Arenadata Hyperwave

Универсальная гибридная платформа для хранения, обработки и анализа данных любой структуры и объёма

ARENADATA

План презентации

1. О компании Arenadata
2. Обзор Arenadata Hyperwave
3. Экосистема Arenadata Hyperwave
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

СТАБИЛЬНОСТЬ

ПАО Группа Arenadata:

- **9 лет** на рынке
- **600+** сотрудников
- **6 млрд ₽** выручка ’24
- **30 млрд ₽** капитализация

150+ КЛИЕНТОВ

- ВТБ, ПСБ, ГПБ, Т-Банк
- ГПН, Росатом, Норникель
- X5, Магнит, Ашан, Hoff
- Мегафон, МТС, HH.ru
- ФНС, ДИТ МСК, Росреестр

ПРОДУКТЫ

- Мы создаем **тиражируемые продукты,** которые внедряют наши партнеры и клиенты
- **170+ релизов** в год
- Обеспечиваем совместимость с **upstream opensource**

ONPREM & CLOUD

- On-premise и ПАКи
- Cloud: VK Cloud, Cloud.ru, Selectel, MWS, ТІ Cloud, K2 Cloud, Beeline Cloud
- Гетерогенные ландшафты

#DATA

Дата-платформа в доменах:

- **Аналитика, Транзакции, Интеграция, Data Governance**
- Техническая поддержка 1, 2 и 3 линии и сервисы от вендора

ЭКСПЕРТЫ

- **150+** технологических партнеров и интеграторов
- **4 000+** экспертов подготовлено в центре обучения Arenadata

ВКЛАД В OPENSOURCE

- **Greengage**: преемник Greenplum
- **ClickHouse**: #1 контрибутор в РФ
- **Greengage, Kyuubi, Spark, Kafka, SSM, PXF**: коммитим в ядро
- **ADCM, Picodata**: собственные open source-проекты
- Открытая онлайн документация

РЕГУЛЯТОРЫ

- Отечественные ОС
- Реестр Минцифры
- ФСТЭК УД4
- ГосТех, ГЕОП

Greenplum · ClickHouse · Apache Impala · S3 · Spark · Apache Ozone · hadoop · Prometheus · PostgreSQL · Flink · Grafana · Apache Ranger · GREENGAGE · ICEBERG · TARANTOOL · trino · APACHE KNOX · ANSIBLE · kafka · Open Metadata · Kyuubi · nifi · HUE · Apache Airflow · HIVE

Наш успех – доверие клиентов

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

Банки и страховые: 40+ клиентов

ВТБ · ГПБ · Т-Банк · Московская Биржа · МКБ · Отп Банк · ПСБ · РСХБ · РГС

Промышленность, ТЭК, Транспорт: 40+ клиентов

Северсталь · ALROSA · Газпром нефть · Норникель · Росатом · Почта России · СИБУР · ФОСАГРО

Госсектор: 30+ клиентов

ФНС · ДИТ Москвы · Минцифры · МинЗрав · Росреестр · ФОМС · СФР

Retail, FMCG, Фарма, Telecom, Digital, Media: 30+ клиентов

X5 Group · Магнит · Ашан · Hoff · Мегаполис · Бургер Кинг · Рив Гош

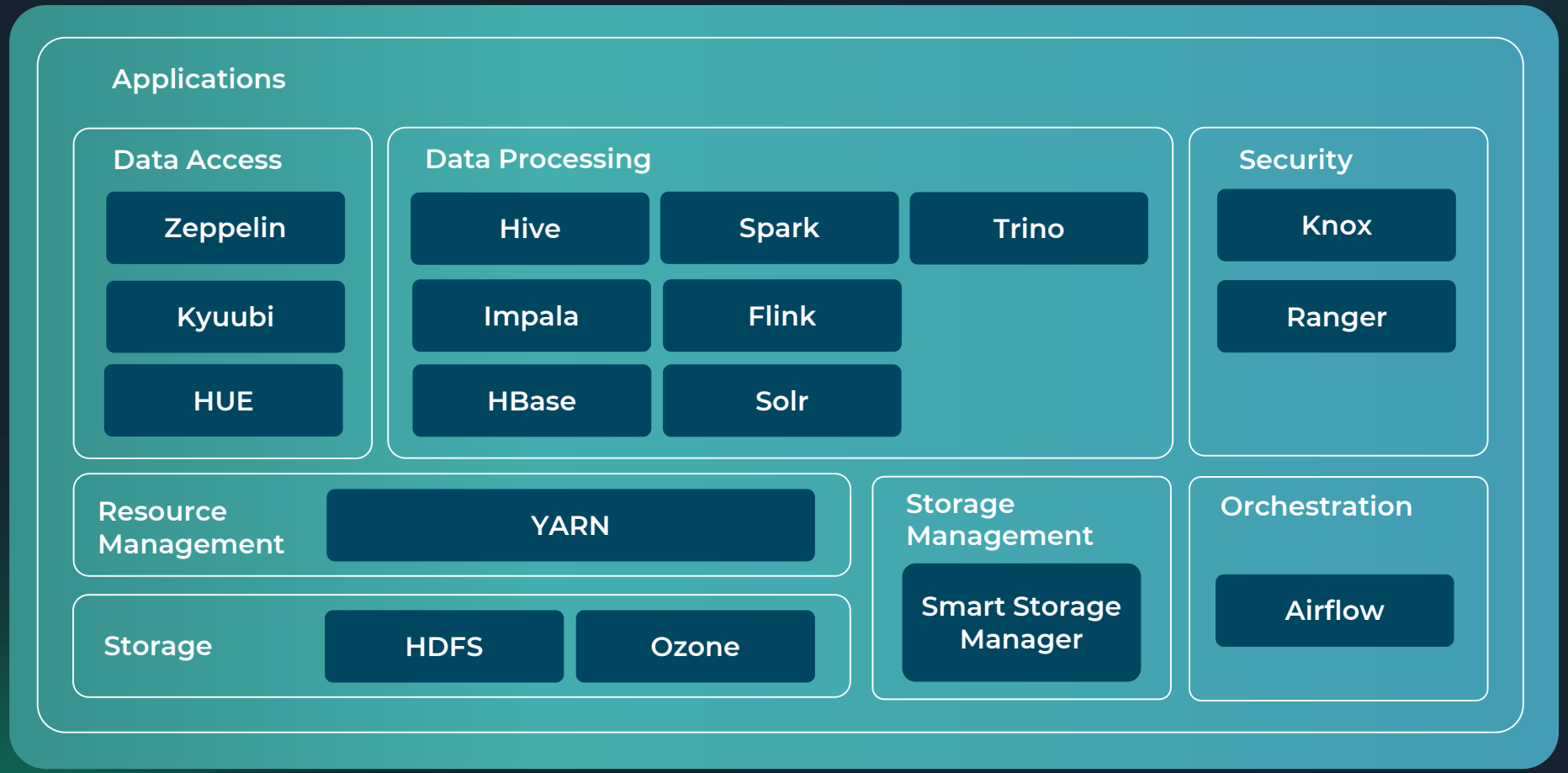
МегаФон · МТС · hh · Ситидрайв · Циан · НМГ · Skillbox

Обзор Arenadata Hyperwave

ARENADATA HYPERWAVE

Гибридная платформа данных

Универсальная платформа для хранения, обработки и анализа данных любой структуры и объёма



- Децентрализация компонентов
- Устранение жёсткой привязки к hadoop-сервисам
- Гибридная архитектура
- Федеративный доступ к данным
- Развитие Observability
- Основа Data Lake, Lakehouse, Data Mesh

Технические возможности Arenadata Hyperwave

Хранение данных

- Распределённое отказоустойчивое хранение
- Объектное хранилище с S3-совместимым интерфейсом
- Поддержка колоночных форматов для аналитических нагрузок
- Табличный формат Iceberg с поддержкой ACID-транзакций
- Горизонтальное масштабирование до экзабайт данных
- Поддержка горячего/холодного хранения
- Автоматизация управления хранением данных (многоуровневое хранение, репликация, сжатие, настройка политик Erasure Coding и др.)

Пользовательские интерфейсы

- Единые веб-интерфейсы для доступа ко всей платформе
- Интерактивная аналитика
- Совместимость с BI-инструментами
- Multitenant поддержка

Обработка и анализ данных

- Multi Engine обработка данных
- Единая платформа для пакетной и потоковой загрузки и аналитики
- Открытые и проприетарные коннекторы ко всем популярным системам
- Интеграция с единым метастором
- Федеративные распределённые запросы к разнородным источникам
- Low-latency для OLAP-запросов поверх HDFS/Ozone/S3
- Поддержка полнотекстового, векторного и геопространственного поиска

Безопасность

- Централизованное управление RBAC-политиками
- Интеграция с LDAP/AD/Kerberos
- Колоночное маскирование
- Шифрование данных и чувствительных конфигураций
- Трассировка всех операций и централизованный аудит системы
- Сертификация для корпоративных сред

Решаемые задачи

Хранение, обработка и анализ данных любой структуры и объёма

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

Построение озёр данных, lakehouse и data mesh

- 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 систем зарубежных вендоров

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

Экосистема Arenadata Hyperwave

Архитектурные паттерны и сценарии использования

ARENADATA

Data Lake

Подход предполагает наличие выделенного хранилища данных для оперативной отчётности, отдельного решения для быстрых SQL-запросов и озера данных для дешёвого хранения исторических данных и удобной интеграции.

Интеграция данных

Сбор и согласование данных из разнородных источников (IoT-датчики, веб-логи, финансовые транзакции, соцсети) в едином озере.

Анализ больших данных

Предварительная обработка и хранение «сырых» данных в озере данных, глубокий статистический и BI-анализ — в корпоративном хранилище.

Аналитика логов

Хранение сырых лог-файлов в озере и последующая их агрегация в хранилище данных для корреляции событий и оповещений об инцидентах.

Lakehouse

Универсальная платформа, объединяющая мощность классического хранилища с гибкостью озера данных. Позволяет выполнять любые нагрузки – от батчевой аналитики до стриминговых вычислений и машинного обучения.

GenAI и LLM

GenAI и LLM используют огромные объёмы неструктурированных данных. ADH предоставляет инструменты и инфраструктуру для их обработки и выполнения сложных запросов.

Обнаружение и предотвращение мошенничества

Потоковая обработка транзакций и событий с низкой задержкой, применение ML-моделей на лету

Анализ больших данных

ADH может использоваться для обработки и анализа больших объёмов данных. Полученная информация ценна для анализа поведения пользователей, рыночных трендов и других показателей.

Интеграция данных

Интеграция данных из различных источников и форматов в единое хранилище помогает бизнесу устранить разрозненность информации и обеспечить её согласованное представление.

