



Корпоративная платформа хранения
и обработки больших данных

Миграция с Hortonworks

на Arenadata Hadoop

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



Факты об ARENADATA



Специалисты из топовых data-driven компаний:

Pivotal

Yandex



Microsoft



Tinkoff

MIRANTIS

@ mail.ru
group

20+

Отраслевых лидеров РФ
выбрали нас в качестве
центральной платформы
данных

25 ПБ

Суммарный объём
данных, который наши
клиенты хранят в
платформе Arenadata



Развитая сеть партнёров
как в России, так и за
рубежом



Партнёрство
с ведущими вендорами
data-решений



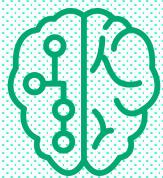
Открытый исходный код

- 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
- Выпускаем **собственное ПО** в Community-редакциях



open source

Время Hortonworks подошло к концу



Пришёл момент задуматься:

компания Cloudera, которая приобрела Hortonworks, 31 декабря 2021 года прекращает поддержку как самого продукта, так и бесплатной версии системы управления Ambari. Пора задуматься о миграции и начать процесс.

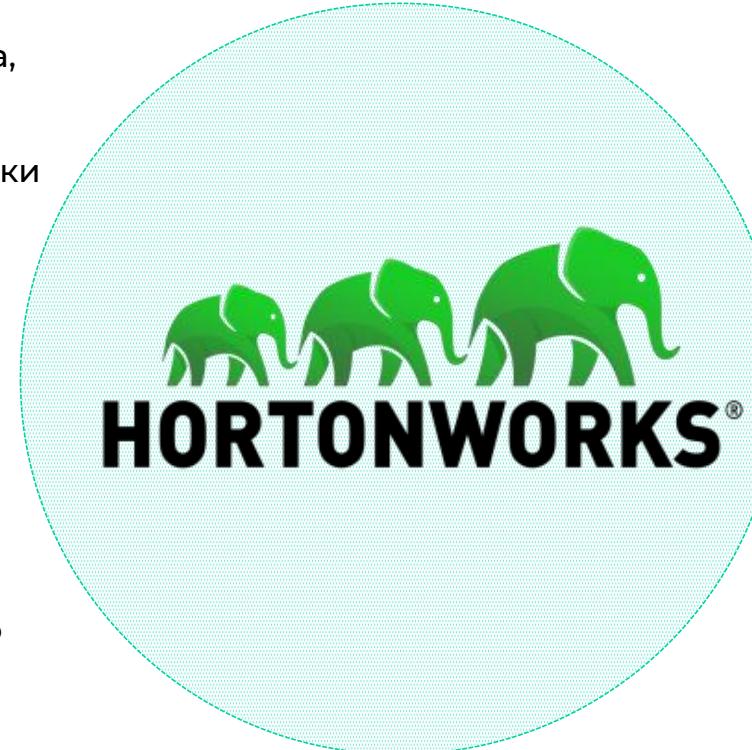
Альтернативы для миграции

CLOUDERA

- 1 Мигрировать на решение от Cloudera, что приведёт к незапланированным финансовым тратам и адаптации к продукту с совершенно другой, с точки зрения концепции, идеологии компании.



- 2 Перейти на открытую «ванильную» сборку Apache Hadoop и другие открытые компоненты экосистемы Hadoop, например Spark и Hbase. Это потребует обеспечить полный цикл сборки продукта и его поддержки, а также разработать собственную систему управления сервисами.



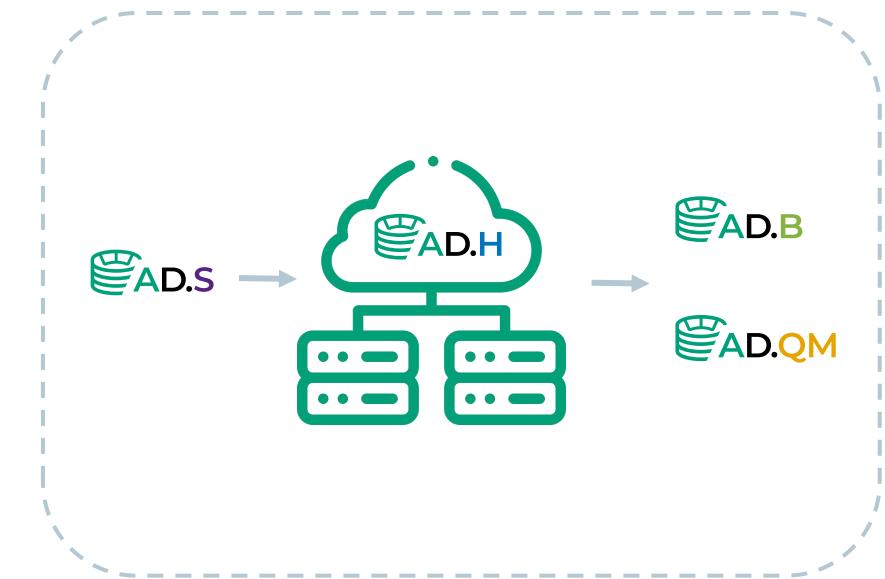
- 3 Мигрировать на Arenadata Hadoop, чей дистрибутив весьма близок решениям Hortonworks.



