

Arenadata Hadoop

Экосистема для хранения, обработки и анализа неограниченного объёма данных любого типа



План презентации

1. О компании Arenadata
2. Обзор Arenadata Hadoop
3. Экосистема Arenadata Hadoop
4. Сравнение ADH vs Open Source
5. Платформа сбора и хранения данных Arenadata
Enterprise Data Platform

О компании Arenadata

Arenadata

лидер в области платформ больших данных для цифровой трансформации



TARANTOOL



СТАБИЛЬНОСТЬ

АДС Холдинг:

- **6 лет** на рынке
- **430** человек
- **4,0 млрд. Р** выручка '23



80+ КЛИЕНТОВ

- ГПН, Росатом, ММК, Норникель, Сибур;
- X5, Магнит, Ашан;
- Тинькофф, ВТБ, ПСБ, ГПБ.
- ФНС, ДИТ МСК, Минздрав;



NO VENDOR LOCK

Мы создаем **тиражируемый продукт, совместимый с upstream** версиями opensource проектов.



ONPREM & CLOUD

- Onpremise и ПАКи;
- CCloud: VK Cloud; Cloud.ru, MTS Cloud, TI Cloud, Croc Cloud, Beeline Cloud;
- Гетерогенные ландшафты;



ФОКУС БИЗНЕСА

100% бизнеса Arenadata – создание СУБД на базе Open Source, поддержка 1, 2 и 3 линии. Архитектурный надзор от вендора.



ЭКСПЕРТЫ

- **60+** технологических партнеров и интеграторов;
- **2.500+** курсов проведено для специалистов и экспертов.



COMMUNITY

- **#1** в мире в сообществе контрибьютеров в Greenplum;
- **#1** в РФ в сообществе контрибьютеров в Clickhouse;
- Профессиональная [документация](#).



РЕГУЛЯТОРЫ

- Отечественные ОС;
- Реестр;
- ФСТЭК;
- ГОСТЕХ, ГЕОП.

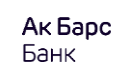
Наш успех – доверие клиентов

80+ клиентов размещают свои хранилища, витрины, озера данных на продуктах Arenadata

Retail & FMCG



Банки, страховые и телеком



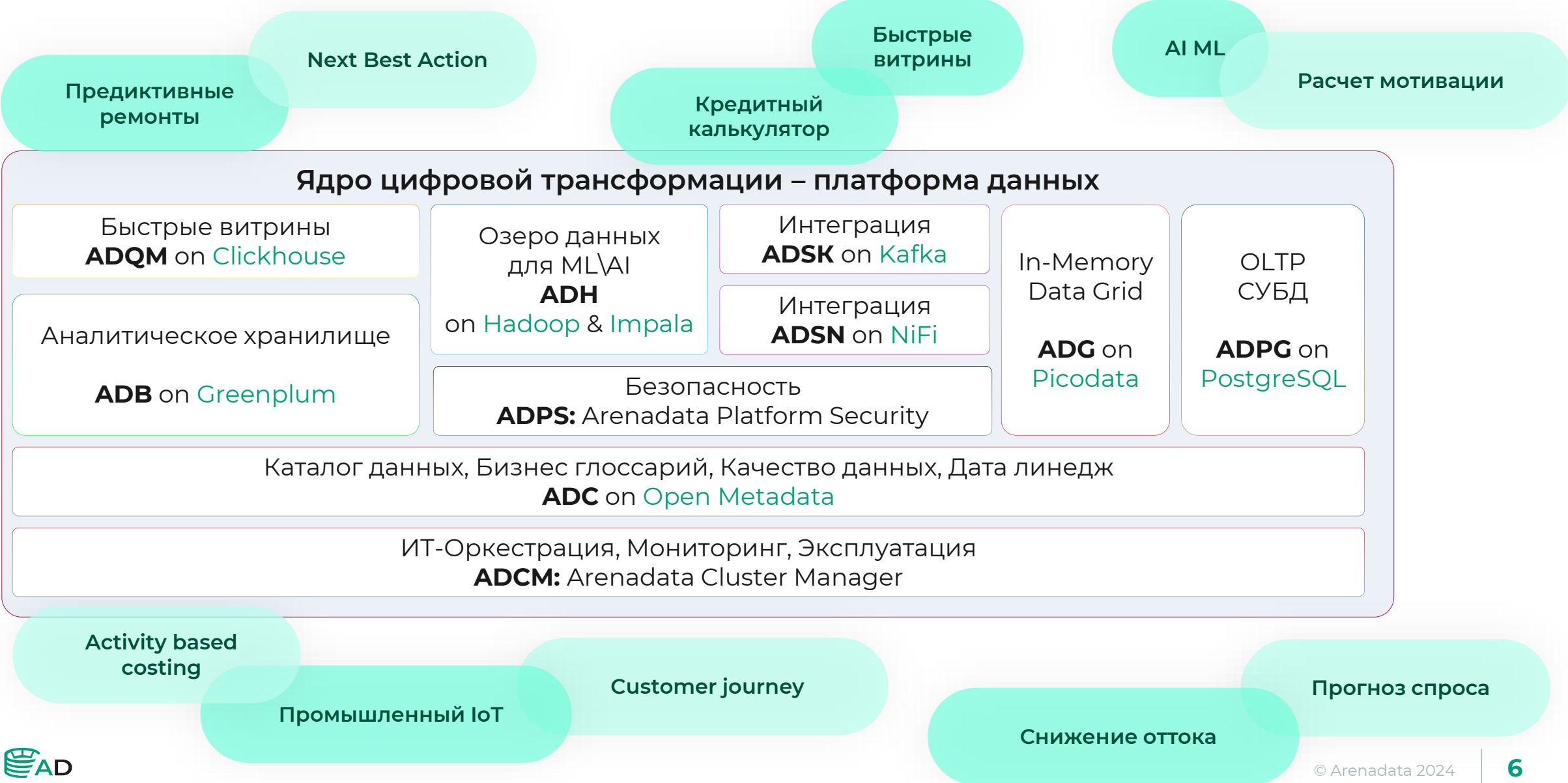
Госсектор



Пром, Энерго, Добыча, Транспорт



Arenadata – платформа данных для цифровой трансформации



Обзор Arenadata Hadoop

Корпоративный дистрибутив для распределённой масштабируемой обработки данных



Arenadata Hadoop (ADH) — корпоративный дистрибутив на базе Apache Hadoop, предназначенный для хранения и обработки слабоструктурированных и неструктурированных данных.

Наиболее эффективен для следующих сценариев:

- обработка больших наборов данных в средах, где размер данных превышает доступную память
- пакетная обработка с задачами, использующими операции чтения и записи на диск
- создание инфраструктуры для анализа больших данных
- анализ исторических и архивных данных
- непрерывный сбор различных метрик и журналов
- распределённое хранилище оперативных данных

Развитие Arenadata Hadoop

Arenadata выпустила первый «ванильный» дистрибутив на базе Apache Hadoop

Arenadata выпустила дистрибутив для корпоративного использования

Arenadata выпустила дистрибутив на базе Hadoop 3 с новой системой управления

- Arenadata Hadoop доступен из облака VK Cloud Solutions
- Arenadata выпустила Platform Security

- Новые функции безопасности
- Поддержка Apache Impala
- Поддержка ОС Astra Linux

2015

2017

2019

2021

2023

2016

2018

2020

2022

Arenadata Hadoop стал первым сертифицированным Hadoop-дистрибутивом от ODPi Compliant Distribution

Провели первую миграцию с Hortonworks на Arenadata Hadoop

Arenadata Hadoop идёт в «облака»