Data Mesh & Multitenancy

Data Mesh превращает данные в продукты, а Arenadata Hyperwave обеспечивает инфраструктуру для этого подхода: доменные команды работают с изолированными данными через единый каталог, сохраняя автономность. Мультитенантность реализована на всех уровнях, что позволяет безопасно делить платформу между командами, партнёрами и окружениями, соблюдая compliance и оптимизируя затраты.

Корпоративные Data Mesh-инициативы

Доменные команды развивают собственные хранилища в рамках единого каталога, сохраняя автономию и единообразие метаданных.

Compliance и управление доступом

Тонкая гранулярная настройка политик безопасности и аудита.

Безопасное разделение среды

Физическая или логическая изоляция кластеров под разные бизнес-юниты, партнёров или окружения. От разделения на уровне хранилища и ресурсов до физического выделения кластера под отдельный домен.

Экосистема Arenadata Hyperwave

Компоненты и сервисы

ARENADATA

Компоненты ADH

		ADH 3.2.4.3	ADH 3.3.6.1	ADH 3.3.6.2	ADH 4.0.0	ADH 4.1.0
Storage	HDFS	3.2.4	3.3.6	3.3.6	3.3.6	3.4.2
	Ozone	-	-	1.4.1	1.4.1	2.0.0
Resource Management	YARN	3.2.4	3.3.6	3.3.6	3.3.6	3.4.2
Coordination	ZooKeeper	3.5.10	3.8.4	3.8.4	3.8.4	3.8.4
Service Management	Smart Storage Manager	1.6.0	2.0.0	2.0.1	2.1.0	2.2.0
Data Access	Zeppelin	0.11.1	0.11.1	0.11.1	0.11.2	0.11.2
	Kyuubi	1.8.1	1.9.0	1.9.0	1.10.1	1.10.2
	HUE	-	4.11.0	4.11.0	4.11.0	4.11.0
Data Processing	Hive	3.1.3	4.0.0	4.0.0	4.0.1	4.0.1
	Spark2	2.3.2	2.3.2	2.3.2	-	-
	Spark3	3.4.2	3.5.1	3.5.1	3.5.4	3.5.6
	Spark4	-	-	-	-	4.0.0
	Impala	4.4.0	4.4.0	4.4.0	4.5.0	4.5.0
	Flink	1.18.1	1.19.1	1.19.1	1.20.1	1.20.2
	Flink2	-	-	-	-	2.0.0
	HBase	2.4.17	2.5.8	2.5.8	2.5.8	2.6.3
	Solr	8.11.2	8.11.3	8.11.3	8.11.3	8.11.3
	Sqoop	1.4.7	1.4.7	1.4.7	-	-
	Trino	-	-	468	468	476
Orchestration	Airflow2	2.6.3	2.6.3	2.6.3	2.6.3	2.6.3
Security	Knox	1.6.0	2.0.0	2.0.0	2.0.0	2.0.0
	Ranger	2.4.0	2.5.0	2.5.0	2.6.0	2.6.0
	Kerberos	latest	latest	latest	latest	latest
Cluster Management	ADCM	latest	latest	latest	latest	latest

Предыдущие версии

ADH 3.2.4.3

ADH 3.3.6.1

ADH 3.3.6.2

Текущая версия

ADH 4.0.0

Будущая версия

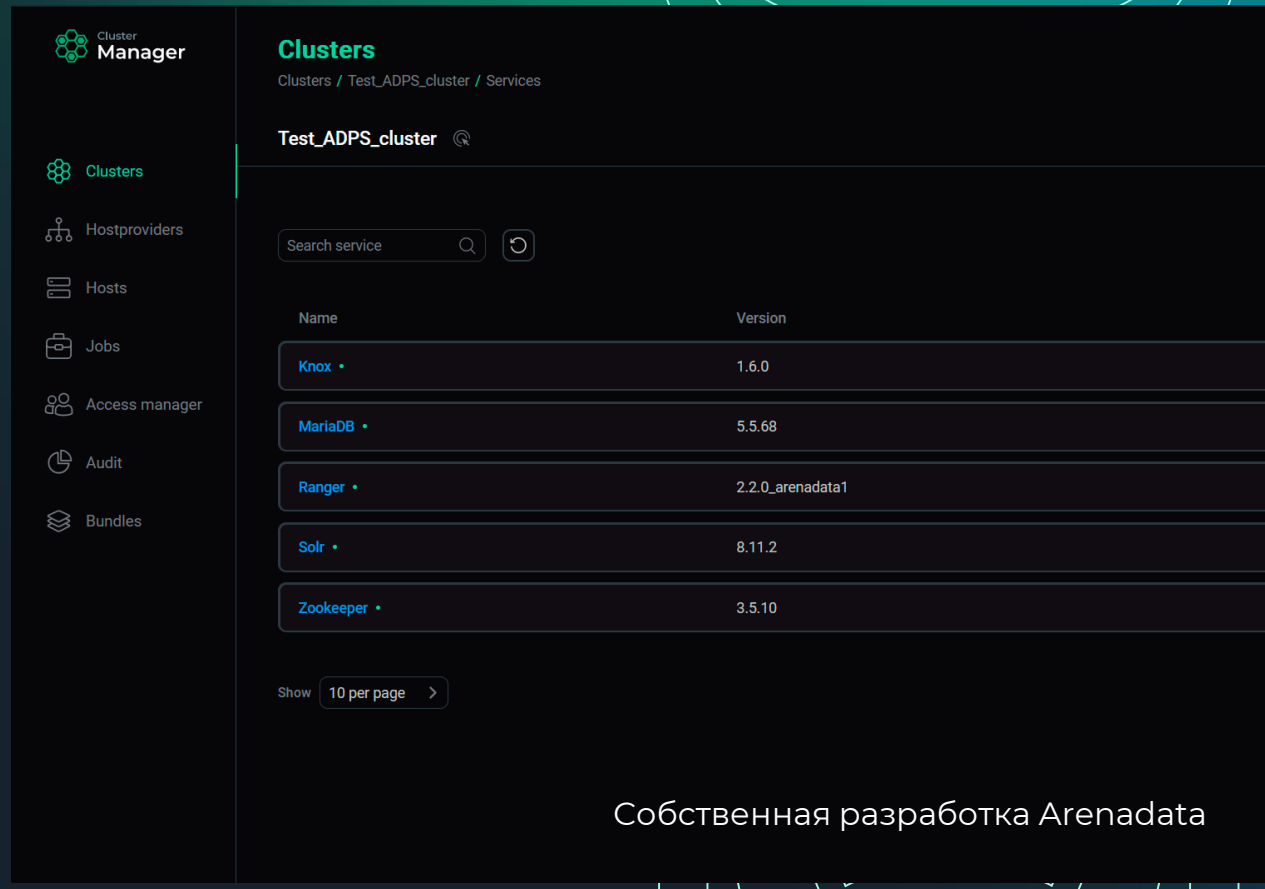
ADH 4.1.0

ARENADATA
PLATFORM SECURITY

Система управления групповыми политиками безопасности

Система централизованного управления политиками безопасности кластера ADH включает:

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



Собственная разработка Arenadata

Поставляется как бесплатное дополнение к enterprise-редакции ADH

Функции безопасности Arenadata Platform Security

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

Защита конфиденциальных данных и соответствие нормативным требованиям

Безопасность периметра

- Apache Knox
- Gateway

Аутентификация

- Kerberos
- LDAP/AD
- FreeIPA
- Samba
- MIT KDC

Защита данных

- SSL
- Шифрование RPC
- Шифрование at Rest

Аудит и мониторинг

- Запросы доступа
- Операции обработки данных
- Изменение данных

Авторизация

- Контроль доступа ко всем сервисам дистрибутива
- Контроль доступа на уровне баз данных, таблиц, столбцов для наборов данных SQL-движков
- Контроль доступа Knox
- Контроль доступа к коллекциям Solr

Impala: распределённая система исполнения SQL-запросов

Impala обеспечивает быстрые интерактивные SQL-запросы к данным, хранящимся в HDFS, HBase или S3-хранилище. Является унифицированной платформой для запросов в режиме реального времени или пакетных запросов.

Высокая скорость обработки запросов

Низкая задержка и высокий уровень параллелизма позволяет эффективнее решать задачи self-service аналитики и ad-hoc запросов

Снижение стоимости обработки данных

Достигается за счёт оптимального использования аппаратного обеспечения

Простое внедрение в имеющуюся инфраструктуру

Impala использует те же метаданные, синтаксис SQL (Hive SQL) и драйвер JDBC, что и Hive

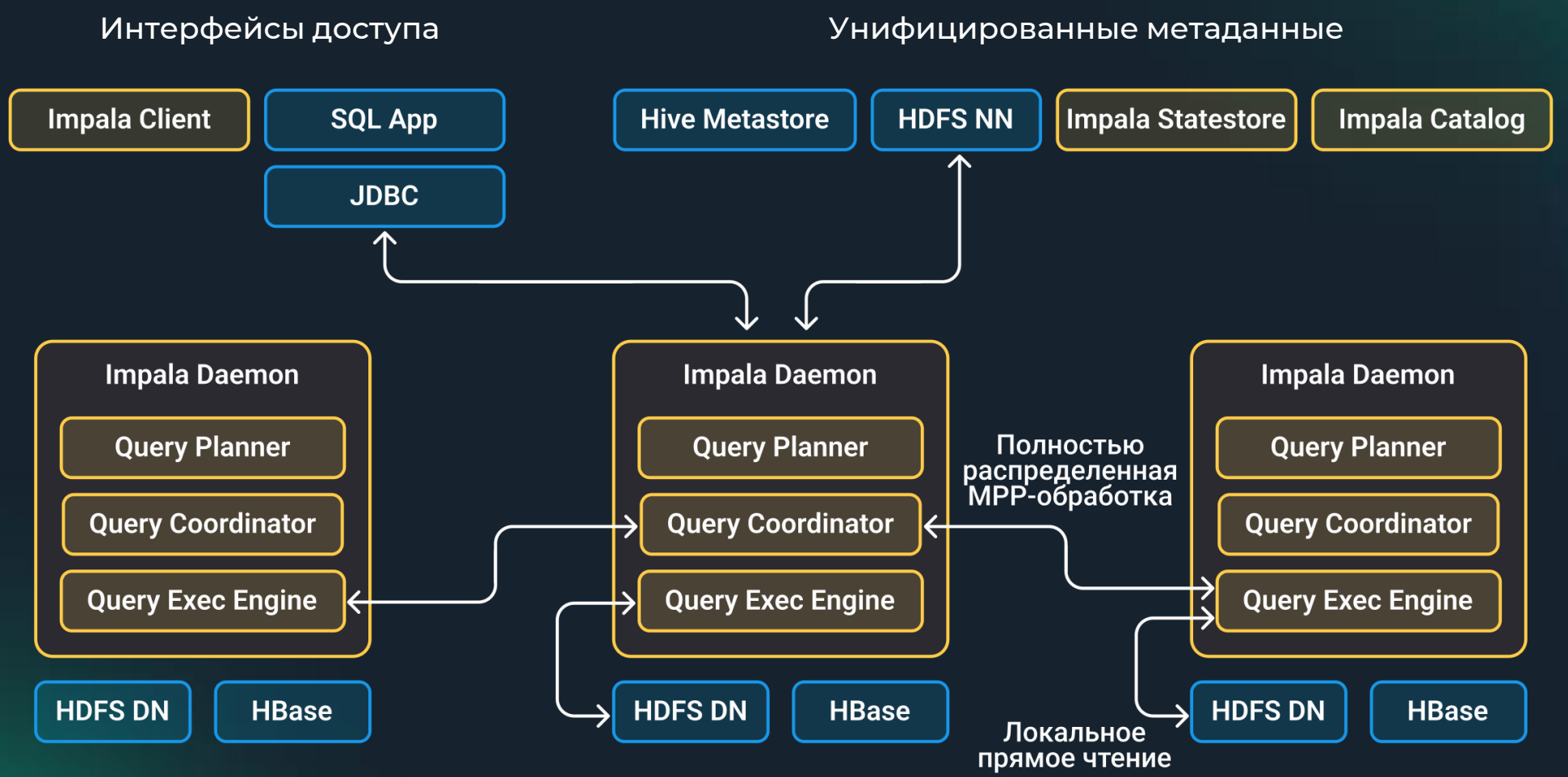
Масштабирование аналитической нагрузки

Развёртывание Impala вне основного кластера позволяет исключить конкуренцию за аппаратные ресурсы