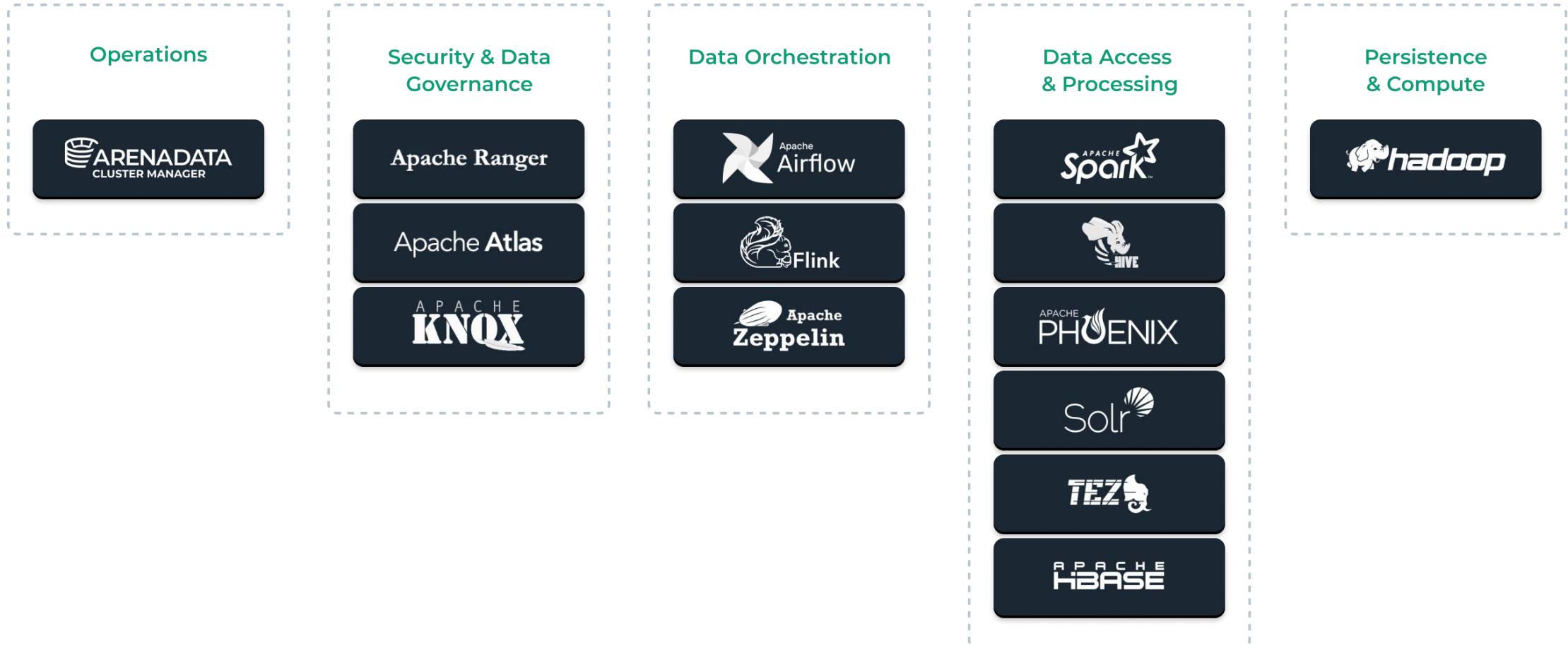
Arenadata Hadoop (ADH) — это полноценный дистрибутив на базе Apache Hadoop, адаптированный для корпоративного использования, предназначенный для хранения и обработки слабоструктурированных и неструктурированных данных.

В состав дистрибутива ADH входят последние актуальные версии всех наиболее популярных инструментов. Некоторые из них существенно доработаны, что гарантирует минимальное количество ошибок ПО, самый полный существующий функционал и корректную интеграцию компонентов между собой.