- Усиление безопасности
- Поддержка ОС Альт 8 СП
- Обновление компонентов

Решаемые задачи



Хранение данных любого типа, включая неструктурированные

- Распределённая обработка данных
- Системы управления документами и контентом
- Хранение и регистрация событий
- Данные датчиков, каталоги товаров
- Резервное копирование других СУБД



Построение озёр и фабрик данных

- Единый центр всех данных компании
- Быстрое развёртывание «песочниц» для пилотных проектов и проверки статистических гипотез
- Работа со всеми аналитическими инструментами в единой среде



Машинное обучение и искусственный интеллект

- Обучение моделей на больших данных
- Распределённое машинное обучение на базе Spark
- Эффективная эксплуатация моделей в SQL-среде с помощью встроенных функций Madlib



Импортозамещение и разгрузка иностранных систем

- Миграция с иностранных систем (Oracle BDA, Cloudera)
- Прозрачная методика перехода, минимум рисков и сохранение всех преимуществ

Минимизируйте санкционные и валютные риски



Arenadata Hadoop высоко зарекомендовал себя на российском рынке. Продукт эффективно замещает такие иностранные системы как Cloudera CDP, Oracle BDA, а также «ванильные» сборки Hadoop и др.



Наши программные решения стали частью проектов миграции с зарубежного ПО в десятках крупнейших организаций, включая ФНС РФ, Счётную палату РФ, Банк ВТБ и ПАО «Газпром нефть».



ORACLE®
BIG DATA APPLIANCE



CLOUDERA
Data Platform

Преимущества Arenadata Hadoop

1

Корпоративная платформа хранения и обработки данных

Широкая экосистема data-сервисов Arenadata позволит заказчикам использовать не только Hadoop, но и полнофункциональные решения для обработки структурированных и потоковых данных.

3

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.

2

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.

4

Интеграция с FreeIPA

Поддержка Kerberos аутентификации и Ranger авторизации на базе FreeIPA (централизованной системы по управлению идентификацией пользователей), как альтернатива интеграции с Active Directory.

Преимущества Arenadata Hadoop

5

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.

7

Набор типовых пакетных сервисов по планированию, установке и аудиту системы

Наши специалисты настроят Arenadata Hadoop (удалённо или on-site), а впоследствии проведут аудит системы и помогут разработать шаги для решения поставленных задач.

6

Возможность влиять на развитие системы

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

8

Поддержка российских ОС

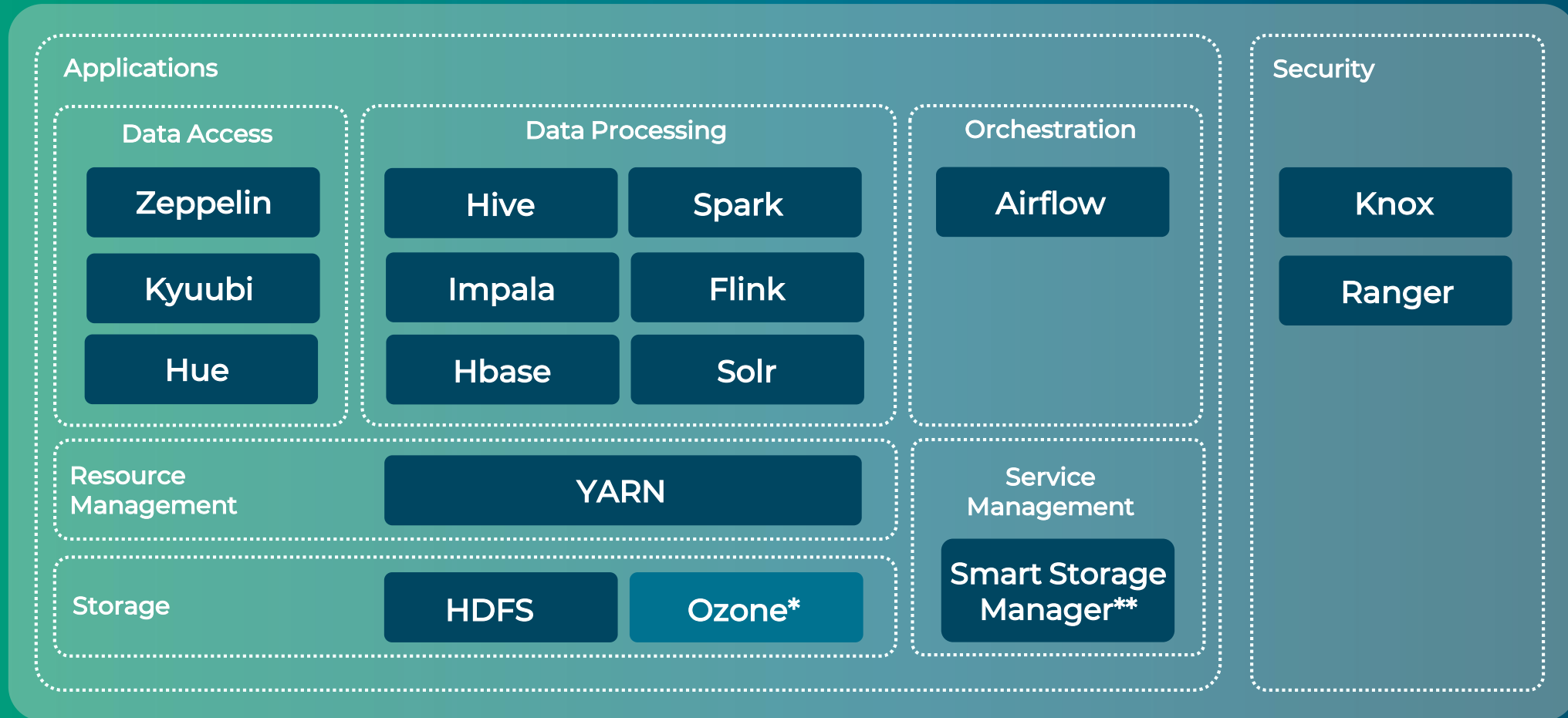
Поддержка российских сертифицированных ОС AltLinux 8 SP (Альт 8 СП), Astra Linux 1.7 SE «Орёл».