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)

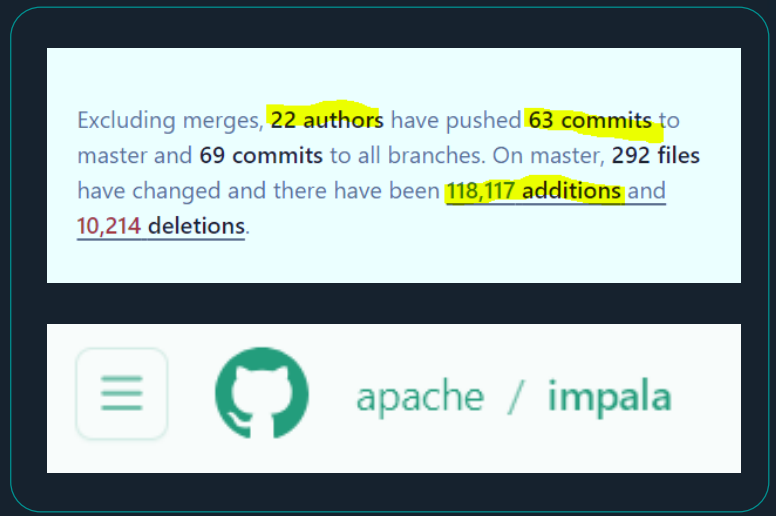
- MPP-архитектура на Hadoop
- Механизм Short-Circuit (только C++ библиотеки)
- Кэширование данных
- Кэширование метаданных
- Вычисления в памяти

Конкурентный доступ – Ad Hoc-analysis

- Минимальная утилизация ОП
- Механизм кэширования данных ОС

Перспективный open source проект

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



Варианты развёртывания Impala

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

1

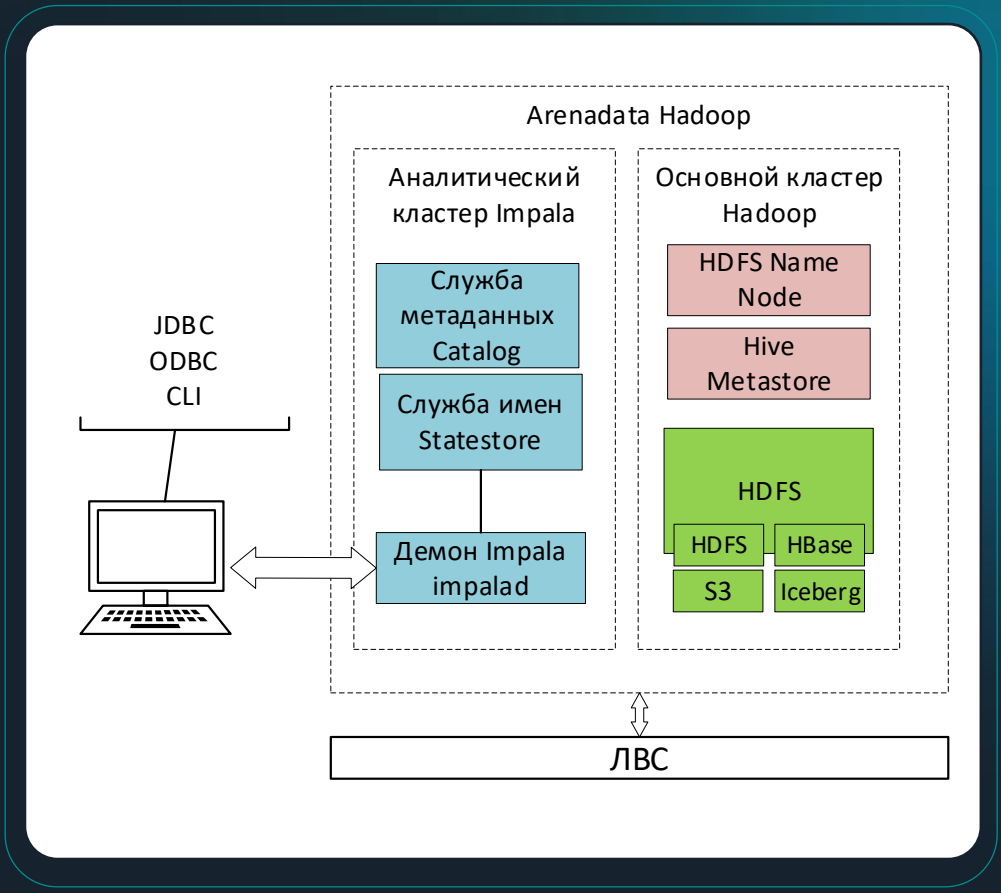
Совместное с HDFS развёртывание это:

- 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

Температурное хранение данных

Оптимизация управления данными в зависимости от их востребованности:

- перемещение наиболее горячих данных в кэш
- 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

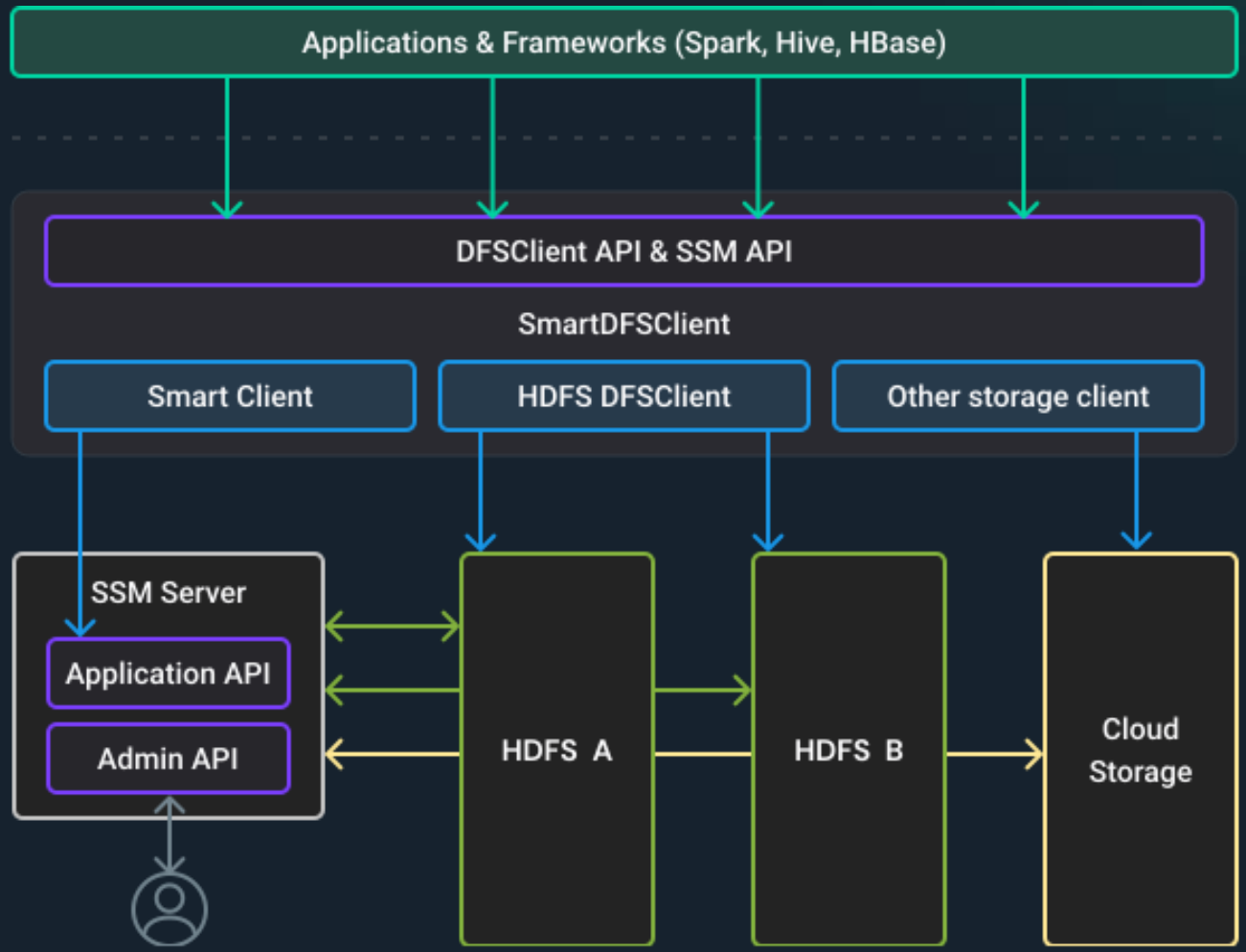
- Снижение стоимости хранения холодных данных
- Повышение производительности чтения горячих данных
- Простая настройка и управление функцией асинхронной репликации
- Снижение накладных расходов и повышение производительности записи и чтения небольших файлов HDFS
- Экономия места в хранилище

