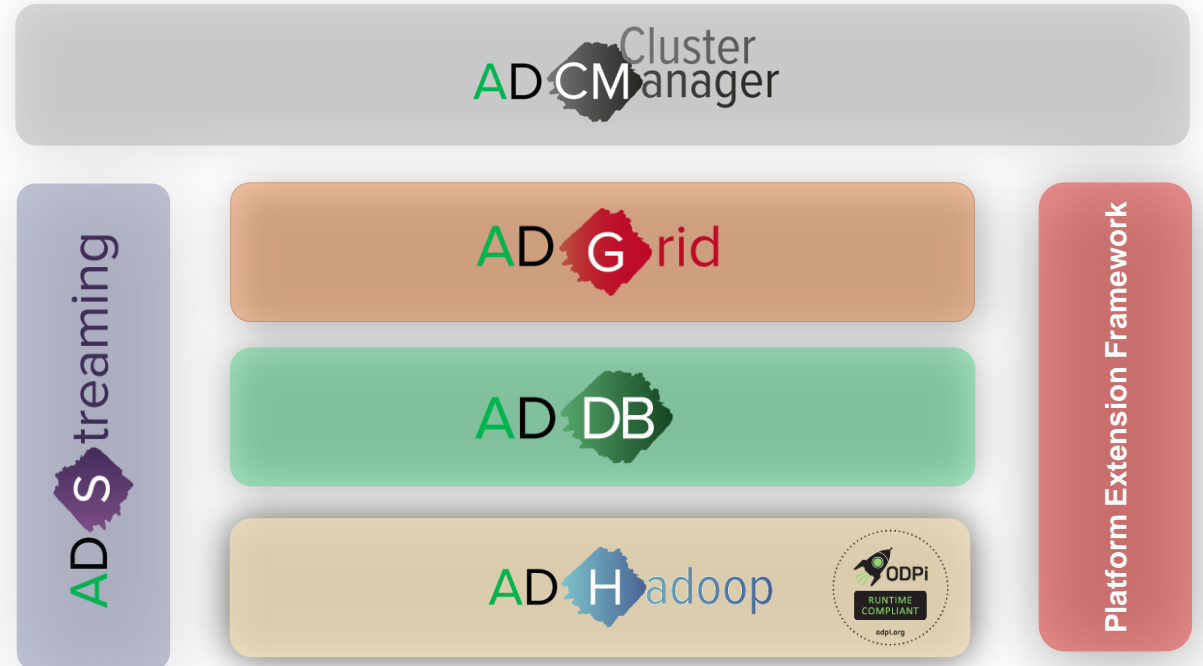


Многофункциональная платформа данных | Arenadata EDP

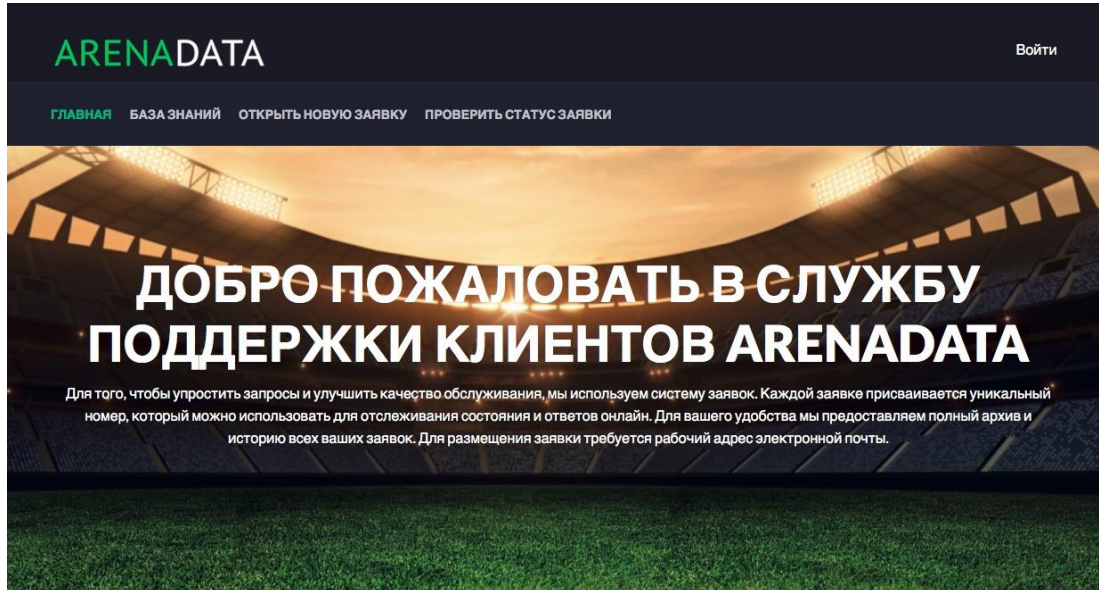


Почему Arenadata Enterprise Data Platform

- Для каждого типа нагрузки используется специально разработанный компонент;
- Параллельная интеграция всех компонентов;
- Линейное масштабирование всех компонентов;
- Единая точка входа обращений к платформе;
- Единое ПО управления;
- Отказоустойчивость всех компонентов;
- Все типы нагрузки;
- Отечественное ПО (реестр МинКомСвязи);
- Простая интеграция со внешними системами;
- Open-source.



Техническая поддержка и сопровождение платформы



Центры технической поддержки 24x7:

- Москва;
- Санкт-Петербург;
- Хабаровск;
- Алматы.

ARENADATA AMBARI

Разделы документации:

- Инструкция по установке Ambari
- Руководство администратора
- Настройка безопасности Ambari
- Информация о релизе

ARENADATA DB

Разделы документации:

- Архитектура ADB
- Инструкция по установке
- Расчет полезной емкости диска
- Руководство администратора
- Рекомендации по работе с ADB

ARENADATA HADOOP

Разделы документации:

- Инструкция по установке Hadoop
- Настройка безопасности Hadoop
- Руководство администратора HDFS
- Информация о релизе

ANALYTIC WORKSPACE

Разделы документации:

- Установка сервиса
- Руководство администратора
- Руководство пользователя

Вендорский консалтинг и профессиональные сервисы

Smart Start

Интеграция продуктов Arenadata в действующую инфраструктуру данных

- Выдача рекомендаций по подготовке вычислительной инфраструктуры: сайзинг, параметры настройки операционной системы, требования к сети;