Экосистема Arenadata Hadoop

Компоненты и сервисы



Экосистема Arenadata Hadoop



* В разработке

** Technology Preview

Состав компонентов текущей и будущих версий

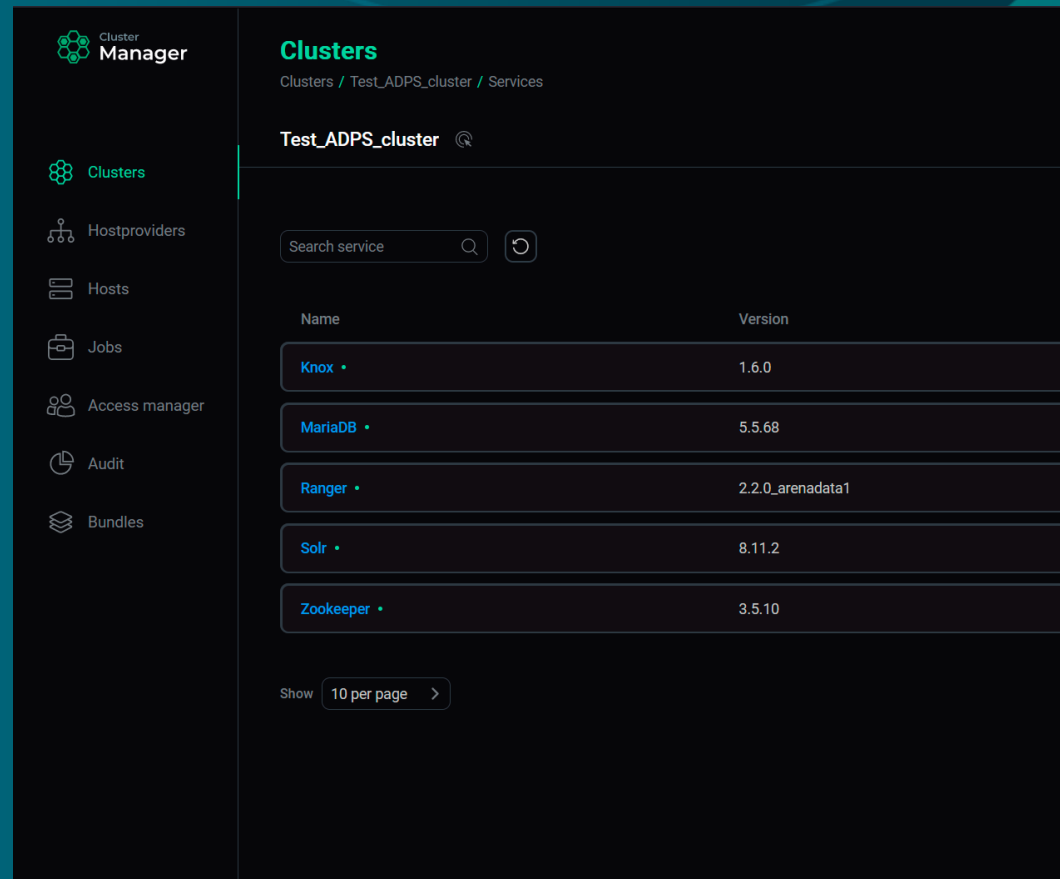
	ADH 1.6	ADH 3.2.4.1	ADH 3.2.4.2	ADH 3.2.4.3	ADH 3.3.6.1	ADH 3.3.6.2
HDFS&YARN	2.8.5	3.2.4	3.2.4	3.2.4	3.3.6	3.3.6
Hive	2.3.3	3.1.3	3.1.3	3.1.3	4.0.0	4.0.0
Tez	0.9.1	0.10.1	0.10.1	0.10.1	0.10.1	0.10.1
HBase	1.3.5	2.4.17	2.4.17	2.4.17	2.5.8	2.5.8
Phoenix	4.14.0	5.1.3	5.1.3	5.1.3	5.1.3	5.1.3
Impala		4.2.0	4.2.0	4.4.0	4.4.0	4.4.0
Zeppelin	0.8.0	0.8.1	0.8.1	0.11.1	0.11.1	0.11.1
Zookeeper	3.4.12	3.5.10	3.5.10	3.5.10	3.5.10	3.5.10
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7	1.4.7	1.4.7
Airflow		—	—	—	—	—
Airflow2		2.6.3	2.6.3	2.6.3	2.6.3	2.6.3
Solr	7.3.1	8.11.2	8.11.2	8.11.2	8.11.2	8.11.2
Kyuubi			1.8.0	1.9.0	1.9.0	1.9.0
SSM			1.6.0	1.6.0	2.0.0	2.1.0
Spark2	2.3.1	2.3.2	2.3.2	2.3.2	2.3.2	2.3.2
Spark3		3.3.2	3.4.2	3.4.3	3.5.1	3.5.1
Flink	1.6.1	1.17.1	1.17.1	1.18.1	1.19.0	1.19.1
Hue				4.11.0	4.11.0	4.11.0
Knox	0.14	1.6.0	1.6.0	1.6.0	1.6.0	1.6.0
Ranger	0.7.1	2.2.0	2.4.0	2.4.0	2.4.0	2.4.0
Kerberos	latest	latest	latest	latest	latest	latest
ADCM		latest	latest	latest	latest	latest
Kafka	1.0.2	ADS	ADS	ADS	ADS	ADS
NiFi	1.6.0	ADS	ADS	ADS	ADS	ADS

- ADH 1.6
- ADH 3.2.4.1
- ADH 3.2.4.2
- ADH 3.2.4.3
- ADH 3.3.6.1 - In Development
- ADH 3.3.6.2 - Scheduled
- ADS - Moved to Arenadata Streaming
- - Deprecated

Система централизованного управления политиками безопасности кластера ADH включает:

- аутентификацию с использованием Kerberos, интеграция с LDAP/Active Directory
- интеграцию с Apache Ranger и Apache Knox для обеспечения безопасного доступа к кластерам Hadoop
- журналы аудита и отчёты

Поставляется как бесплатное дополнение к Enterprise редакции ADH версии 2.1+



The screenshot shows the Cluster Manager interface. On the left is a navigation sidebar with icons for Clusters, Hostproviders, Hosts, Jobs, Access manager, Audit, and Bundles. The main content area is titled 'Clusters' and shows the path 'Clusters / Test_ADPS_cluster / Services'. Below this is a search bar for services and a table listing services and their versions.

Name	Version
Knox	1.6.0
MariaDB	5.5.68
Ranger	2.2.0_arenadata1
Solr	8.11.2
Zookeeper	3.5.10

At the bottom of the table, there is a 'Show 10 per page' dropdown menu.

Функции безопасности Arenadata Platform Security

Комплексный подход к безопасности, включающий защиту периметра, управление доступом на основе политик, авторизацию и безопасный доступ к платформе и ее сервисам.

Защита конфиденциальных данных и соответствие нормативным требованиям.



Безопасность периметра

- Apache Knox
- Gateway



Аутентификация

- Kerberos
- LDAP/AD
- FreeIPA
- MIT KDC



Защита данных

- SSL
- Шифрование RPC
- Шифрование at Rest



Аудит и мониторинг

- Запросы доступа
- Операции обработки данных
- Изменение данных



Авторизация

- Контроль доступа HDFS, YARN, Hbase
- Контроль доступа на уровне баз данных, таблиц, столбцов для наборов данных Hive
- Контроль доступа Knox
- Контроль доступа к коллекциям Solr

Impala: MPP-СУБД в экосистеме Hadoop



Impala — распределённая система исполнения SQL-запросов в экосистеме Hadoop. Сервис предназначен для интерактивной обработки данных на сверхбольших объёмах данных.

Основные преимущества



Высокая скорость обработки запросов в озере данных

Низкая задержка и высокий уровень параллелизма в экосистеме Hadoop, что позволяет эффективнее решать задачи self-service аналитики и ad-hoc запросов



Простое внедрение в имеющуюся инфраструктуру

Impala использует те же метаданные, форматы файлов и драйверы подключения, что и Hive



Масштабирование аналитической нагрузки

Развёртывание Impala вне основного кластера позволяет исключить конкуренцию за аппаратные ресурсы