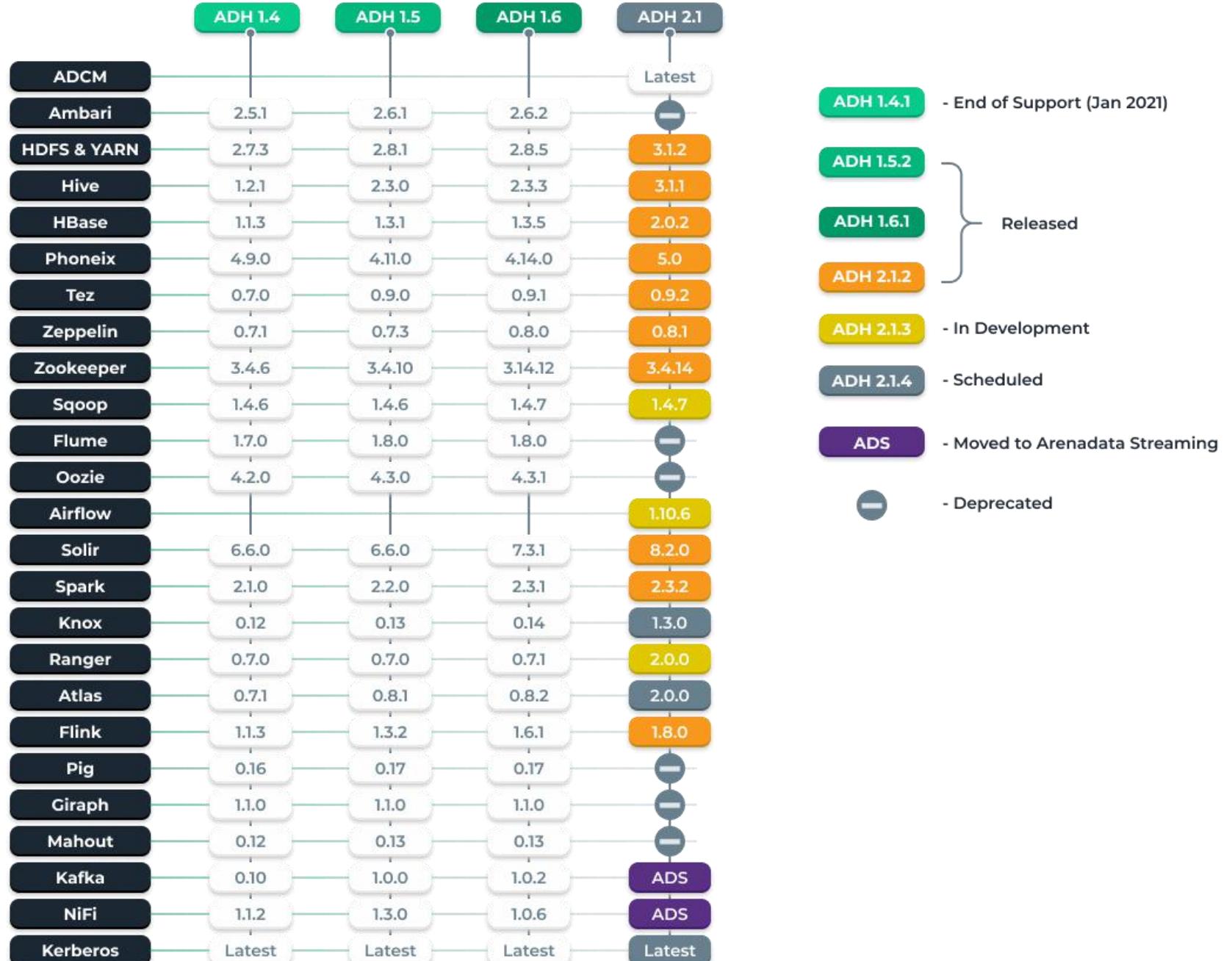
Arenadata Hadoop включает собственную систему управления, развёртывания и мониторинга Arenadata Cluster Manager, позволяющую устанавливать сервисы Hadoop и управлять ими как on-premise, так и в облаке.