- Проверка выделенной инфраструктуры;
- Установка ПО Arenadata;
- Проверка работоспособности;
- Настройка базовой функциональности системы управления и мониторинга кластера;
- Настройка базовых пользовательских ролей и прав доступа;
- Тестовая загрузка образцов данных из источников;
- Помощь и консультации в подключении аналитических инструментов;
- 16 человек-часов консультаций наших экспертов по работе с данными;
- Практический семинар (workshop) для администраторов по сопровождению кластера.

Information Security

Комплексное обеспечение информационной безопасности

- Аутентификация пользователей с помощью корпоративной службы каталогов (Active Directory \ LDAP);
- Настройки авторизации на основе ролевых моделей;
- Настройки аудита событий ИБ;
- Разработка инструкции администратора ИБ;
- Обеспечение perimeter security кластера ADH с помощью Knox;
- Настройка защищенных подключений к платформе;
- Настройка защищенных соединений между компонентами платформы и узлами кластеров;
- Настройка шифрования хранимых данных;
- Интеграция с Kerberos;
- Разработка прочей документации ИБ.

Data Governance Consulting

Фиксация подходов к управлению данными

- Рекомендации экспертов по формированию концепции управления данными;
- Детальная проработка десяти кейсов нашими экспертами;
- Рекомендации по документарному оформлению Data Governance;
- Рекомендации по профилированию данных;
- Фиксация основных ролей и соответствующих им сотрудников;
- Выбор решения для автоматизированного управления данными;
- 16 человек-часов консультаций наших экспертов;
- Практический семинар (workshop) для администраторов по управлению данными.