Снижение стоимости обработки данных

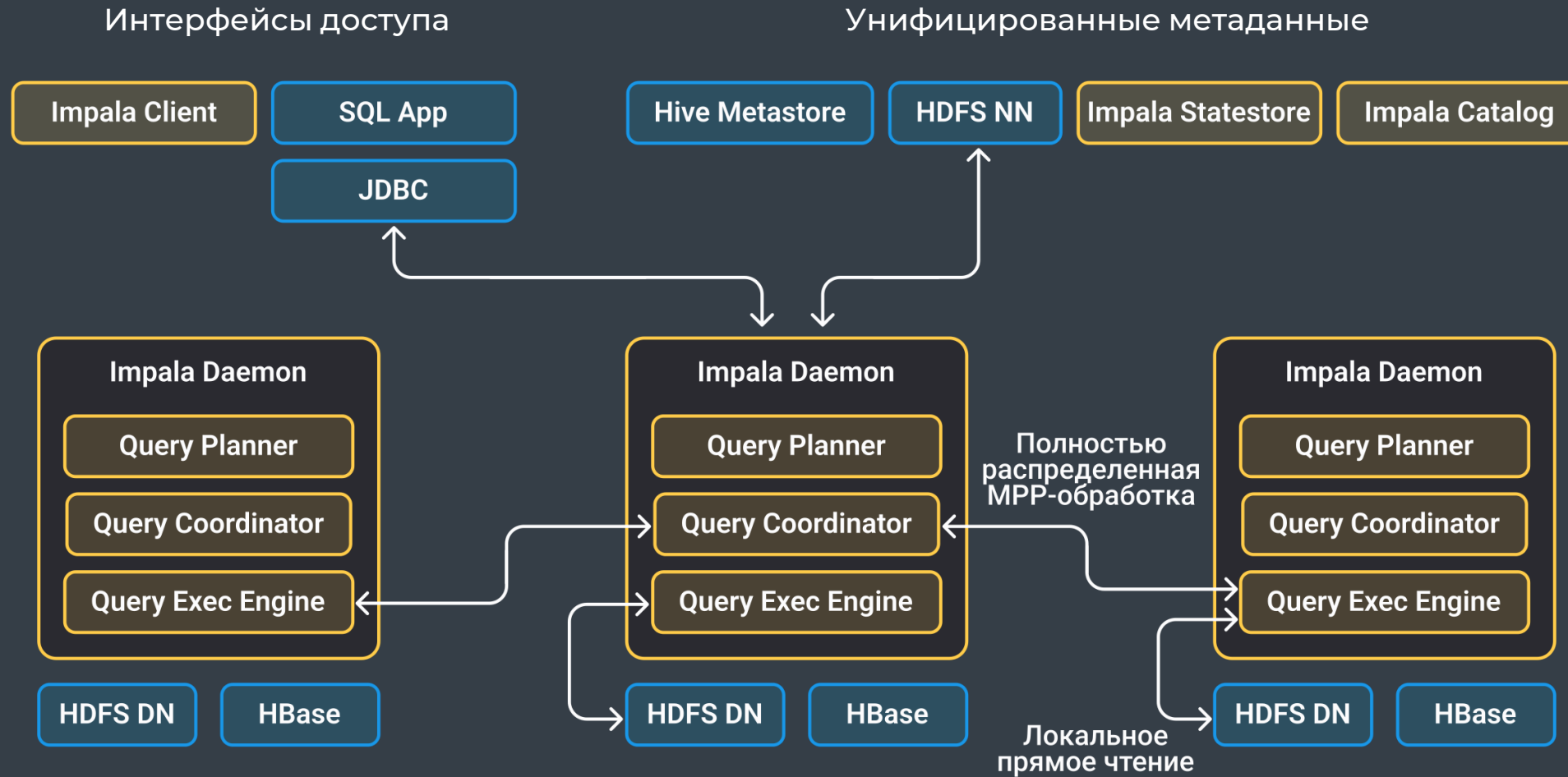
Достигается за счёт оптимального использования аппаратного обеспечения



Оптимизация ландшафта

Благодаря локальной реализации отдельных сценариев ad-hoc и self-service аналитики

Архитектура Impala



Почему Impala? Технические аргументы



Быстрый SQL – ELT на больших данных

(TPC-DS/H SF>5000)

- MPP-архитектура на Hadoop
- Механизм Short-Circuit (только C++ библиотеки)
- Кэширование данных
- Кэширование метаданных
- Вычисления в памяти



Перспективный opensource проект

- Активно развивается
- Hadoop Native (HDFS, Ozone)
- Проверенный production ready сервис



Конкурентный доступ – Ad Hoc-analysis

- Минимальная утилизация ОП
- Механизм кэширования данных ОС

Excluding merges, 22 authors have pushed 63 commits to master and 69 commits to all branches. On master, 292 files have changed and there have been 118,117 additions and 10,214 deletions.



apache / impala

Варианты развёртывания Impala

1. Совместно с HDFS на одном узле данных
2. Отдельный аналитический кластер Impala

1

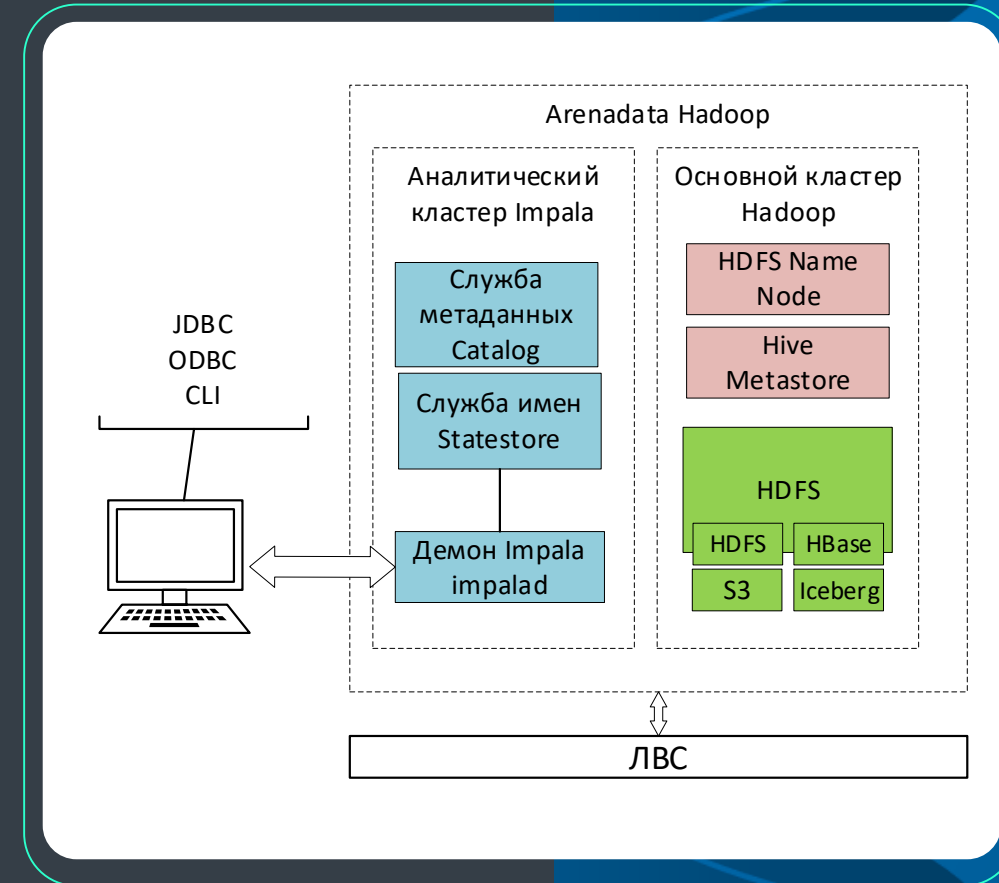
Совместное с HDFS развёртывание это:

- максимально производительный MPP
- использование преимуществ механизма прямого доступа к блокам данных (Short-Circuit) и HDFS cache

2

Развёртывание отдельного аналитического кластера Impala позволит:

- использовать существующую инфраструктуру хранения и обработки
- дополнить производственный кластер быстрой Ad Hoc-аналитикой
- ускорить критические ELT-процессы



Smart Storage Manager: интеллектуальное управление данными в экосистеме Hadoop

Температурное хранение данных

Оптимизация управления данными в зависимости от их востребованности:

- перемещение наиболее горячих данных в кеш
- перемещение горячих данных в SSD
- архивация холодных данных

Асинхронная репликация

Репликация данных между Hadoop-кластерами или между Hadoop-кластером и облачным хранилищем:

- отслеживание изменений данных
- синхронизация в реальном времени
- реализация сценариев аварийного восстановления

Настройка политик Erasure Coding и сжатия данных

- гибкая настройка включения Erasure Coding и управление файлами с разными политиками EC с помощью правил
- сжатие данных в HDFS без ограничения доступа к ним для внешних приложений

Оптимизация работы с небольшими файлами

- сжатие небольших файлов в файл-контейнер, хранящийся в HDFS
- данные доступны для приложений верхнего уровня

80%

**вычислительных
нагрузок приходится
на обработку**

20%

данных

Smart Storage Manager: преимущества сервиса

- Снижение стоимости хранения холодных данных
- Повышение производительности чтения горячих данных
- Простая настройка и управление функцией асинхронной репликации
- Снижение накладных расходов и повышение производительности записи и чтения небольших файлов HDFS
- Экономия места в хранилище



Оптимизация
затрат



Повышение
производительности



Надёжность

Кyuubi: SQL gateway в экосистеме Hadoop



Кyuubi — распределённый многопользовательский шлюз для предоставления SQL для DWH и DataLake.

