

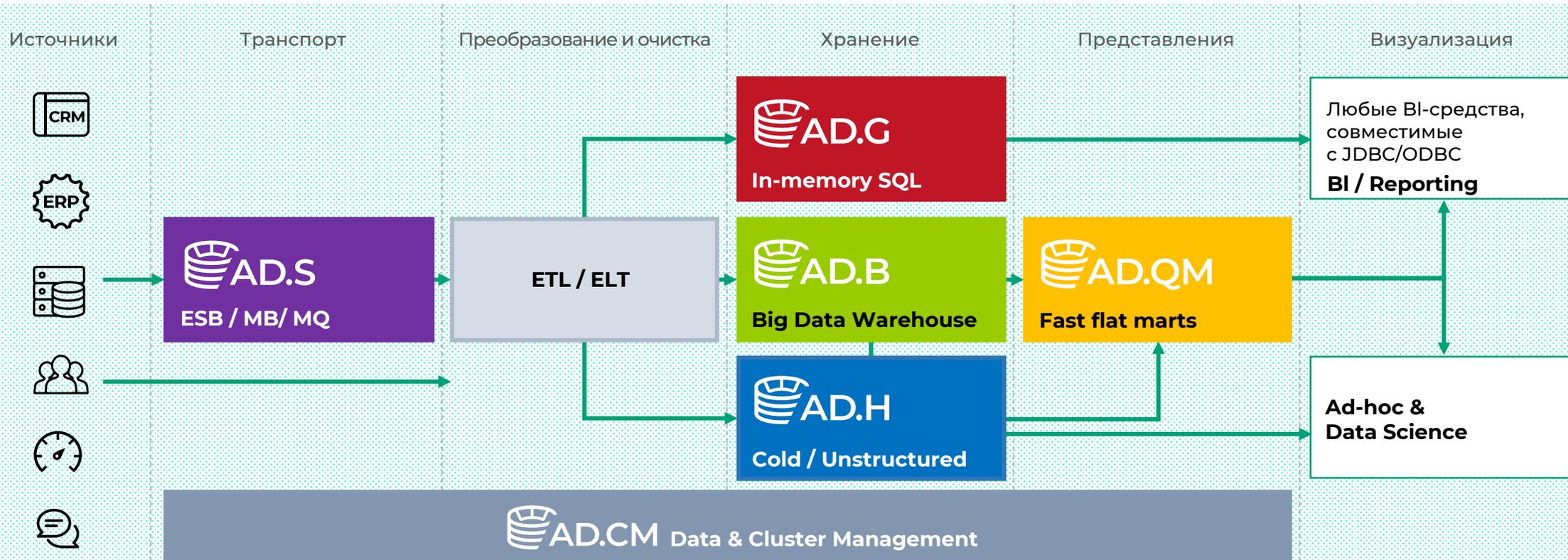


Корпоративная платформа хранения  
и обработки больших данных

# Arenadata Hadoop

Хранение и обработка слабоструктурированных и  
неструктурированных данных

# Arenadata Enterprise Data Platform



On-premise

PAAS

Public Cloud

Private Cloud

# 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.

# Что внутри?

 AD.H Cold / Unstructured	
 AD.B Big Data Warehouse	
 AD.QM Fast flat marts	
 AD.G In-memory SQL	
 AD.S ESB / MB/ MQ	
 AD.CM Data & Cluster Management	

Каждый добавляемый компонент платформы доработан, чтобы стать её частью.