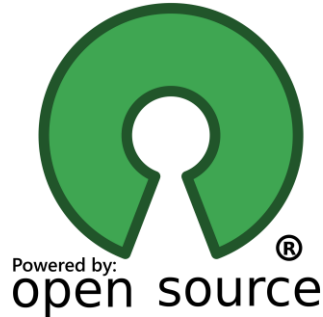
Аудит и архитектурный контроль

Помощь наших архитекторов на проектах прикладных внедрений

- Обследование источников данных;
- Обследование потребителей данных;
- Обследование интеграционной инфраструктуры;
- Помощь в постановке задачи подрядчику;
- Помощь в определении метрик оценки объема проекта;
- Контроль планирования проекта;
- Архитектурный контроль исполнения проекта, в том числе анализ и обсуждение подходов, функциональных схем и конкретных артефактов проектирования;
- Оценка экспертизы сотрудников подрядчика (по желанию заказчика);
- Подготовка заключений по итогам аудита / арх. контроля.

Arenadata Hadoop (ADH):

- Корпоративный дистрибутив платформы хранения данных Apache Hadoop;
- Построен полностью на открытых компонентах проекта Apache Software Foundation, лицензия Apache Software License 2.0;
- Российское программное обеспечение, регистрация в Реестре Министерства Связи;
- Сертифицирован в 2016 году по спецификации и требованиям ODPI (международное сообщество разработчиков проектов хранения данных, часть Linux Foundation).



Поддержка 24x7 и 8x5.
On site или удаленно



Русский язык
документации и
поддержки.
Рублевые цены



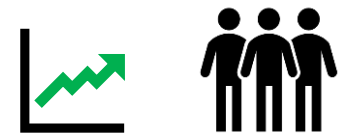
ПАК для x86 / Power



Любой формат
установки ПО



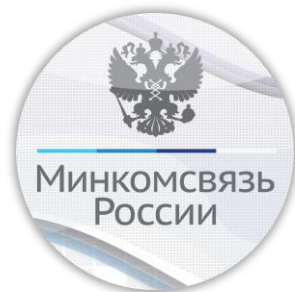
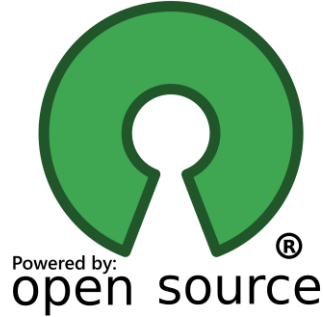
Возможность доработки и
сборки под заказчика



Профессиональные
сервисы

Arenadata DB (ADB):

- Массово-параллельная аналитическая СУБД;
- Apache Software License 2.0;
- Российское ПО, регистрация в Реестре Министерства Связи;
- Ядро платформы Arenadata EDP;
- Оптимальное решение для детального слоя хранилища данных (DDS, Detailed Data Store).



Поддержка 24x7 и 8x5. On site или удалённо.

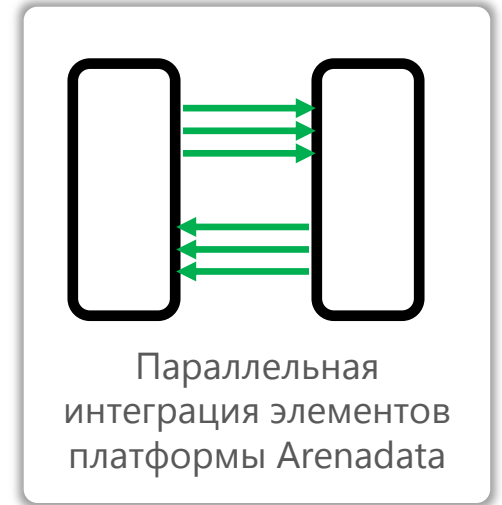
Русский язык документации и поддержки. Рублёвые цены.