Кyuubi создает распределённые механизмы SQL запросов поверх различных вычислительных платформ, например, Apache Spark, Flink, Hive, Impala и др., чтобы получать и обрабатывать большие наборы данных из разнородных источников.

Основные возможности



Многопользовательский доступ

Сквозная поддержка доступа нескольких пользователей к данным через единую систему аутентификации и авторизации



Высокая доступность (HA)

Балансировка нагрузки через Zookeeper обеспечивает высокую доступность enterprise-уровня и неограниченно высокий уровень параллелизма клиентов



Несколько рабочих нагрузок

Поддержка разнородных рабочих нагрузок в рамках одной платформы, одной копии данных и одного интерфейса SQL

Сферы применения Кууби



Интерактивная аналитика

- Быстрая аналитика для интерактивного анализа больших данных
- Распределённые механизмы SQL-запросов поверх различных вычислительных платформ (Spark, Flink, Impala и др.)
- Доступ через JDBC/ODBC
- Возможность генерации запросов через SQL или инструменты BI
- Совместное использование ресурсов и быстрый отклик за счёт распараллеливания запросов



Пакетная обработка

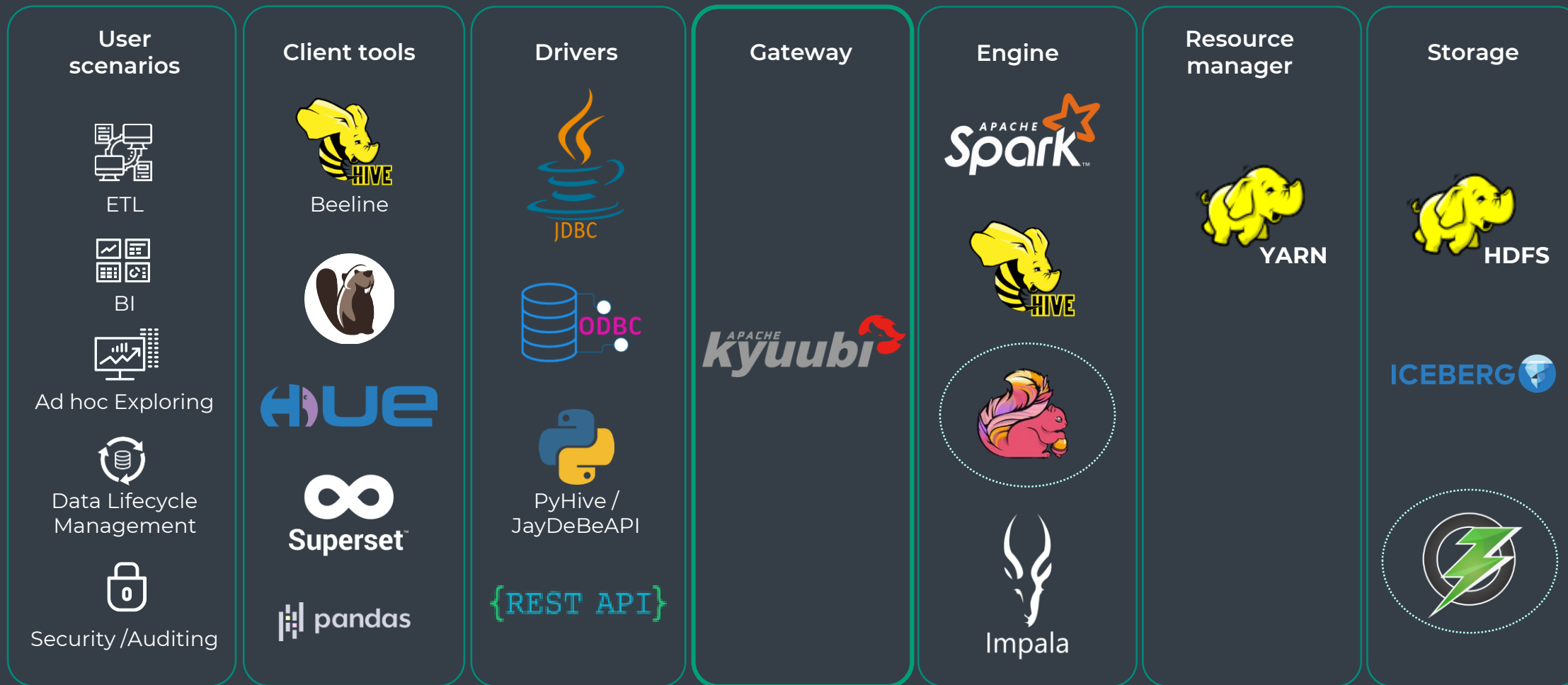
- Кууби предоставляет интерфейс SQL, удобный для пакетной обработки ETL
- Кууби и его движки работают с многочисленными источниками данных независимо от хранилища
- Изоляция вычислительных ресурсов



Data Lake & Lakehouse

- Возможность выполнения запросов к традиционным хранилищам (Hive/HDFS) и современным озёрам данных.
- Централизованная картина данных
- Возможность запрашивать разнородные источники данных
- Аутентификация и авторизация (поддержка Kerberos)