Оптимизация затрат

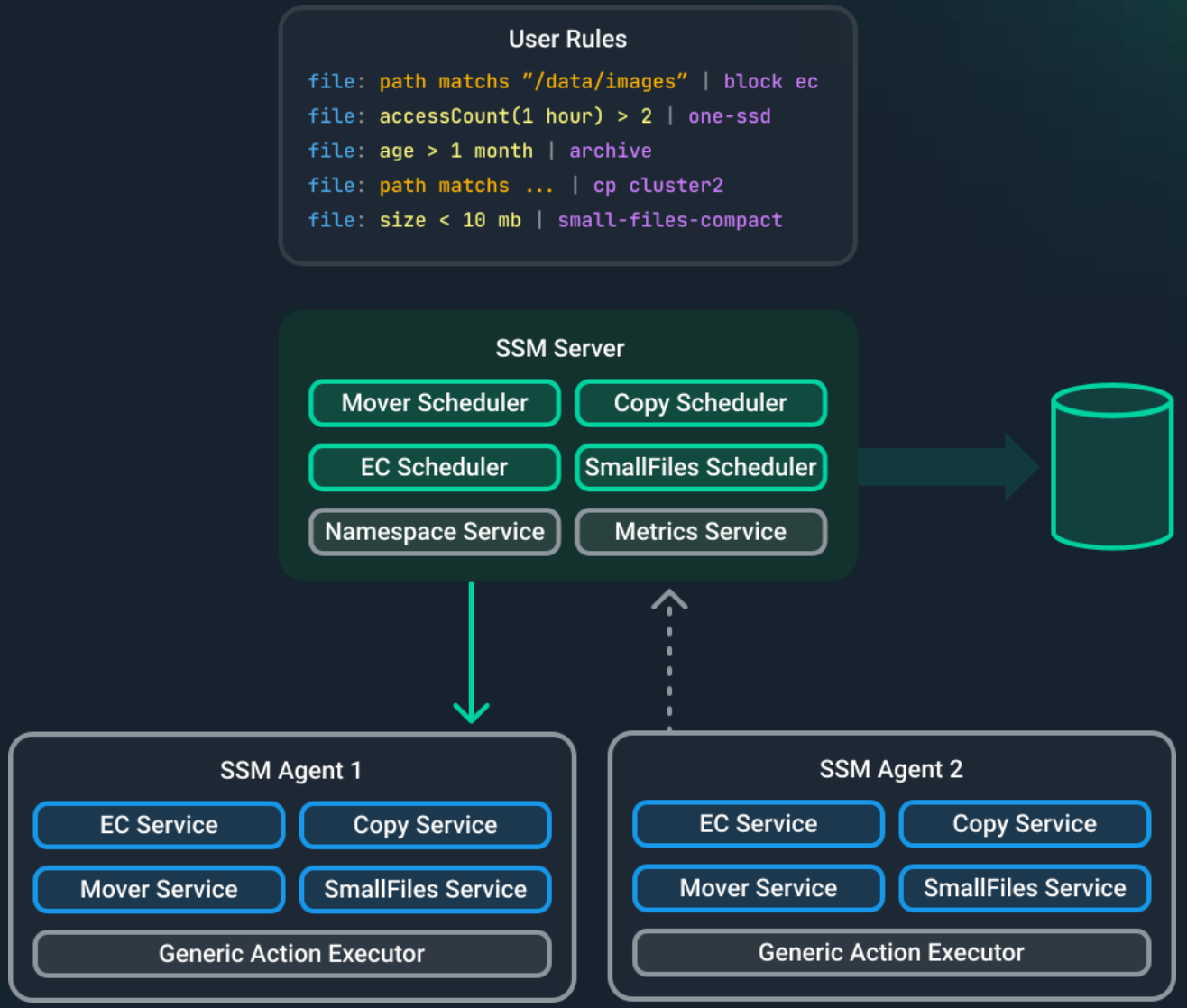
Повышение производительности

Надёжность

Архитектура Smart Storage Manager



Взаимодействие компонентов Smart Storage Manager



Kyuubi: SQL gateway для Data Lake и Lakehouse

Kyuubi — распределённый многопользовательский шлюз для предоставления SQL для Data Lake и Lakehouse.

Kyuubi 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 Kyuubi

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

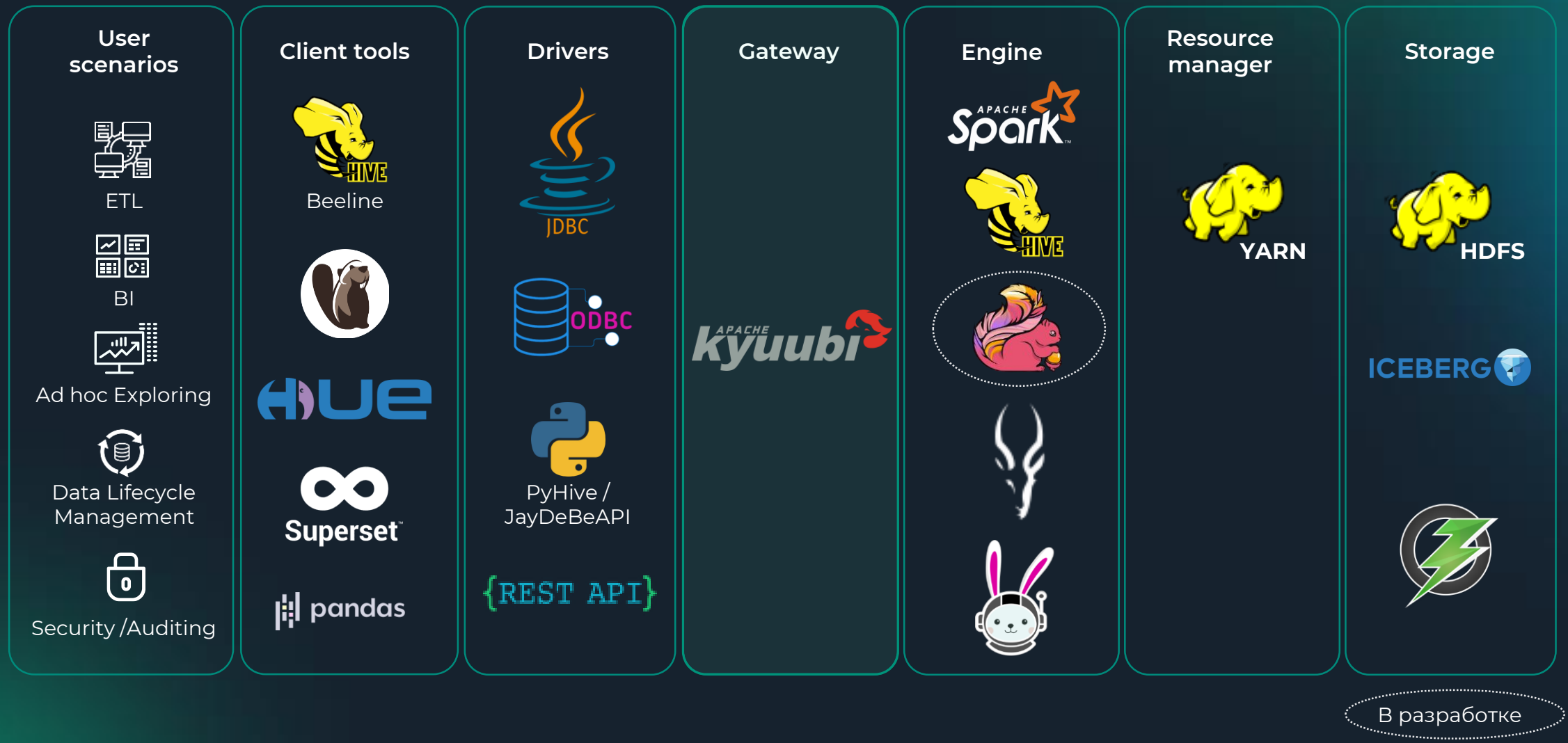
Пакетная обработка

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

Data Lake & Lakehouse

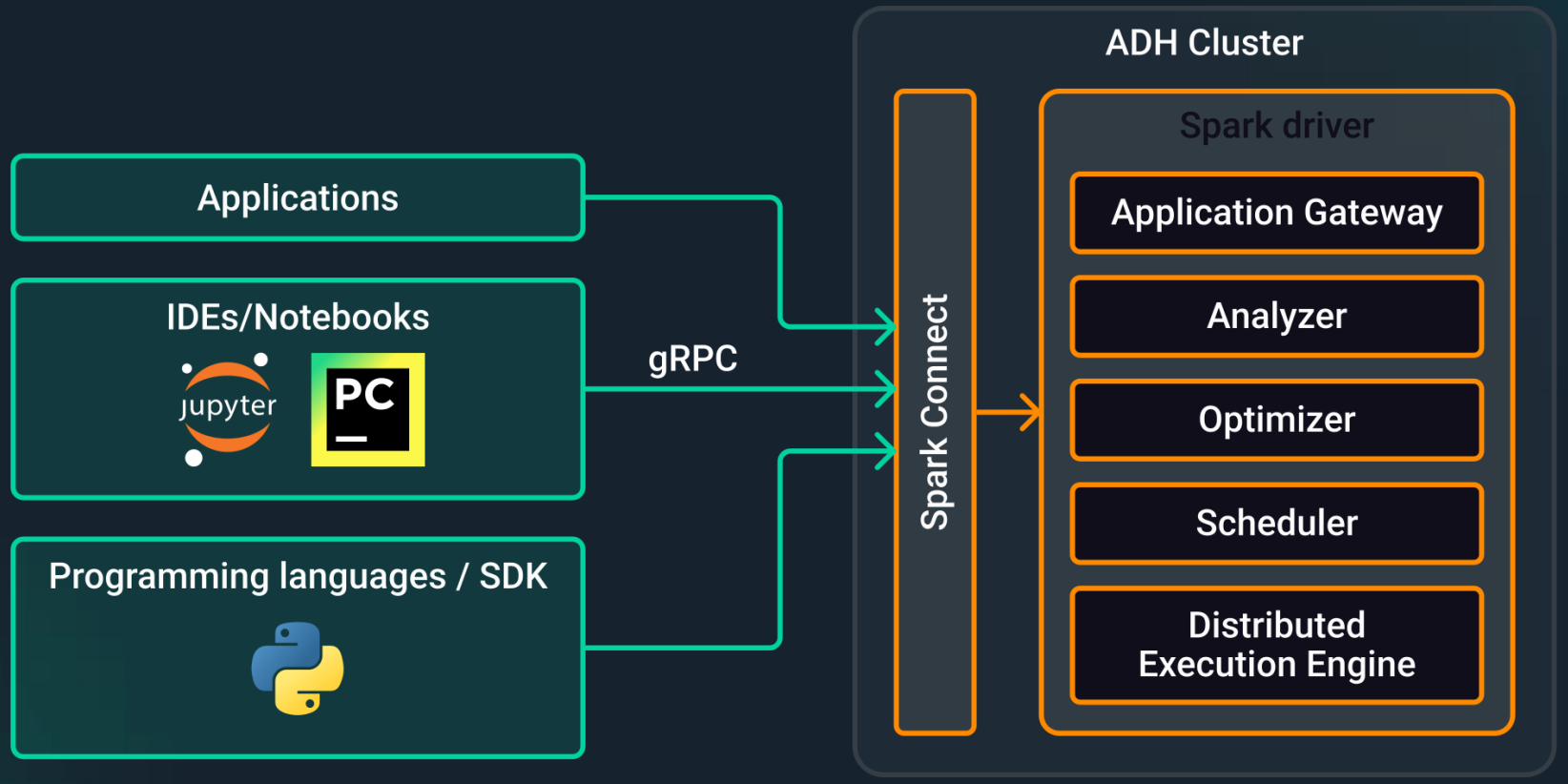
- Возможность выполнения запросов к традиционным хранилищам (Hive/HDFS) и современным озёрам данных
- Централизованная картина данных
- Возможность запрашивать разнородные источники данных
- Аутентификация и авторизация (поддержка Kerberos)