ПАК для x86 / Power

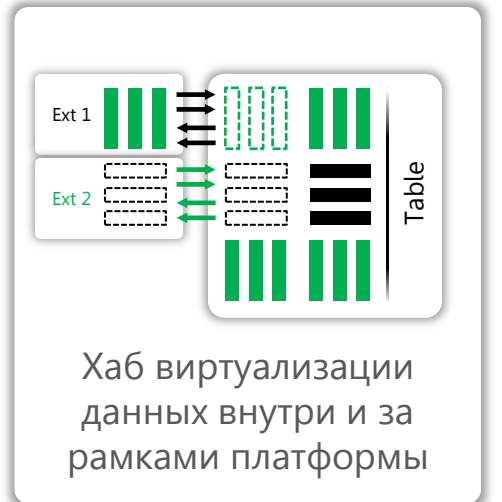
Любой формат установки ПО

Возможность доработки и сборки под заказчика

Профессиональные сервисы

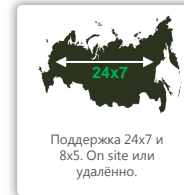
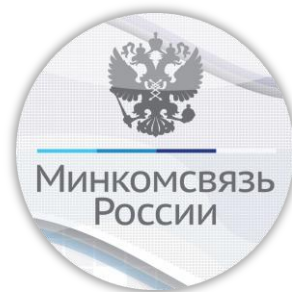
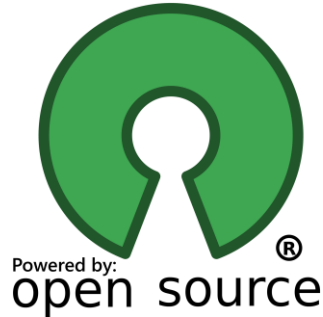


ANSI SQL 2008, колоночное или строковое хранение

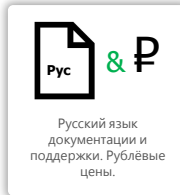


Arenadata Grid (ADG):

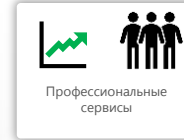
- Линейно-масштабируемая кластерная СУБД для вычислений в оперативной памяти;
- Apache Software License 2.0;
- Российское ПО, регистрация в Реестре Министерства Связи;
- Прорывная функциональность для in-memory решений:
 - ANSI SQL 99 с поддержкой ACID;
 - Работа с данными как в режиме key-value, так и с помощью SQL;
 - Все на Java: просто разрабатывать и поддерживать прикладные решения;
 - Persistence, партиционирование и другие редкие для in-memory СУБД опции.



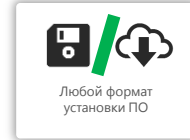
Поддержка 24x7 и 8x5. On site или удалённо.



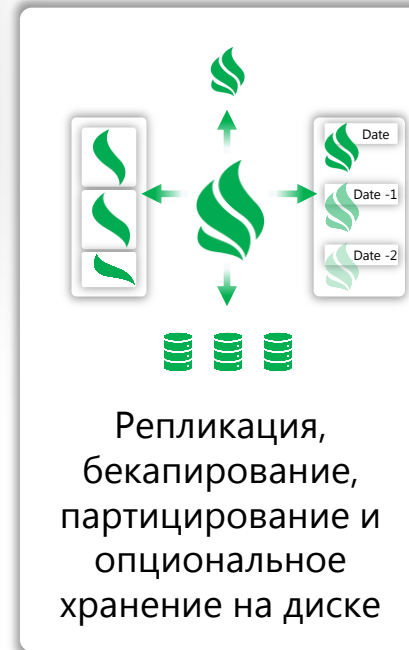
Русский язык документации и поддержки. Рублёвые цены.