## Производится:

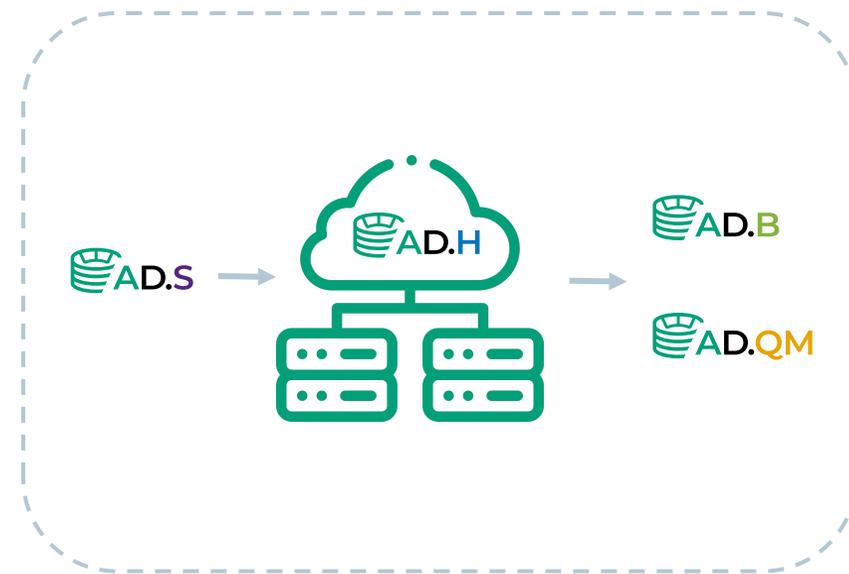
- Анализ кода
- Интеграция с модулем мониторинга
- Проверки отказоустойчивости
- Подбор компетентного персонала поддержки
- Доработка возможностей параллельной интеграции
- Множественные прочие проверки и доработки



**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

## Operations



## Security & Data Governance

Apache Ranger

Apache Atlas

APACHE  
KNOX

## Data Orchestration



Apache  
Zeppelin

## Data Access & Processing



APACHE  
PHOENIX

Solr

TEZ

APACHE  
HBASE

## Persistence & Compute



# Состав компонентов текущей и будущих версий

	ADH 1.4	ADH 1.5	ADH 1.6	ADH 2.1
ADCM				Latest
Ambari	2.5.1	2.6.1	2.6.2	—
HDFS & YARN	2.7.3	2.8.1	2.8.5	3.1.2
Hive	1.2.1	2.3.0	2.3.3	3.1.1
HBase	1.1.3	1.3.1	1.3.5	2.0.2
Phoneix	4.9.0	4.11.0	4.14.0	5.0
Tez	0.7.0	0.9.0	0.9.1	0.9.2
Zeppelin	0.7.1	0.7.3	0.8.0	0.8.1
Zookeeper	3.4.6	3.4.10	3.14.12	3.4.14
Sqoop	1.4.6	1.4.6	1.4.7	1.4.7
Flume	1.7.0	1.8.0	1.8.0	—
Oozie	4.2.0	4.3.0	4.3.1	—
Airflow				1.10.6
Solir	6.6.0	6.6.0	7.3.1	8.2.0
Spark	2.1.0	2.2.0	2.3.1	2.3.2
Knox	0.12	0.13	0.14	1.3.0
Ranger	0.7.0	0.7.0	0.7.1	2.0.0
Atlas	0.7.1	0.8.1	0.8.2	2.0.0
Flink	1.1.3	1.3.2	1.6.1	1.8.0
Pig	0.16	0.17	0.17	—
Giraph	1.1.0	1.1.0	1.1.0	—
Mahout	0.12	0.13	0.13	—
Kafka	0.10	1.0.0	1.0.2	ADS
NIFi	1.1.2	1.3.0	1.0.6	ADS
Kerberos	Latest	Latest	Latest	Latest

- ADH 1.4.1 - End of Support (Jan 2021)
- ADH 1.5.2 } Released
- ADH 1.6.1 }
- ADH 2.1.2 }
- ADH 2.1.3 - In Development
- ADH 2.1.4 - Scheduled
- ADS - Moved to Arenadata Streaming
- - Deprecated

**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 image displays two screenshots of the Arenadata Cluster Manager (ADCM) web interface. The top screenshot shows the 'ADH' (Hadoop 2.1) configuration page, which lists various services and their status. The bottom screenshot shows the 'HOSTS' page, which lists the infrastructure components used for the clusters.