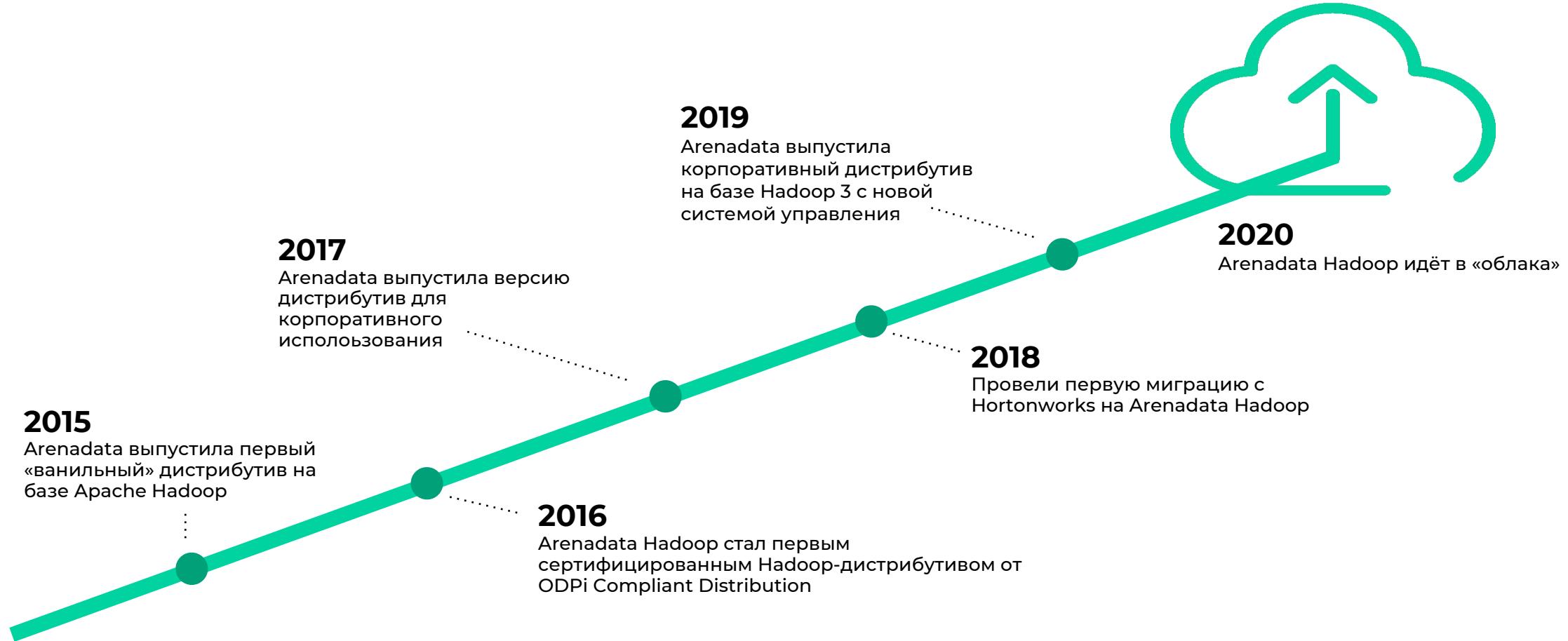
Знакомые вам компоненты



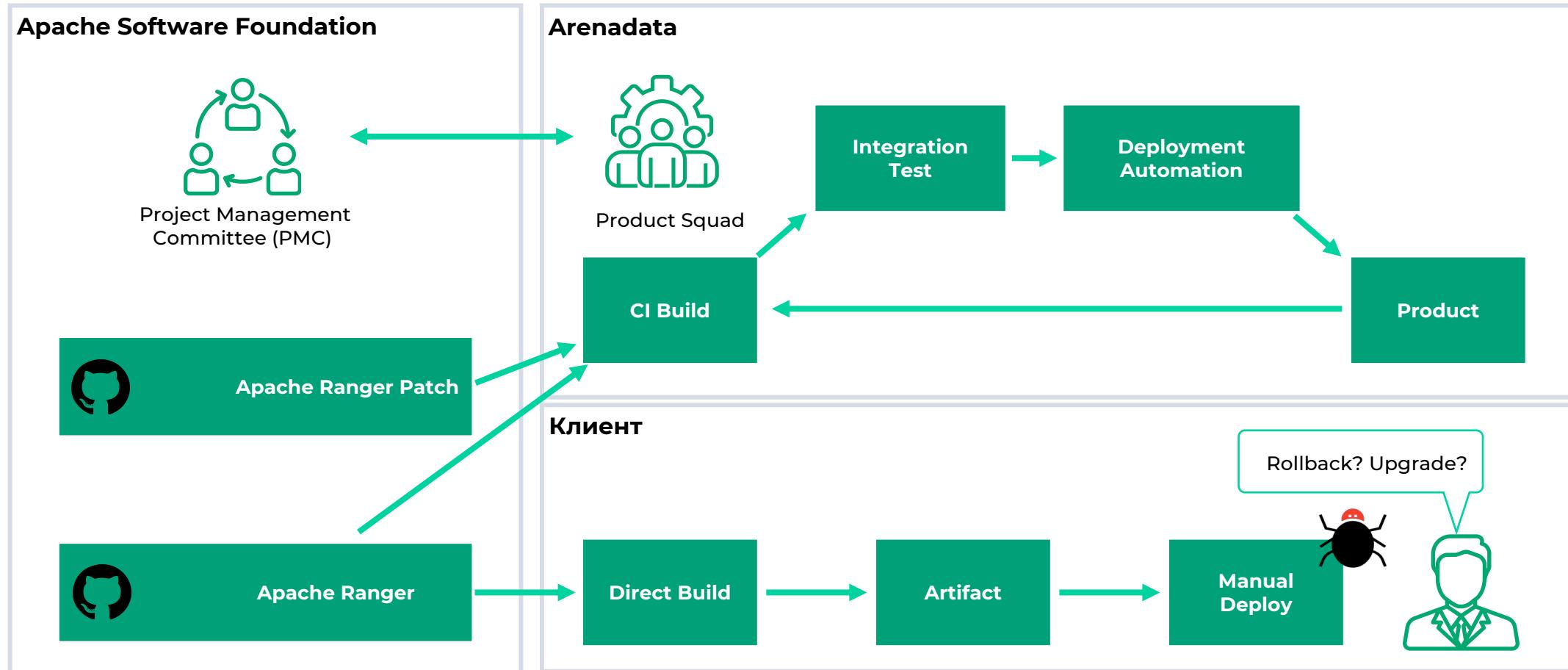
Состав компонентов текущей и будущих версий



История Arenadata Hadoop



Отличия ADН от Apache Hadoop Upstream



Преимущества Arenadata Hadoop



Корпоративная платформа хранения и обработки данных

Arenadata создала богатую экосистему data-сервисов, которая позволит заказчикам использовать не только Hadoop, но и полнофункциональные решения для обработки структурированных и потоковых данных.



Сборка на базе открытых проектов Apache

ADH является полностью open-source дистрибутивом Hadoop, поэтому нашим клиентам никогда не придётся столкнуться с такой проблемой, как vendor lock-in. Помимо этого, мы продолжаем поддерживать legacy-сервисы, в числе которых Flume, Sqoop, Pig, Mahout.



Пакет утилит для полной офлайн-установки

Arenadata Hadoop обеспечивает полный набор возможностей и инструментов для автоматических установки и настройки компонентов как на «голом железе», так и на виртуальных машинах (в облаке). Средства мониторинга и управления конфигурацией кластера позволяют оптимизировать производительность для всех компонентов системы.



Собственная система управления

Мы предоставляем открытую систему автоматического развертывания и управления Arenadata Cluster Manager. Она является Multi-cloud системой, и может быть развернута на любой имеющейся инфраструктуре, включая публичные облака.