Профессиональные сервисы

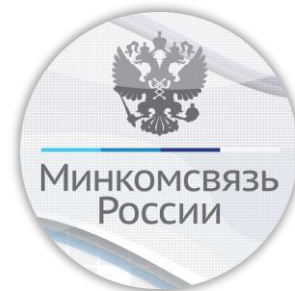
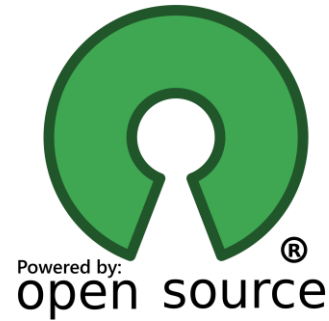


Любой формат установки ПО



Arenadata Streaming (ADS):

- Распределенная платформа потоковой обработки данных на базе Apache Kafka и Apache NiFi, адаптированная для корпоративного использования;
- Apache Software License 2.0;
- Российское программное обеспечение, регистрация в Реестре Министерства Связи запланирована на 2019;
- 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 обработки данных в реальном времени. Позволяет делать это в визуальном интерфейсе за считанные минуты.



На регистрации



Поддержка 24x7 и 8x5. On site или удаленно



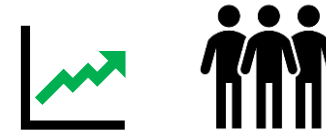
Русский язык документации и поддержки. Рублевые цены