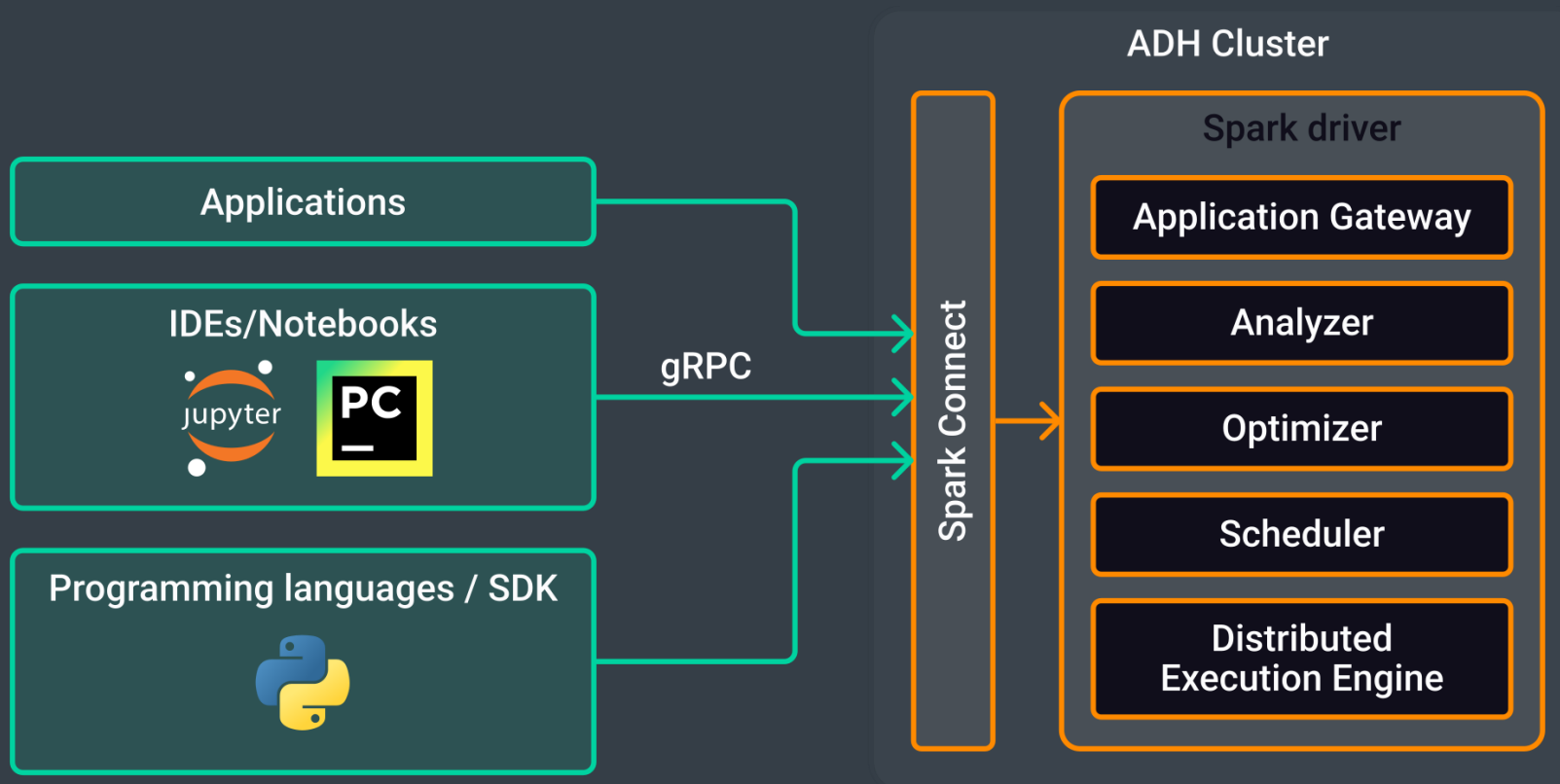
Кууби: место в платформе данных



* В разработке

Spark Connect: удалённое управление кластером Spark

Spark Connect — компонент сервиса Spark3, который, выполняя функции тонкого клиента, обеспечивает удалённое подключение к кластерам Spark. С помощью Spark Connect можно удалённо управлять кластером Spark, например, используя привычную IDE на обычном пользовательском ноутбуке.



Верхнеуровневая архитектура Spark Connect

Hue: веб-интерфейс экосистемы Hadoop для анализа данных



Hue (HUE: Hadoop User Experience) позволяет пользователям запрашивать, находить и анализировать имеющиеся данные без потери контекста.

В дистрибутиве ADH он содержит предустановленные интерпретаторы SQL для Impala, Hive, Kyuubi и Spark SQL.

Для кого: бизнес-аналитики, дата-инженеры и дата-сайентисты, администраторы баз данных, SQL-разработчики

Основные преимущества



Низкий порог входа

Интеллектуальные компоненты автозаполнения и редактор SQL-запросов упрощает работу с источниками данных



Универсальность

Hue подключается практически к любой СУБД или хранилищу данных через нативные коннекторы Thrift или SQLAlchemy, которые необходимо добавить в ini-файл



Функциональность

Файловый браузер для создания, копирования, перемещения файлов и директорий внутри HDFS, автоматизированный scheduler и платформа для разработки

Интерфейс Hue

The screenshot displays the Hue web interface for interacting with a data warehouse. It is divided into several main sections:

- Left Sidebar:** Contains navigation icons for Dashboard, Scheduler, Documents, Files, S3, Tables, Indexes, Jobs, Streams, HBase, Security, and Importer. A support link and email (romain@gethue.com) are also present.
- Top Bar:** Shows the Hue logo, a 'Query' dropdown, a search bar for saved documents, and a 'Jobs' button.
- Editor:** The central area where SQL queries are written. The current query is:

```
15 WHERE a.key = 'shipping' and a.zip_code = '76710';
16
17
18
19 -- Compute total amount per order for all customers
20 SELECT
21   c.id AS customer_id,
22   c.name AS customer_name,
23   o.order_id,
24   v.total
25 FROM
26   customers c,
27   c.orders o,
28   (SELECT SUM(price * qty) total FROM o.items) v;
```
- Tables:** A list of available tables in the 'default' database, including customers, kbs_logs, sample_07, sample_08, web_logs, and various log tables.
- Query History:** Shows a list of recent queries with their IDs and completion status.
- Results (106):** A table displaying the output of the first query:

	customer_id	customer_name	order_id
1	75012	Dorothy Wilk	4056711
2	75012	Dorothy Wilk	J882C2
3	17254	Martin Johnson	I72T39
4	12532	Melvin Garcia	PB6268
5	12532	Melvin Garcia	B8623C
6	12532	Melvin Garcia	R9S838
7	42632	Raymond S. Vestal	HS3124
8	42632	Raymond S. Vestal	BSS902
9	77913	Betty J. Giambrone	DN8815
10	77913	Betty J. Giambrone	XR2771
- Bottom Panel:** Shows a second query and its visualization:

```
1 select startstation,
2 count(*) as trips
3 from `201402_trip_data`
4 group by startstation,
5 endstation
6 order by trips desc
7 ;
8
```

The visualization is a bar chart showing the number of trips for different startstations. The Y-axis is labeled 'trips' and ranges from 0 to 1,33k. The X-axis lists startstations: Harry Bridges Plaza, Townsend at 7th, San Francisco Caltr..., Market at Sansome, and Embarcadero at San... The bars are colored blue, green, grey, pink, and teal respectively.

Iceberg: открытый формат таблиц для больших хранилищ данных



Apache Iceberg позволяет работать с файлами в data lake как со структурированными таблицами через SQL-запросы.

Обеспечивает ACID-транзакции, перемещение во времени (time travel), изменение схемы (schema evolution), изменение партиции (partition evolution) и другие возможности работы с данными.

Основные преимущества



Совместимость

Работает поверх ORC, Avro или Parquet и легко интегрируется с популярными движками Spark, Impala, Hive и др. Работает с любым облачным хранилищем и снижает нагрузку в HDFS.



Масштабируемость

Предназначен для огромных таблиц с десятками петабайт данных. Позволяет обрабатывать большие объёмы данных без потери производительности.



Целостность

Благодаря атомарности, консистентности и изолированности транзакций, операции в data lake, включая параллельную запись, не приведут к ошибкам или повреждению данных.

Основные возможности таблиц Iceberg

Выражения SQL

Iceberg поддерживает множество SQL-операций, которые позволяют выполнять такие задачи, как обновление строк, объединение данных, целевые удаления и др.

Изменение схемы таблицы

Изменение схемы таблиц влияет только на метаданные, а не на файлы данных. Это позволяет быстро добавлять, удалять, обновлять, реорганизовывать и переименовывать столбцы.

Изменение партиции

В отличие от стандартного партиционирования, таблицы Iceberg позволяют изменять схему партиционирования данных без перезаписи всех данных.

Снепшоты

Снепшоты позволяют сохранить состояние таблицы в конкретный момент. Iceberg ведет лог созданных снепшотов, что позволяет выполнять запросы time travel.