Kyuubi: место в платформе данных



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 для анализа данных

HUE (Hadoop User Experience) позволяет пользователям запрашивать, находить и анализировать имеющиеся данные без потери контекста. В дистрибутиве ADH он содержит преднастроенные 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

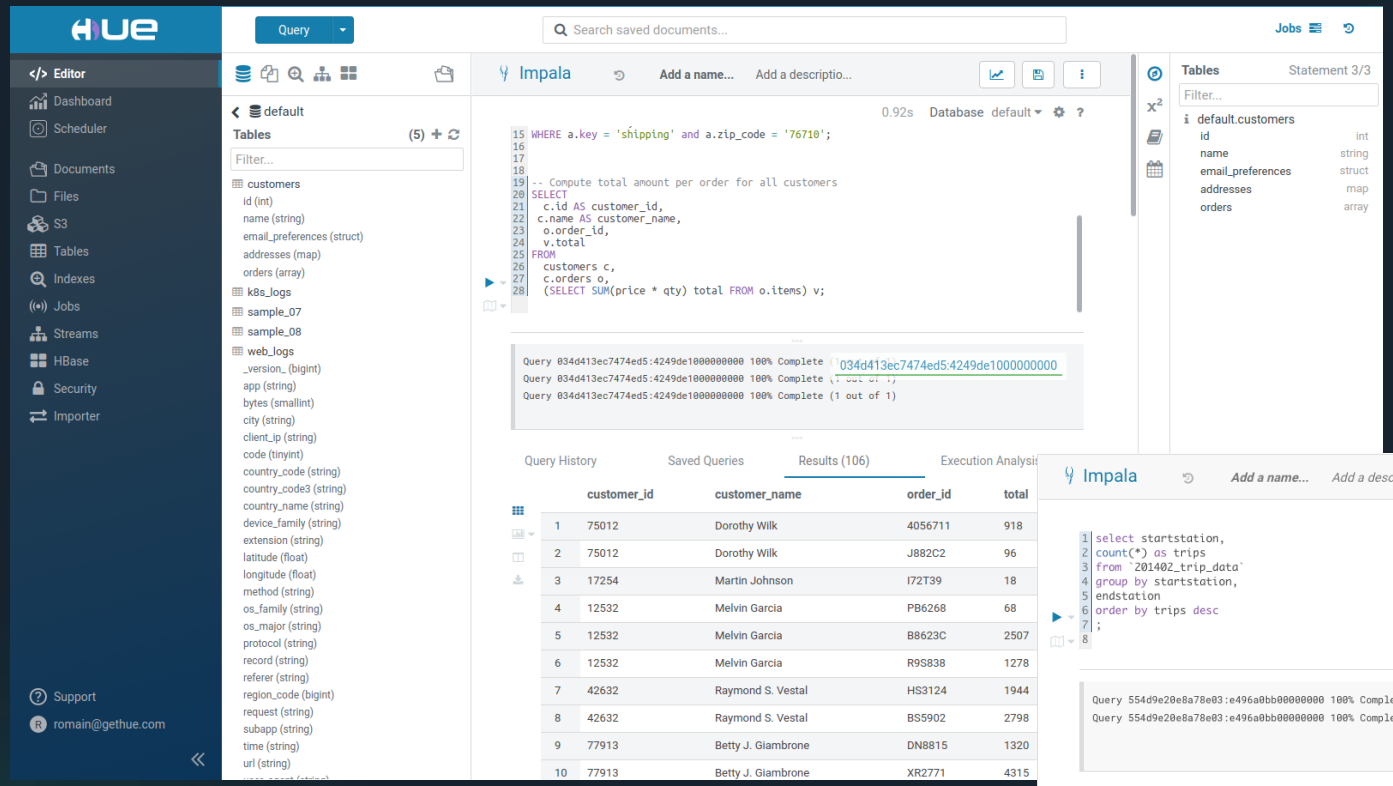
Универсальность

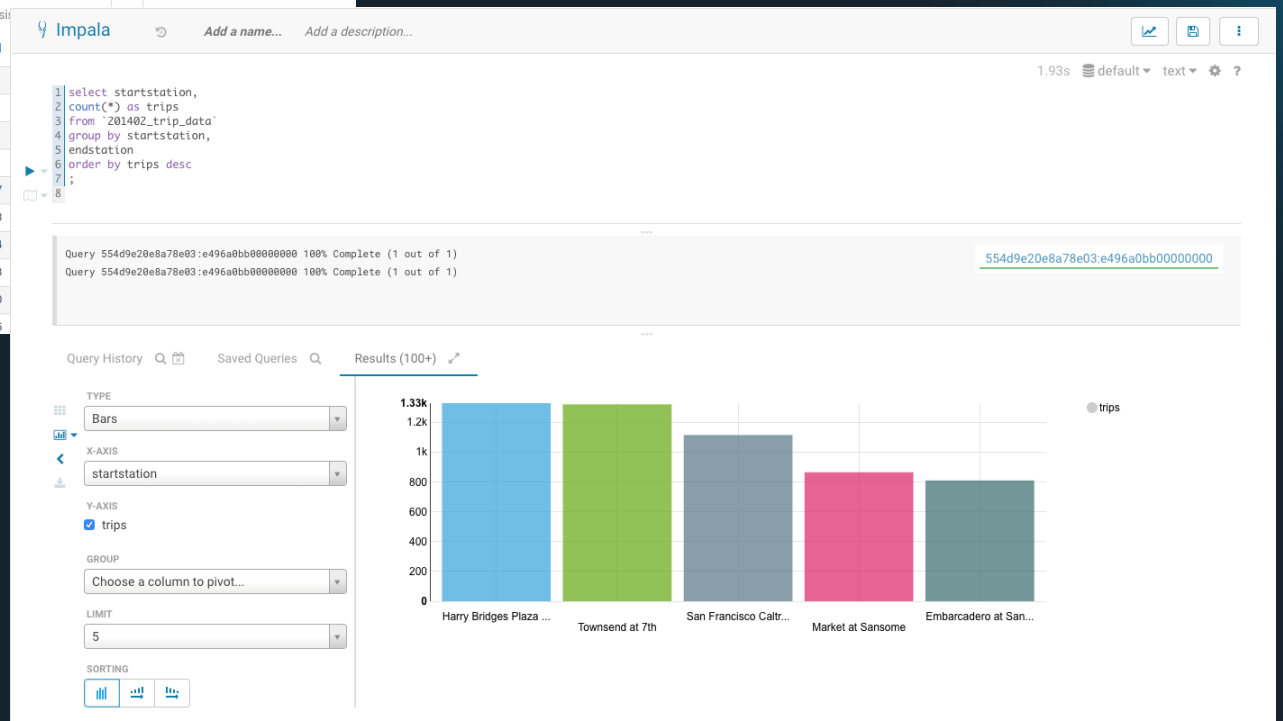
HUE подключается практически к любой СУБД или хранилищу данных через нативные коннекторы Thrift или SqlAlchemy, которые необходимо добавить в ini-файл

Функциональность

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

Интерфейс HUE





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.

Ozone*: новое поколение хранилища для платформы больших данных

Масштабируемое распределённое объектное хранилище

Масштабируемость

Предназначен для работы с миллиардами небольших файлов (в отличие от файловой системы HDFS, ориентированной на хранение файлов большего размера)

Высокая доступность

Отказоустойчивость и быстрое восстановление работы системы после сбоя без потери данных

Поддержка нескольких протоколов

Поддерживает работу как со стандартным HDFS-протоколом, так и S3 API

Лучшие гарантии консистентности

Достигается путём использования консенсус-протоколов, включая RAFT

Безопасность

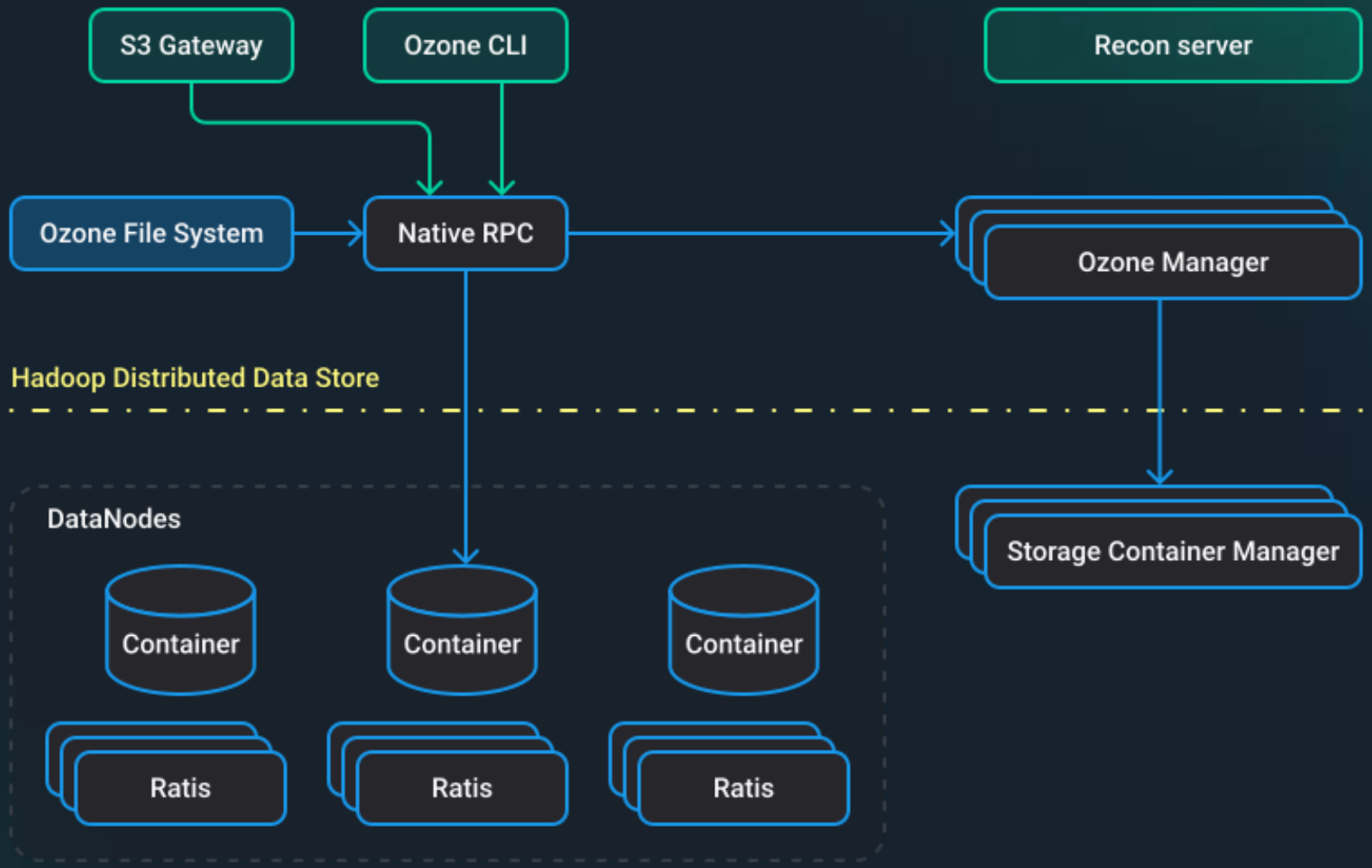
Поддержка Kerberos, Transparent Data Encryption и Ranger для управления безопасностью платформы