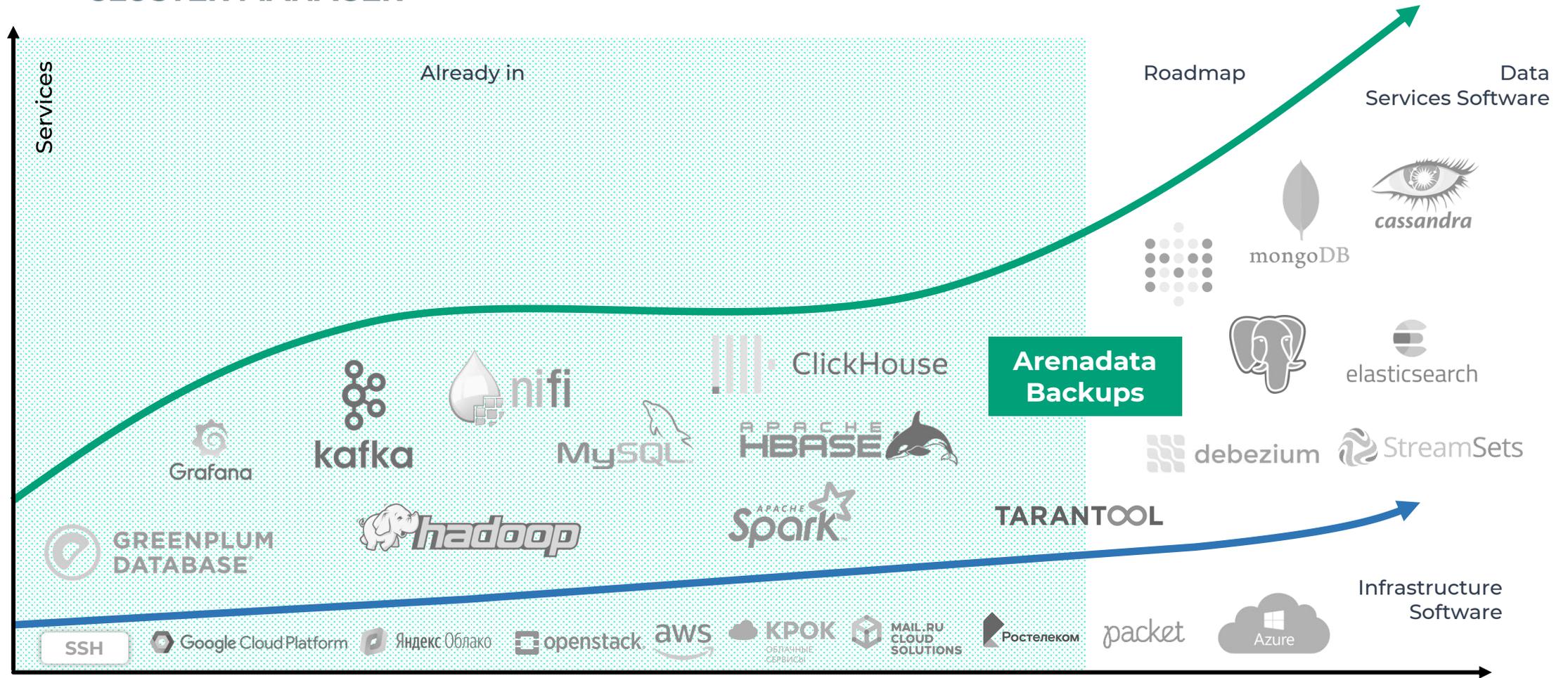
Name	Version	State	Status	Actions	Config
YARN	3.1.2	installed	✓	⚙️	⚙️
HDFS	3.1.2	installed	✓	⚙️	⚙️
Zookeeper	3.4.14	running	✓	⚙️	⚙️
HBase	2.0.2	installed	✓	⚙️	⚙️
Hive	3.1.1	created	✓	⚙️	⚙️
Spark	2.3.2	created	✓	⚙️	⚙️
Zeppelin	0.8.1	locked	✓	⚙️	⚙️