Возможность влиять на развитие функционала системы

Наши клиенты могут влиять на планы по развитию системы и экосистемы в целом, а мы можем взять на поддержку полноценный компонент платформы.



Набор типовых пакетных сервисов по планированию, установке и аудиту системы

Вам не придётся самостоятельно проводить оценку оборудования, которое потребуется для решения поставленной задачи. Наши специалисты настроят Arenadata Hadoop (удалённо или on-site), а в последствии проведут аудит системы и помогут разработать шаги для решения ваших проблем.

Arenadata Hadoop — российское ПО

Регистрация в
Реестре
Министерства
Связи

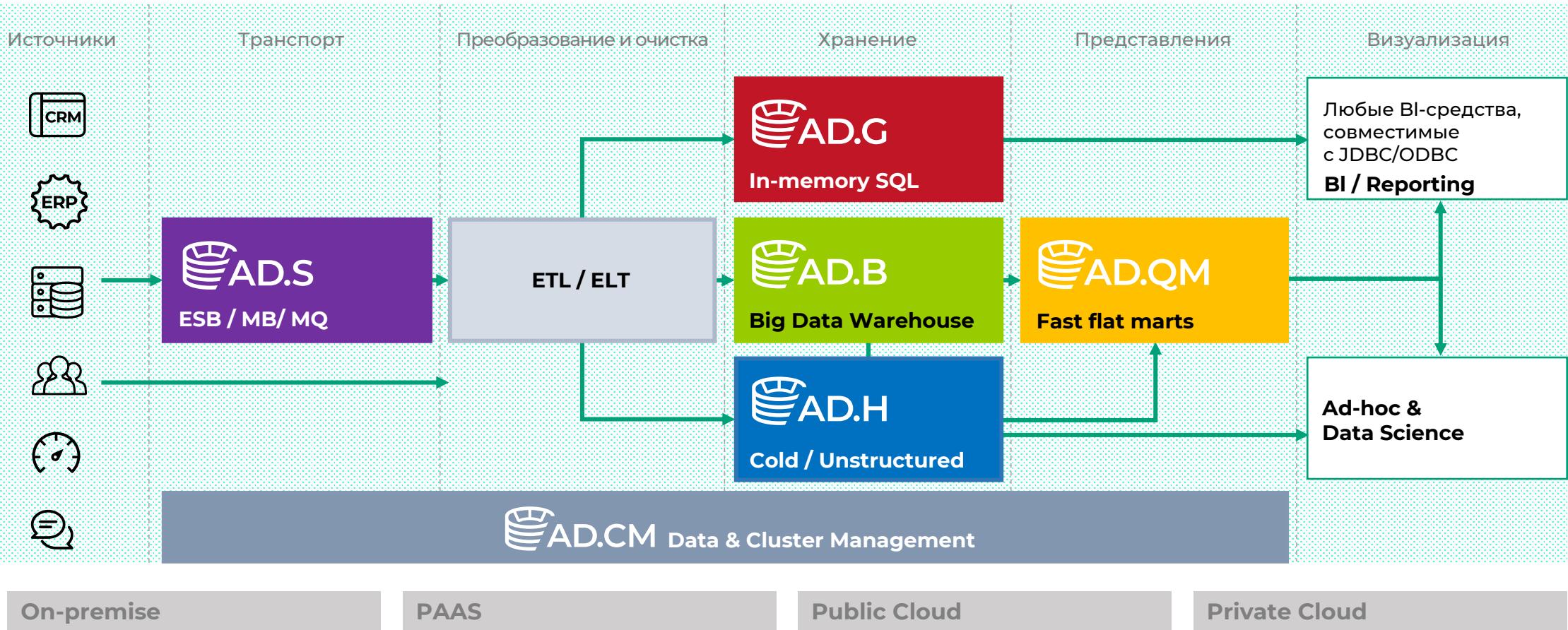
Центры разработки
в РФ - Москва,
Санкт-Петербург,
Хабаровск

Отсутствие
валютных рисков -
цены в рублях

Поддержка,
документация и
учебные курсы на
русском языке



ADН — часть корпоративной платформы данных EDP



Что внутри Arenadata EDP?