Совместное развёртывание

Может работать вместе с HDFS на одних и тех же хостах

* Поставляется в составе enterprise-редакции ADH

Архитектура Ozone



Сравнение ключевых возможностей HDFS и Ozone

	HDFS	Ozone
Модель данных	Файловое хранилище с иерархической структурой	Хранилище объектов, работает с неструктурированными данными и оптимизировано для облака
Репликация данных	Репликация данных между узлами DataNode для предоставления отказоустойчивости	Программно определяемое хранилище, позволяет настраивать политики репликации и обеспечивать избыточность данных
Масштабируемость	Хорошая масштабируемость для больших задач по обработке данных	Разработан с целью предоставить ещё более хорошую масштабируемость, чем в HDFS
Управление пространством имён	Единое пространство имён для всего кластера	Несколько пространств имён для различных задач
Хранилище объектов	нет	да
Поддержка S3 и других протоколов хранилищ объектов	нет	да
Управление доступом	Права в стиле POSIX	Права в стиле S3 и управление доступом на уровне бакета
Аутентификация и авторизация	Kerberos	Kerberos, Ozone Token
Согласованность данных	Согласованность рано или поздно достигается	Высокая согласованность за счёт таких протоколов, как RAFT

Области применения HDFS и Ozone

Рабочие нагрузки с меньшими требованиями к хранению небольших объектов без возможности их объединения

HDFS

Файловая система по умолчанию со следующими преимуществами:

- поддержка хранения большого количества данных
- быстрое определение и реагирование на аппаратные сбои
- поддержка потоковой передачи данных
- упрощённая модель согласованности
- высокая отказоустойчивость и лёгкость восстановления
- предназначена для коммерческого оборудования

Среды, где требуется масштабируемость для небольших файлов и совместимость с S3

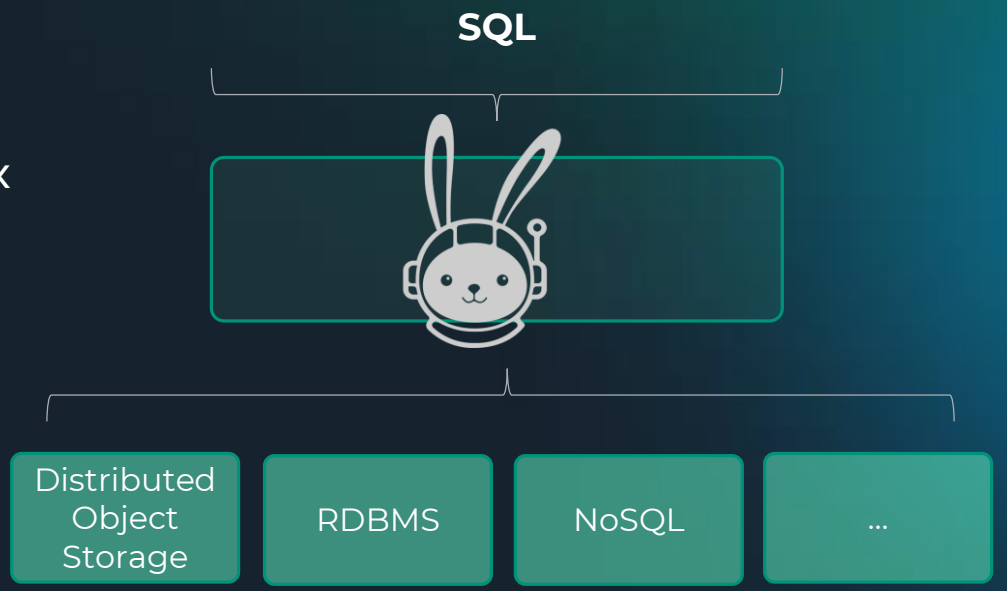
Ozone

Новое поколение хранилища для платформы больших данных:

- высокая согласованность данных
- разработано для хранения более 100 миллиардов объектов в одном кластере
- отличная масштабируемость благодаря многоуровневой архитектуре
- столь же высокая отказоустойчивость и лёгкость восстановления, как у HDFS
- может работать вместе с HDFS на одних и тех же хостах

Trino: единая точка доступа для запросов к хранилищам и озёрам данных

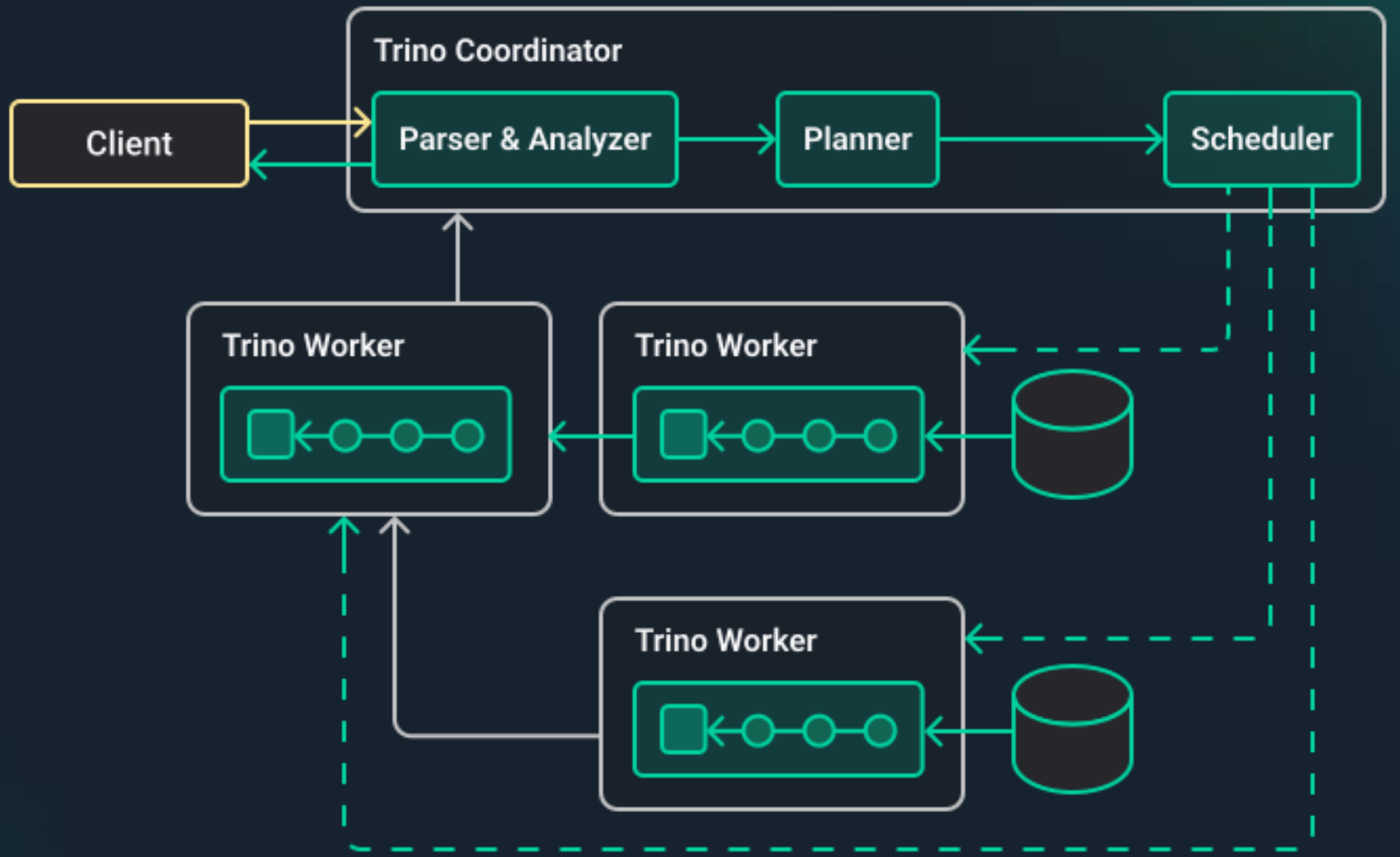
Trino — инструмент для обработки огромных объёмов данных с использованием распределённых федеративных запросов. Позволяет запрашивать разрозненные источники данных в одной системе с помощью одного и того же SQL — объектные хранилища, базы данных, файловые системы — в одном запросе