Перемещение во времени и откат изменений

Функция «перемещения во времени» позволяет выполнять запросы к более ранним версиям таблиц. Функция отката изменений позволяет пользователям восстанавливать состояние таблицы в определенный момент времени.

Транзакционная согласованность

Iceberg поддерживает ACID-транзакции, что обеспечивает безопасные параллельные записи в кластере и предотвращает влияние операций записи на операции чтения. Когда происходят изменения, Iceberg генерирует новую, неизменяемую версию файлов данных и метаданных таблицы.

Быстрое выполнение запросов

Iceberg повышает скорость выполнения запросов за счёт возможности организации инкрементальной обработки данных, быстрого планирования сканирования и фильтрации неактуальных данных.

Arenadata Hadoop vs “ваниль”

Преимущества Arenadata Hadoop в сравнении
с Open Source Hadoop

Преимущества Arenadata Hadoop в сравнении с Open Source



Качественная сборка совместимых компонентов

Hadoop — это множество сервисов, призванных взаимодействовать между собой. Самостоятельная сборка из исходников требует существенных вложений в RnD, либо будет выполнена без оглядки на совместимость, что скажется на стоимости эксплуатации и повлечет за собой простои.



Техническая поддержка от вендора с SLA, подтвержденным контрактом

Гарантии на поддержку платформы от вендора, включая штрафы, указанные в договоре, в сравнении с мотивацией команды собственных экспертов.



Универсальный оркестратор гибридного ландшафта

Предоставляем Arenadata Cluster Manager, объединяющий Hadoop и комплиментарные сервисы. ADCM можно развернуть на любой инфраструктуре, включая публичное облако.



Дополнительные возможности Enterprise версии

Высокопроизводительные коннекторы позволяют интегрировать Hadoop, ClickHouse, Greenplum и Tarantool между собой, а также с внешними системами.



Безопасность

Единая, интегрированная во все компоненты платформы, система безопасности Arenadata Platform Security на базе Kerberos, Ranger и Knox — в сравнении с частным решением, которое нужно постоянно дорабатывать и обновлять.

Преимущества Arenadata Hadoop в сравнении с Open Source



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), проведут аудит системы и помогут определить дальнейшие шаги.



Пакет утилит для полной офлайн-установки

Arenadata Hadoop включает набор инструментов для автоматической установки и настройки компонентов, как на «чистом железе», так и в облаке. Средства мониторинга и управления конфигурацией кластера позволяют оптимизировать производительность для всех компонентов системы.



Реестр отечественного ПО

Наличие регистрации продукта в реестре отечественного ПО не только отражает возможность снижения внешних рисков, но и напрямую влияет на стоимость для клиента за счёт экономии на НДС.



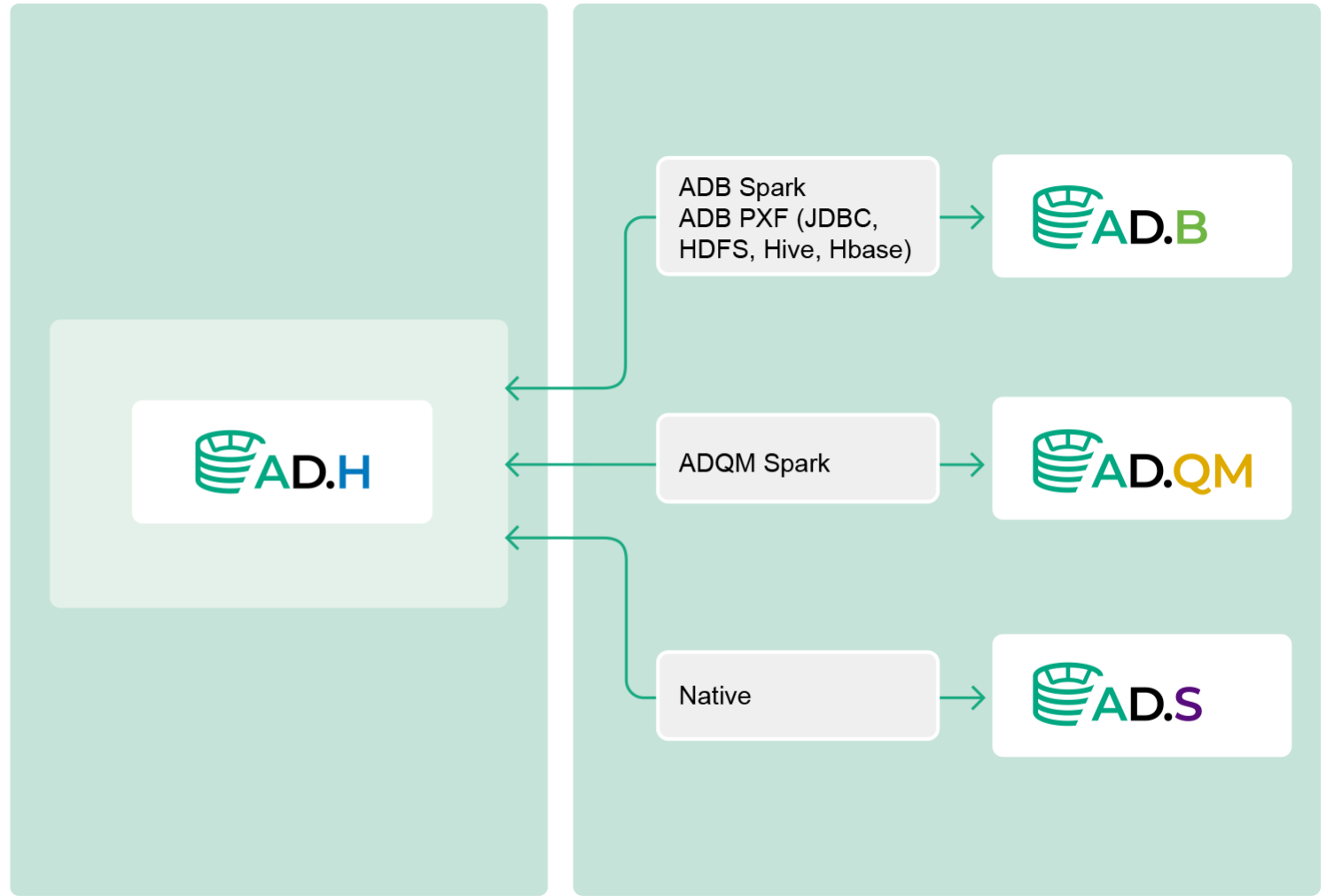
Документация

Оригинальная документация на русском и английском языках поможет облегчить процесс планирования, установки и настройки кластера Hadoop.

Коннекторы в платформе

Продукт Arenadata Hadoop обеспечен всеми необходимыми коннекторами для работы с другими компонентами платформы Arenadata EDP:

- Arenadata DB
- Arenadata Streaming
- Arenadata QuickMarts



* ADB PXF Hive и HBase коннектор работают только на чтение

Коннекторы в платформе

ADB Spark3 Connector

Обмен данными между Apache Spark и Arenadata DB

Интеграционное решение обеспечивает высокоскоростной параллельный обмен данными между Spark 3 и DWH на базе Arenadata DB (ADB).

ADQM Spark3 Connector

Обмен данными между Apache Spark и Arenadata QuickMarts

Многофункциональный коннектор с поддержкой параллельных операций чтения/записи между Spark 3 и Arenadata QuickMarts (ADQM).