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

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

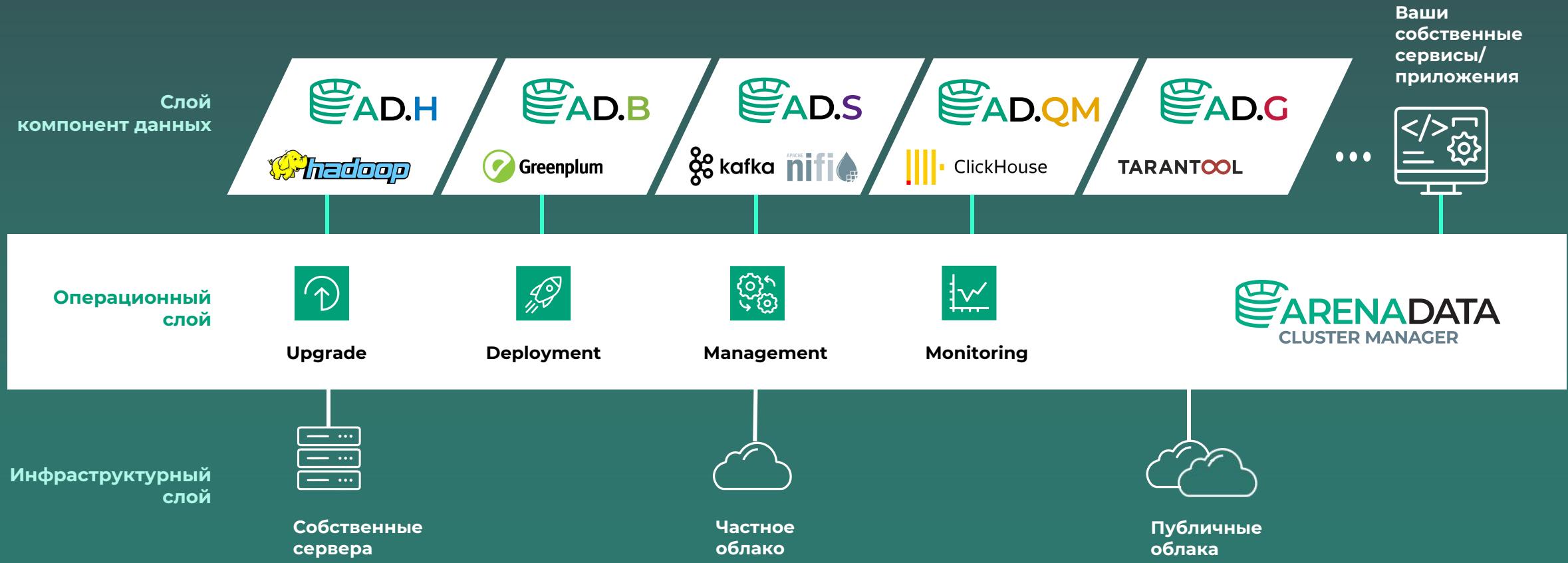
- Поддержка R, Python и Madlib.
- Distributed ML на базе Spark.
- Возможность эффективной эксплуатации моделей в SQL-среде с помощью встроенных функций Madlib.



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.

Гибридное корпоративное облако





Arenadata Cluster Manager (ADCM) - универсальный оркестратор гибридного ландшафта. Позволяет быстро устанавливать, настраивать все data-сервисы компании и управлять ими независимо от инфраструктуры.

Возможности работы с сервисами:

- Установка и настройка
- Обновление
- Управление
- Мониторинг
- Настройка прав доступа
- Интеграция с другими сервисами
- Контейнеризация

Возможности работы с инфраструктурой:

- Создание и удаление виртуальных машин
- Конфигурирование ОС
- Мониторинг
- Управление пользователями
- Настройка прав доступа

Дополнительные возможности:

- Настройка прав безопасности в ADCM
- Открытый ADCM API
- Открытый формат для создания новых бандлов - добавьте ваши собственные сервисы/инфраструктуру



Arenadata Cluster Manager

- Arenadata Cluster Manager — удобный инструмент для установки, настройки, мониторинга и обновления как ПО Arenadata, так и сторонних сервисов
- Функционал приложений ADCM можно использовать для установки, настройки, обновления, управления, мониторинга и интеграции компонентов
- Модульная структура платформы предоставляет необходимый функционал для работы с приложениями и оборудованием по мере необходимости
- Вы значительно сокращаете время на установку и настройку элементов data-ландшафта, тестирование бизнес-сценариев и проверку гипотез
- Вы можете развернуть нужный продукт с помощью ADCM на удобной вам инфраструктуре - на собственных серверах, полностью в одном или нескольких облаках или распределённо на разные инфраструктуры

The screenshot displays the Arenadata Cluster Manager web interface. At the top, there's a navigation bar with tabs for CLUSTERS, HOSTPROVIDERS, HOSTS, JOBS, and BUNDLES. Below the navigation, there are two main content areas.

CLUSTERS Section: This section is titled "ADH Hadoop 2.1". It lists various services with their names, versions, states, and actions. The services listed are YARN (version 3.1.2, installed), HDFS (version 3.1.2, installed), Zookeeper (version 3.4.14, running), HBase (version 2.0.2, installed), Hive (version 3.1.1, created), Spark (version 2.3.2, created), and Zeppelin (version 0.8.1, locked). Each service row includes a green checkmark icon, a "Details" icon, and a "Config" icon.

Name	Version	State	Status	Actions	Config
YARN	3.1.2	installed	✓	[Details]	[Config]
HDFS	3.1.2	installed	✓	[Details]	[Config]
Zookeeper	3.4.14	running	✓	[Details]	[Config]
HBase	2.0.2	installed	✓	[Details]	[Config]
Hive	3.1.1	created	✓	[Details]	[Config]
Spark	2.3.2	created	✓	[Details]	[Config]
Zeppelin	0.8.1	locked	✓	[Details]	[Config]