FOQN	Provider	Cluster	State	Status	Actions	Config
adb-os	Yandex	North Dnieper	stopped	⚠️	⚙️	🗑️
adb-pytl	Yandex	Assign to cluster	failed	⚠️	⚙️	🗑️
adb6	Yandex	Assign to cluster	created	⚠️	⚙️	🗑️
adb6-1	Yandex	Assign to cluster	stopped	⚠️	⚙️	🗑️
adb6-2	Yandex	Assign to cluster	stopped	⚠️	⚙️	🗑️
adb6-3	Yandex	Assign to cluster	stopped	⚠️	⚙️	🗑️
alb-linux-repo	ssh	Assign to cluster	created	⚠️	⚙️	🗑️
ambari-akornyaev	Yandex	Assign to cluster	running	⚠️	⚙️	🗑️
ci-agent-ycc-arenadata	ssh	Assign to cluster	created	⚠️	⚙️	🗑️
dro-naud-load-adb-10	ssh	Assign to cluster	created	⚠️	⚙️	🗑️

# ARENADATA Roadmap

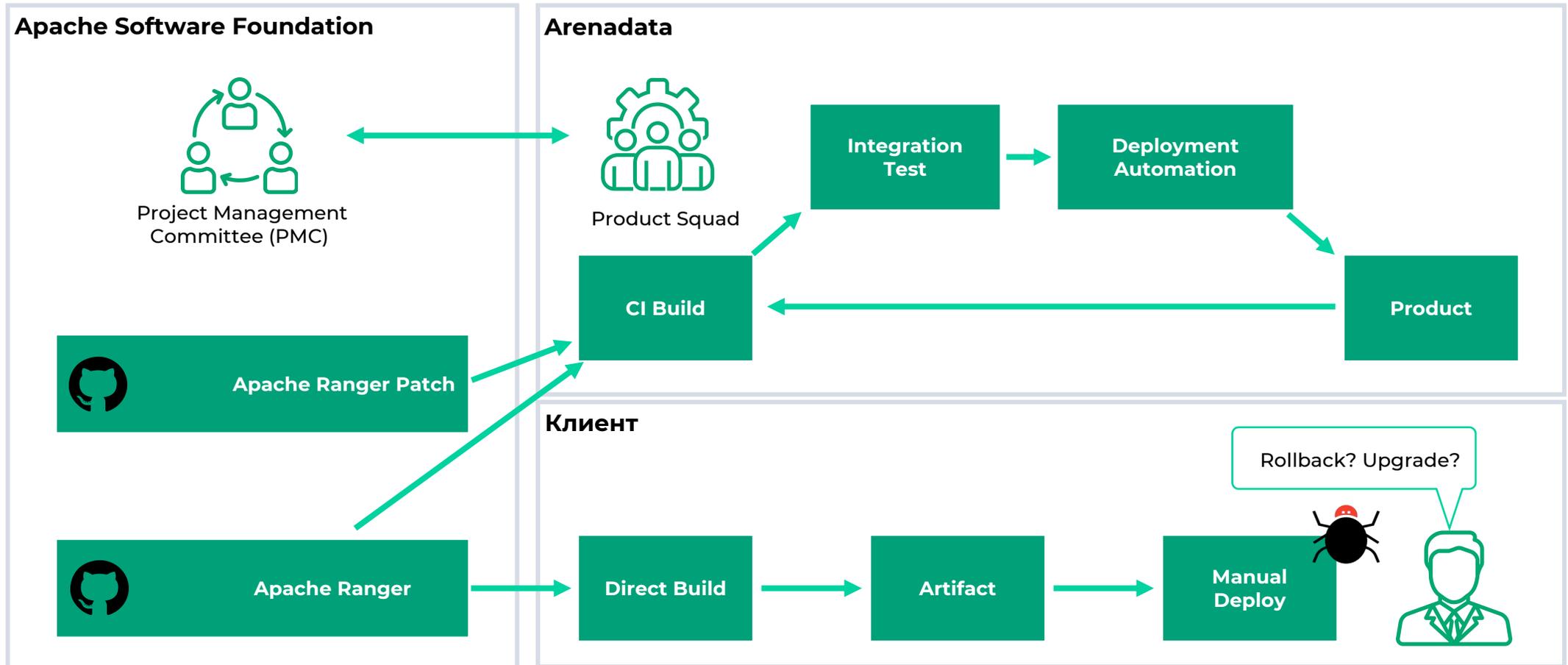
CLUSTER MANAGER



# История Arenadata Hadoop



# Отличия ADH от Apache Hadoop Upstream



# 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

# 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 автоматической 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 развёртывания и управления Arenadata Cluster Manager. Она является Multi-cloud системой, и может быть развёрнута 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