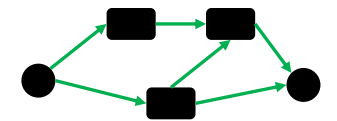
Доработка и сборка под заказчика



Любой формат установки ПО

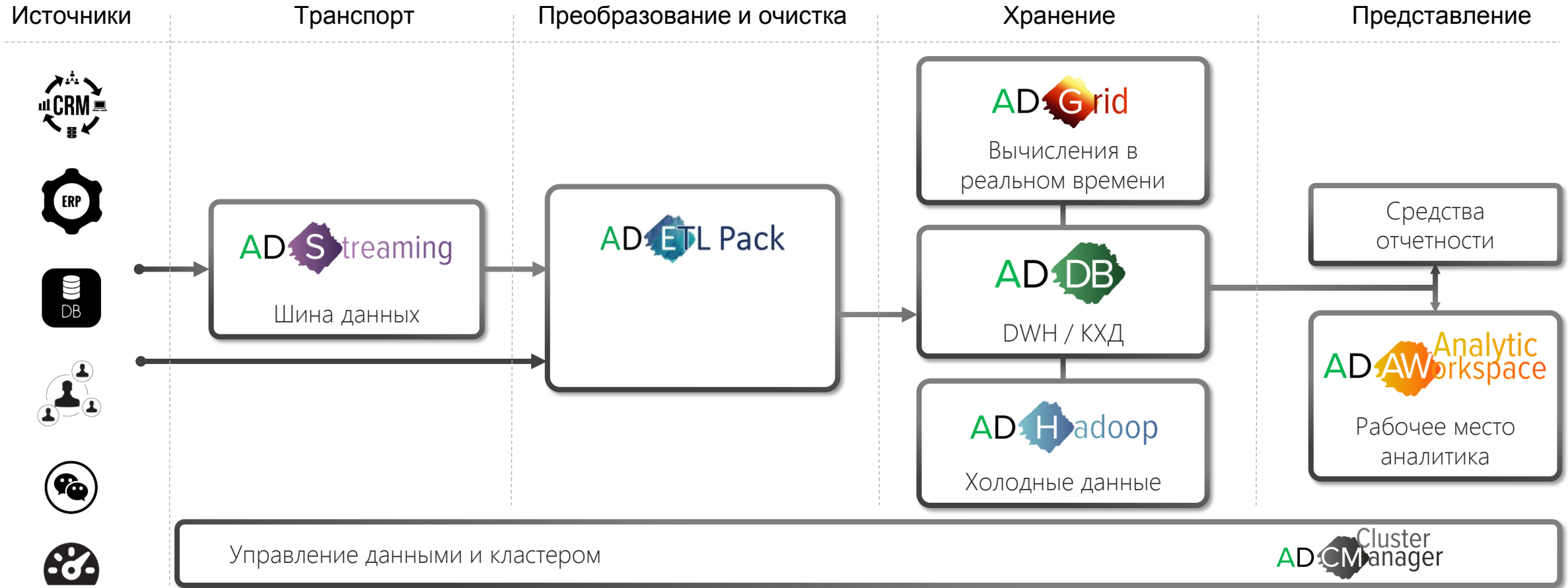


Профессиональные сервисы



Визуальный интерфейс

Многофункциональная платформа данных | Arenadata EDP



x86-64



On-premise

Public Cloud

PaaS

Private Cloud

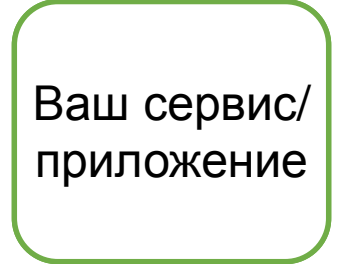
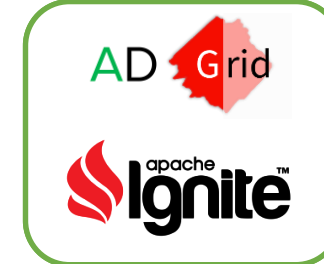


Яндекс Облако

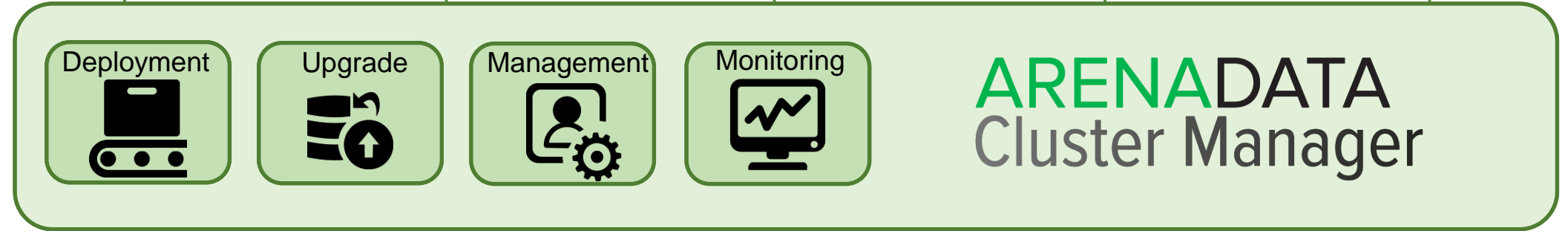


Гибкость платформы Arenadata

Уровень
Приложений &
Сервисов



Операционный
уровень



Уровень
инфраструктуры



Разворачивается на ИТ-Инфраструктуре любого типа

Готова для интеграции сторонних приложений

Arenadata – наша миссия

The background image shows a modern airport terminal with multiple escalators. A directional sign is visible, pointing left to 'Bat Cave' and right to 'Concourse A Terminal', 'Luggage Claim', and 'Hotel'. The sign also includes Japanese text: 'ターミナル', '手荷物受取所', and 'ホテル'. The overall scene is brightly lit with natural light coming from the right.

Предоставить эффективный инструмент развёртывания и управления всеми data-сервисами компании независимо от используемой инфраструктуры

Мы достигаем этого за счёт разделения data-сервисов и инфраструктуры на независимые слои

Благодаря этому становится возможным использовать достоинства всех типов современной инфраструктуры

Развитие экосистемы

Продукты



Сейчас

II-III кв.

III-IV кв.



ARENA DATA