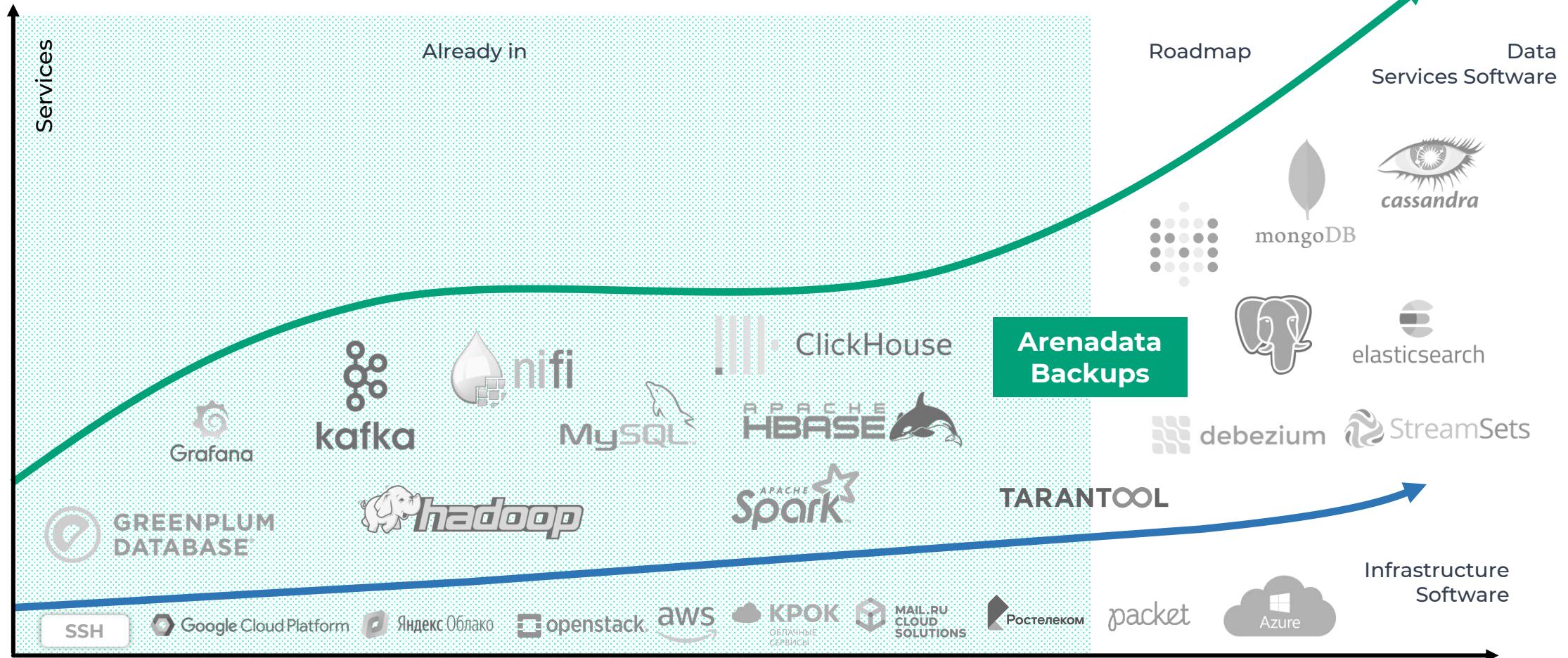
HOSTS Section: This section is titled "HOSTS". It lists hosts with their FQDNs, providers, clusters, states, and actions. The hosts listed are adb-os (Provider: Yandex, Cluster: North Dnieper, State: stopped), adb-pv1 (Provider: Yandex, Cluster: Assign to cluster, State: failed), adb6 (Provider: Yandex, Cluster: Assign to cluster, State: created), adb6-1 (Provider: Yandex, Cluster: Assign to cluster, State: stopped), adb6-2 (Provider: Yandex, Cluster: Assign to cluster, State: stopped), adb6-3 (Provider: Yandex, Cluster: Assign to cluster, State: stopped), alt-linux-repo (Provider: ssh, Cluster: Assign to cluster, State: created), ambari-akonyaev (Provider: Yandex, Cluster: Assign to cluster, State: running), ci-agent-yc-carenadata (Provider: ssh, Cluster: Assign to cluster, State: created), and dro-nsud-load-adb-10 (Provider: ssh, Cluster: Assign to cluster, State: created).

FQDN	Provider	Cluster	State	Status	Actions	Config
adb-os	Yandex	North Dnieper	stopped	!	[Details]	[Config]
adb-pv1	Yandex	Assign to cluster	failed	!	[Details]	[Config]
adb6	Yandex	Assign to cluster	created	!	[Details]	[Config]
adb6-1	Yandex	Assign to cluster	stopped	!	[Details]	[Config]
adb6-2	Yandex	Assign to cluster	stopped	!	[Details]	[Config]
adb6-3	Yandex	Assign to cluster	stopped	!	[Details]	[Config]
alt-linux-repo	ssh	Assign to cluster	created	!	[Details]	[Config]
ambari-akonyaev	Yandex	Assign to cluster	running	!	[Details]	[Config]
ci-agent-yc-carenadata	ssh	Assign to cluster	created	!	[Details]	[Config]
dro-nsud-load-adb-10	ssh	Assign to cluster	created	!	[Details]	[Config]



Roadmap

CLUSTER MANAGER



Enterprise Multi-cloud платформа данных



- Вы можете развернуть любой компонент платформы Areadata EDP, в том числе ADH, 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.