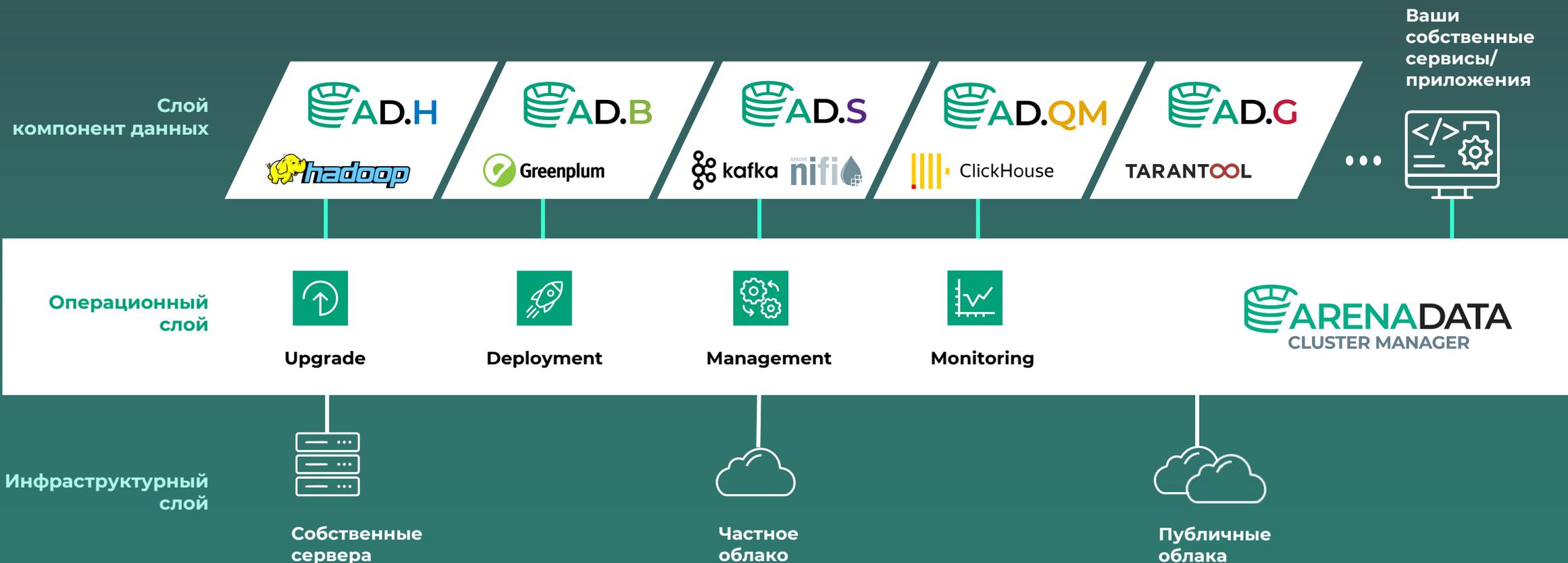
Отсутствие  
валютных рисков -  
цены в рублях

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

Поддержка,  
документация и  
учебные курсы на  
русском языке



# Гибридное корпоративное облако



# Enterprise Multi-cloud платформа данных



- Вы можете развернуть любой компонент платформы Arendata 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.

# Поддержка 24x7

Санкт-Петербург

Москва

Алма-Аты

Хабаровск

Сотрудники  
распределены  
по часовым  
поясам

# Услуги по поддержке клиентов

Vendor support

## VIP-пакет

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

 **ARENADATA**

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



# 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 помогли с вопросами внедрения и безопасности, предлагали собственные решения и оперативно выпускали релизы с учётом специфических требований заказчика.



# 3 шага к Arenadata Hadoop

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](https://arenadata.tech)
- Скачайте бесплатно на [store.arenadata.io](https://store.arenadata.io)
- Получите консультацию по почте [info@arenadata.io](mailto:info@arenadata.io)