Возможности:

- высокая скорость передачи данных
- автоматическое формирование схемы данных
- гибкое партиционирование
- поддержка push-down операторов
- поддержка batch-операций

Arenadata EDP



Arenadata Enterprise Data Platform

ИСТОЧНИКИ

ТРАНСПОРТ

ХРАНЕНИЕ И ПРЕДСТАВЛЕНИЕ ДАННЫХ

ИСПОЛЬЗОВАНИЕ И ВИЗУАЛИЗАЦИЯ

 **AD.C** Data Catalog, Business Glossary & Metadata Management



 **AD.S** ESB / MB / MQ

 **PICODATA** In-memory SQL

 **AD.PG** Relational DBMS

 **AD.QM** Fast flat marts

 **AD.B** Big Data Warehouse

 **AD.H** Cold / Unstructured

Any BI tools compatible with JDBC/ODBC, BI / Reporting

Ad-hoc

Data Science

Other solutions and services

 **AD.CM** Data & Cluster Management

On-premise















PAAS

Public Cloud

Private Cloud



Что внутри?

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

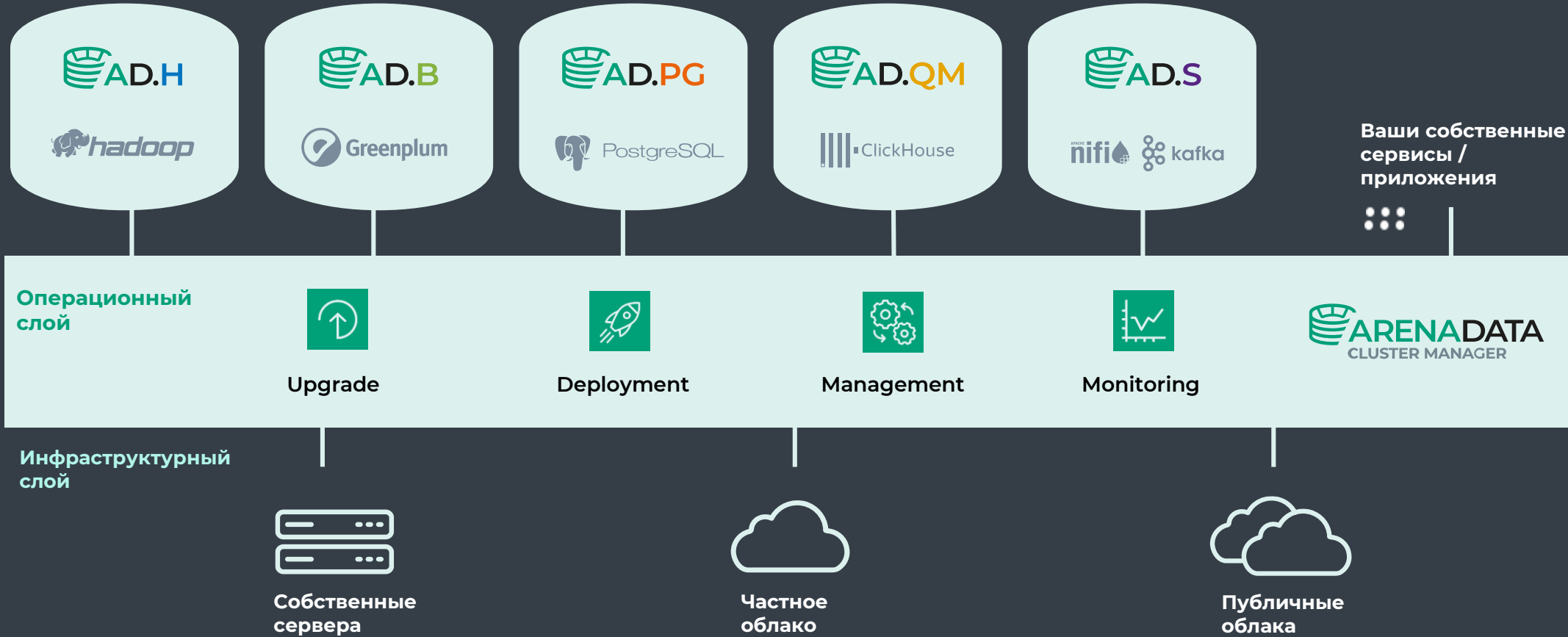
Каждый добавляемый компонент платформы доработан, чтобы стать её частью

Производится:

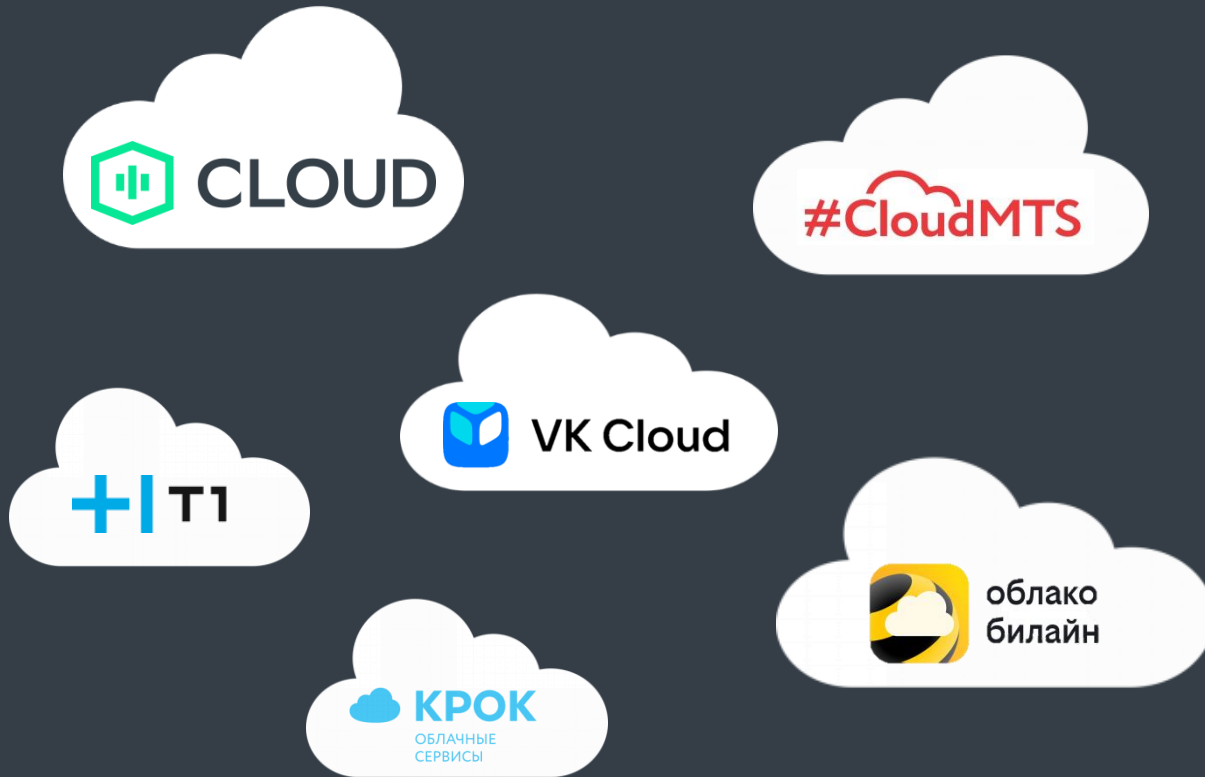
- Анализ кода
- Интеграция с модулем мониторинга
- Проверка работоспособности заявленного функционала
- Подбор компетентного персонала поддержки и погружение в продукт архитекторов Professional Services
- Разработка коннекторов для взаимодействия с другими компонентами платформы
- Формирование дорожной карты доработок продукта

Гибридный корпоративный ландшафт

Слой компонент данных



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 компоненты платформы распределённо на разные инфраструктуры, обеспечив между ними сетевую доступность.

Оставайтесь на СВЯЗИ



Наши новости
в телеграм-канале