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



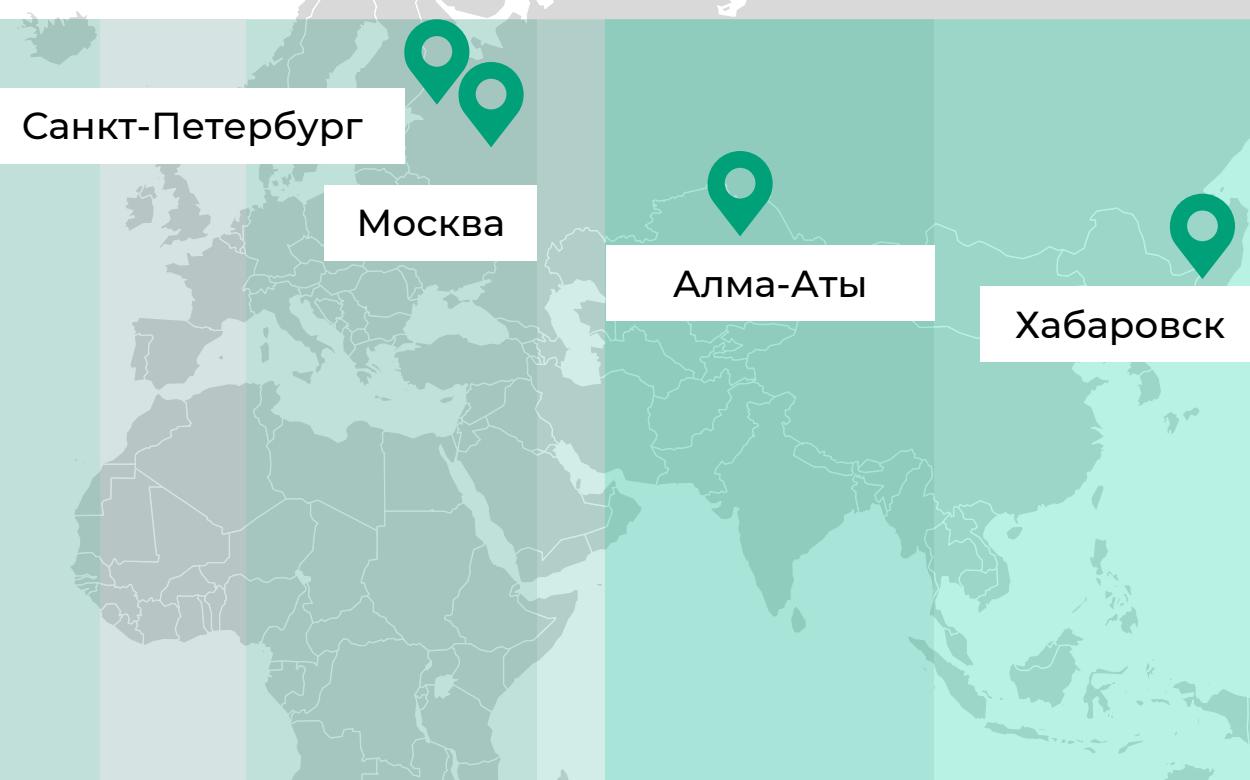
Горизонтальная масштабируемость всех компонент



Open Source

Поддержка 24x7

Сотрудники
распределены
по часовым
поясам



Услуги по поддержке клиентов

Vendor support

VIP-пакет

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



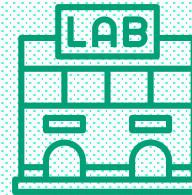
Корпоративная платформа хранения и обработки больших данных

Курсы по Arenadata Hadoop



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



X5 Retail Group

МАГНИТ

М.Видео
нам не всё равно

ВТБ



АСНА
ассоциация независимых аудиторов

iq option
Ultimate trading experience

БКС

CENTERCREDIT
БАНК ТРАДИЦИОННЫХ ЦЕННОСТЕЙ

The logo of the Department of Information Technologies of the City of Moscow, featuring a red and brown graphic element.
Департамент
информационных
технологий
города Москвы

The logo for Gazprom Neft, featuring a stylized flame above the text "ГАЗПРОМ НЕФТЬ".

The logo for Rostelecom, featuring a purple geometric shape above the text "Ростелеком".

НЛМК

The logo for Magnitogorsk Metallurgical Combine, featuring a stylized "M" and the text "МАГНИТОГОРСКИЙ МЕТАЛЛУРГИЧЕСКИЙ КОМБИНАТ".

The logo for Beeline.kz, featuring a yellow and black bee-like icon above the text "Beeline.kz".

The logo for RT labs, featuring a colorful abstract graphic element above the text "RT labs".

The logo for MediaLogika, featuring a stylized figure and the text "МЕДИАЛОГИЯ".

The logo for Kazdream, featuring a stylized blue and white swoosh graphic above the text "KAZDREAM".

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 отчёtnости и аналитики по миллионам клиентов, сокращение ТСО 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 управленической отчёtnости, объединение трёх «песочниц» и их миграция с SAP HANA на Arenadata DB.

Мнение клиентов



Дмитрий Алексеев
заместитель
директора
департамента
технологического
развития
общебанковских
систем ВТБ



Михаил Петров
директор
департамента
Цифровой
трансформации
Счётной палаты РФ



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.



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



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 партнёры

ВНЕДРЕНИЕ



Platformix



ОБЛАЧНЫЕ ПРОВАЙДЕРЫ



ТЕХНОЛОГИЧЕСКИЕ ПАРТНЁРЫ



3 шага к миграции с Hortonworks

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