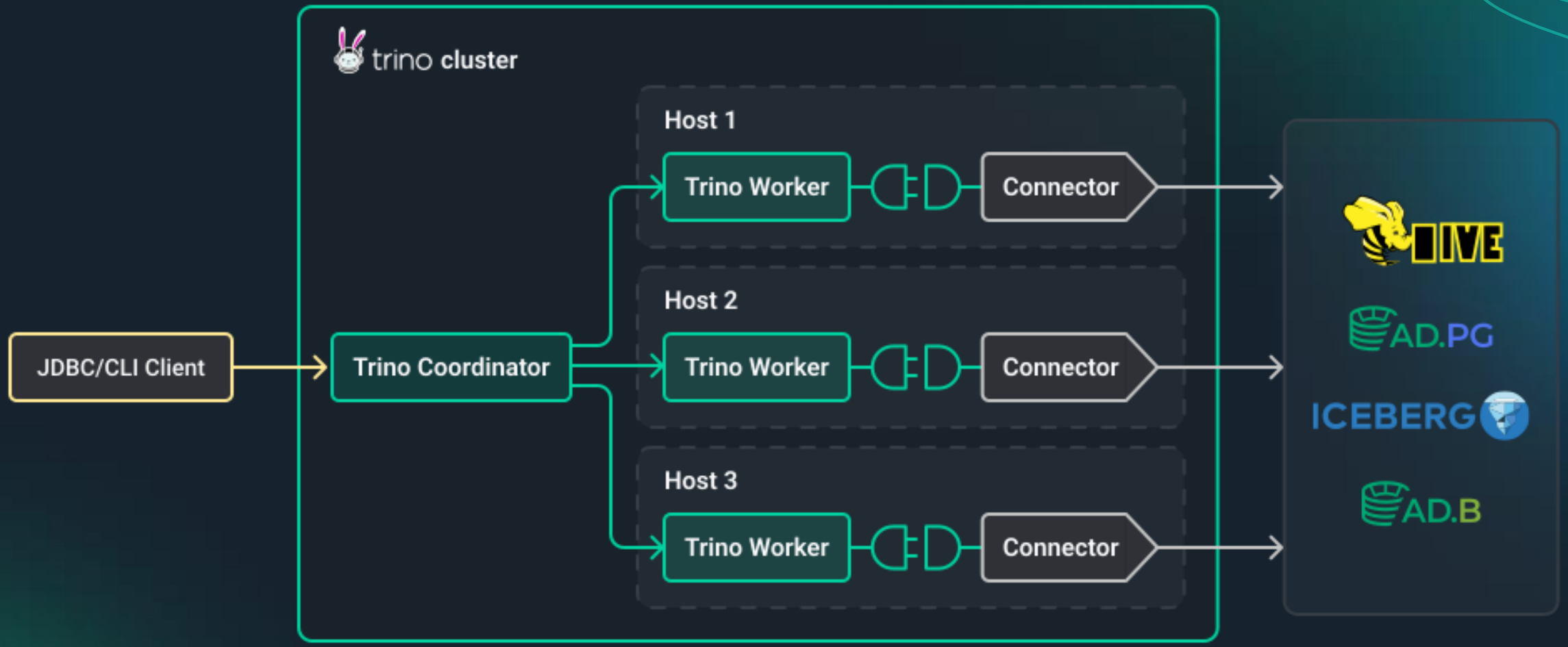
Основные преимущества

- перенос аналитической нагрузки из дорогих в обслуживании и трудно масштабируемых КХД в озёра данных
- уменьшение дублирования данных
- новые сценарии анализа данных при сокращении расходов на инфраструктуру

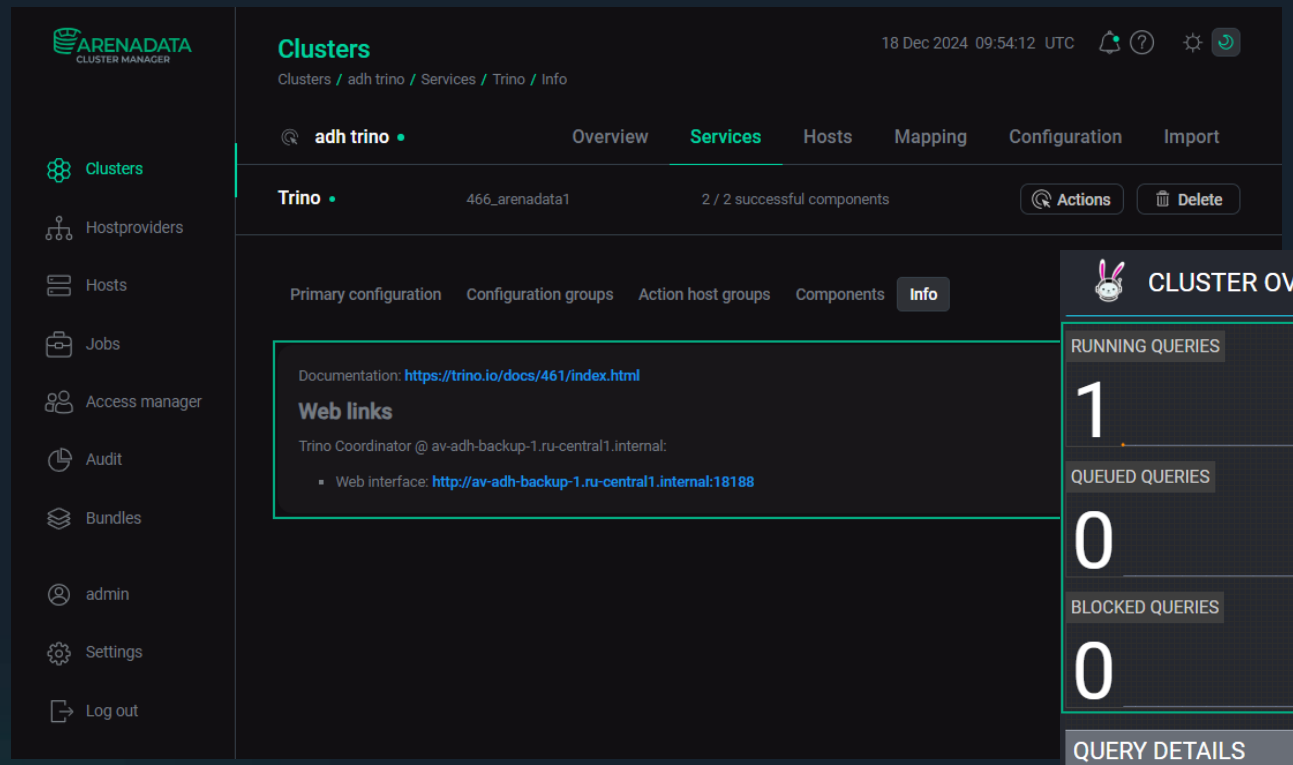
Архитектура Trino



Коннекторы в Trino

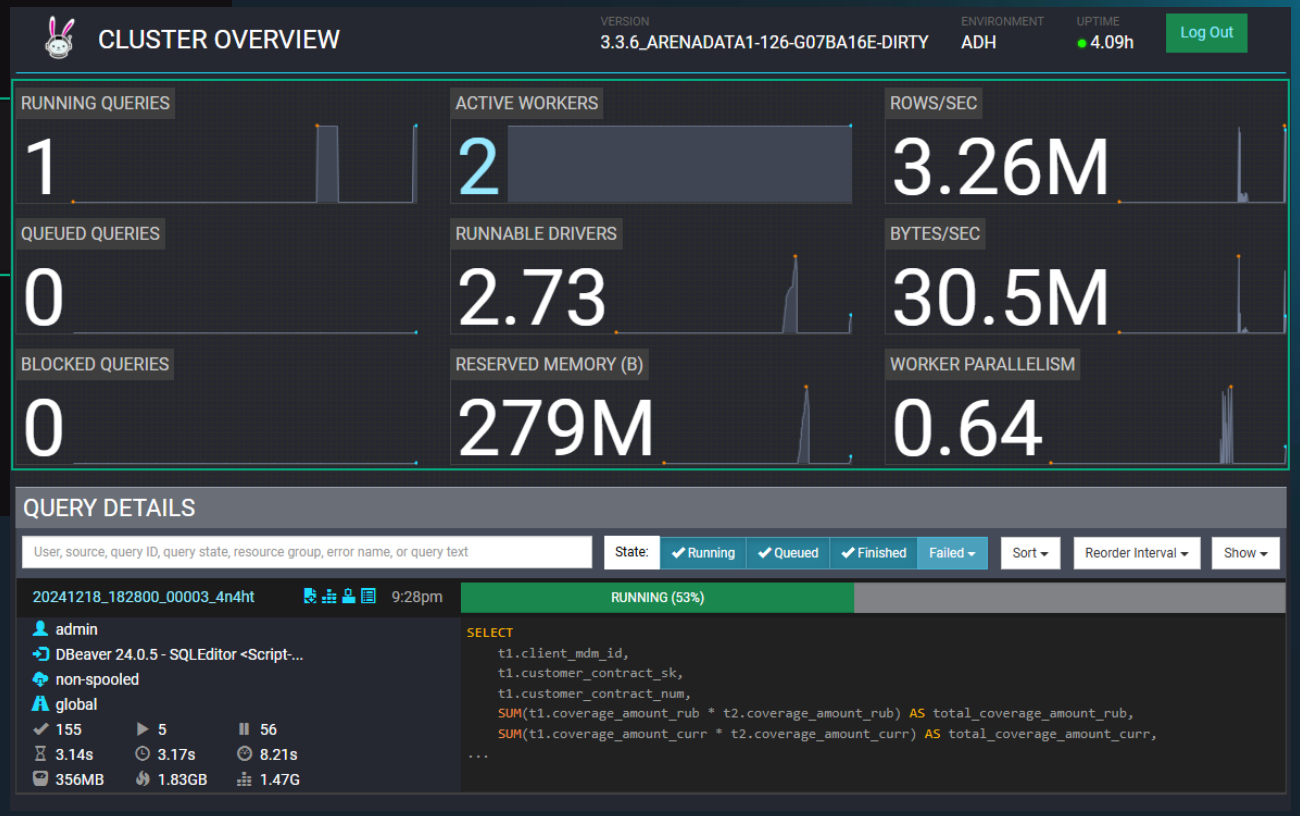


Trino UI



Ссылка на веб-интерфейс Trino в веб-интерфейсе ADCM

Панель мониторинга Trino



Arenadata Hyperwave vs «ваниль»

Преимущества Arenadata Hyperwave
в сравнении с Open Source

ARENADATA

Преимущества Arenadata Hyperwave в сравнении с Open Source

Качественная сборка совместимых компонентов

Дистрибутив ADH включает последние стабильные версии компонентов экосистемы Hadoop и ряд других OSS инструментов и проприетарных решений. Самостоятельная сборка сопоставимой по функциональности платформы требует существенных вложений в RnD, либо будет выполнена без оглядки на совместимость, что скажется на стоимости эксплуатации и повлечёт за собой простои.

Техническая поддержка от вендора с SLA, подтверждённым 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, универсальный инструмент для установки, настройки и обновления всех продуктов Arenadata на любой инфраструктуре с удобным и современным графическим интерфейсом.

Дополнительные возможности Enterprise версии

Высокопроизводительные коннекторы позволяют интегрировать Arenadata Hyperwave с другими продуктами Arenadata и 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 Hyperwave в сравнении с Open Source

Набор типовых пакетных сервисов по планированию, установке и аудиту системы

Вам не придётся самостоятельно проводить оценку оборудования для решения поставленной задачи. Наши специалисты настроят Arenadata Hyperwave (удалённо или on-site), проведут аудит системы и помогут определить дальнейшие шаги.

Пакет утилит для полной офлайн-установки

Arenadata Hyperwave 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.

Российское ПО

Продукт внесён в реестр сертифицированных средств защиты информации ФСТЭК России (по 4 УД) и в реестр отечественного ПО.

Документация

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

Коннекторы в платформе

ADB Spark Connector

Обмен данными между Apache Spark и Arenadata DB

Интеграционное решение обеспечивает высокоскоростной параллельный обмен данными между Spark3 и DWH на базе Arenadata DB (ADB)

ADQM Spark Connector

Обмен данными между Apache Spark и Arenadata QuickMarts

Многофункциональный коннектор с поддержкой параллельных операций чтения/записи между Spark3 и Arenadata QuickMarts (ADQM)

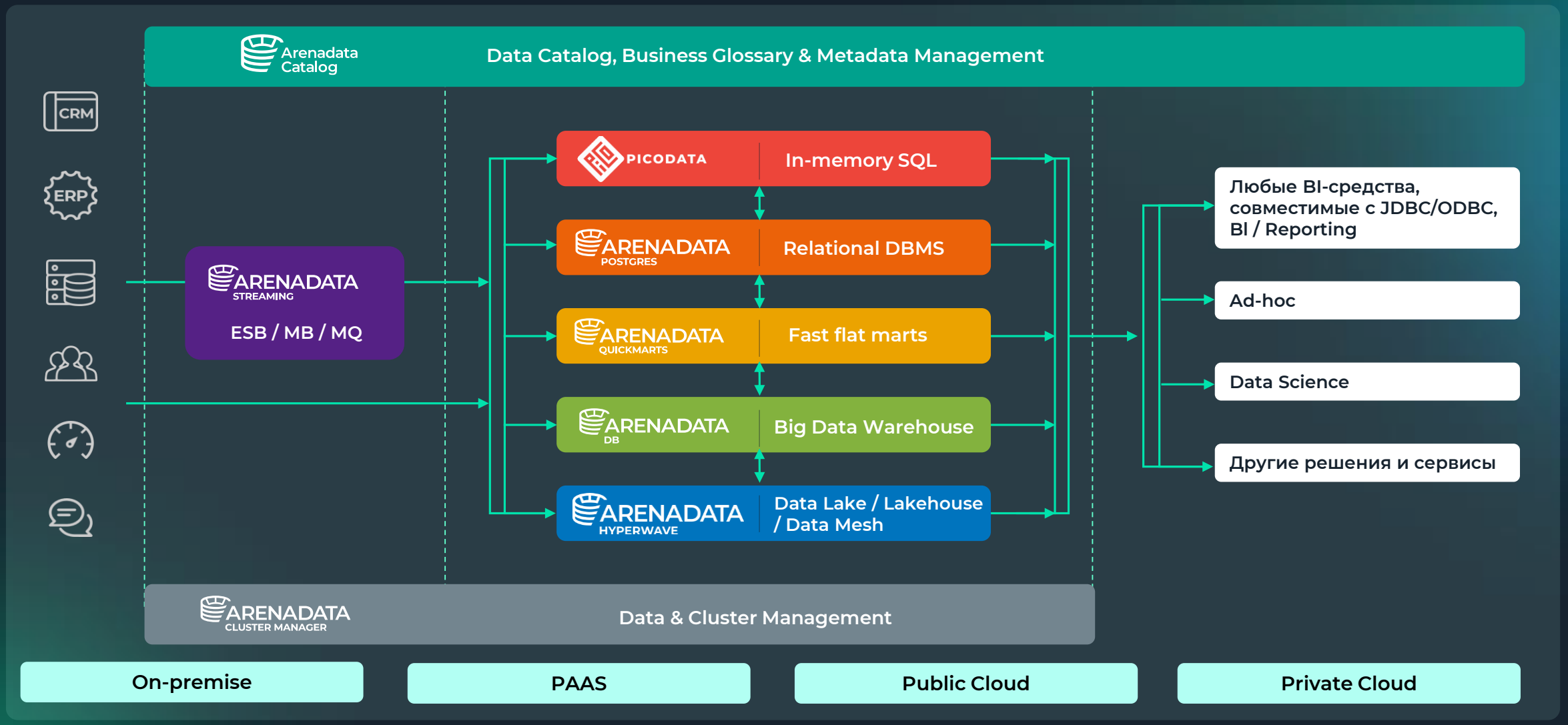
Возможности:

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

Arenadata EDP

ARENADATA

Arenadata Enterprise Data Platform



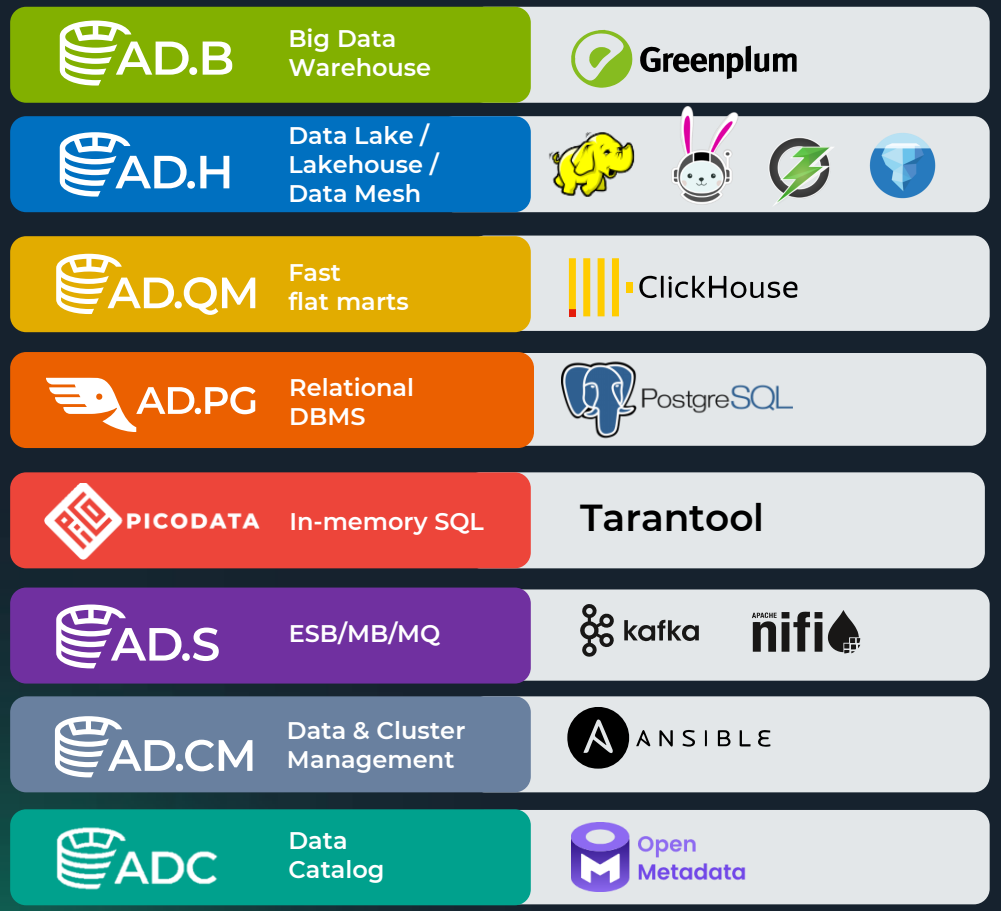
Миграция с иностранных решений

Управление данными и качество данных	Informatica, alteryx, Alation, Collibra, ATAKAMA, IBM	ADC CLEAN DATA
Базы данных в памяти *(резидентные СУБД)*	GridGain, redis, HAZELCAST	PICODATA
Аналитические СУБД в режиме реального времени *(аналитические СУБД)*	Microsoft SQL Server	AD.QM
Хранилища данных *(аналитические СУБД)*	VERTICA, SAP HANA, Pivotal Greenplum, teradata., ORACLE EXADATA	AD.B
Озёра данных, lakehouse, data mesh	CLOUDERA, ORACLE BIG DATA APPLIANCE	AD.H
Потоковая передача (загрузка) данных	kafka, nifi	AD.S
Транзакционные базы данных *(СУБД общего назначения)*	ORACLE DATABASE, IBM DB2, Microsoft SQL Server	AD.PG

Единая платформа, объединяющая востребованные продукты мирового уровня и позволяющая эффективно замещать целый комплекс решений ушедших зарубежных компаний

AD

Каждый добавляемый компонент платформы доработан, чтобы стать её частью



Обязательный перечень задач для каждого компонента:

- Выполнить статистический анализ кода
- Выполнить интеграцию с модулем мониторинга
- Проверить работоспособность заявленного функционала
- Подобрать персонал поддержки, подготовить и обучить архитекторов и разработчиков
- Разработать коннекторы для взаимодействия с другими компонентами платформы
- Сформировать дорожную карту доработок продукта

AD

Мы осуществляем безопасную разработку

Используем инструменты и методики:

- ✓ Статические анализаторы исходного кода
- ✓ Поиск секретов в исходном коде
- ✓ Безопасность в образах контейнеров
- ✓ Анализатор зависимых компонентов
- ✓ Средства динамического анализа

Разработка в соответствии с ГОСТ Р 56939-2016 «Разработка безопасного программного обеспечения»

Применяются международные практики OWASP Application Security Verification Standard

Разработаны и внедрены правила для каждого языка программирования

Enterprise Multi-cloud платформа данных

Виды развёртки любых компонентов платформы Arenadata EDP:

01 в облаке

02 на bare-metal-оборудовании

03 гибридная ИТ-инфраструктура или multi-clouds

Многоуровневая система технической поддержки

Варианты технической поддержки

24/7	600+	400+	99%+	50+
доступность поддержки	кластеров на поддержке	заявок в месяц	соблюдение SLA в 2024	технических экспертов

Гарантийное обслуживание	Режим работы над обращениями 8x5	Анализ проблем в рамках поступивших инцидентов	Предоставление обновлений ПО	Доступ к базе знаний по решению инцидентов
Базовая техническая поддержка	Режим работы над обращениями 24x7 для Prod	Диагностика и предоставление обходного решения	Предоставление обновлений ПО	Консультации по особенностям ПО
		Анализ проблем в рамках поступивших инцидентов	Решение инцидентов в рамках SLA	Доступ к базе знаний по решению инцидентов
Премиальная техническая поддержка	Услуги, включающие сервисы Базовой технической поддержки и учитывающие индивидуальные потребности Заказчика			

Вендорский консалтинг Arenadata

Минимизация рисков, максимизация отдачи от инвестиций, оперативное устранение проблем

Технический аккаунт-менеджмент

TAM-команда, состоящая из архитектора-эксперта, руководителя проектов, аналитиков и инженеров. Постоянная команда идёт с заказчиком рука об руку и использует свой опыт для значительного снижения проектных и технических рисков, возникающих в ходе внедрения и эксплуатации сложных систем

Технический аудит

Комплексный аудит для анализа узких мест кластера. Детальный разбор и рекомендации по изменениям от архитекторов Arenadata позволяют сформировать среднесрочную стратегию эксплуатации

Программа приоритетного внимания

Наблюдения истории развития ландшафтов данных наших клиентов позволили выстроить эталонный сервисный путь ландшафта данных для крупных заказчиков с учётом характера и интенсивности потребления услуг

Производительность

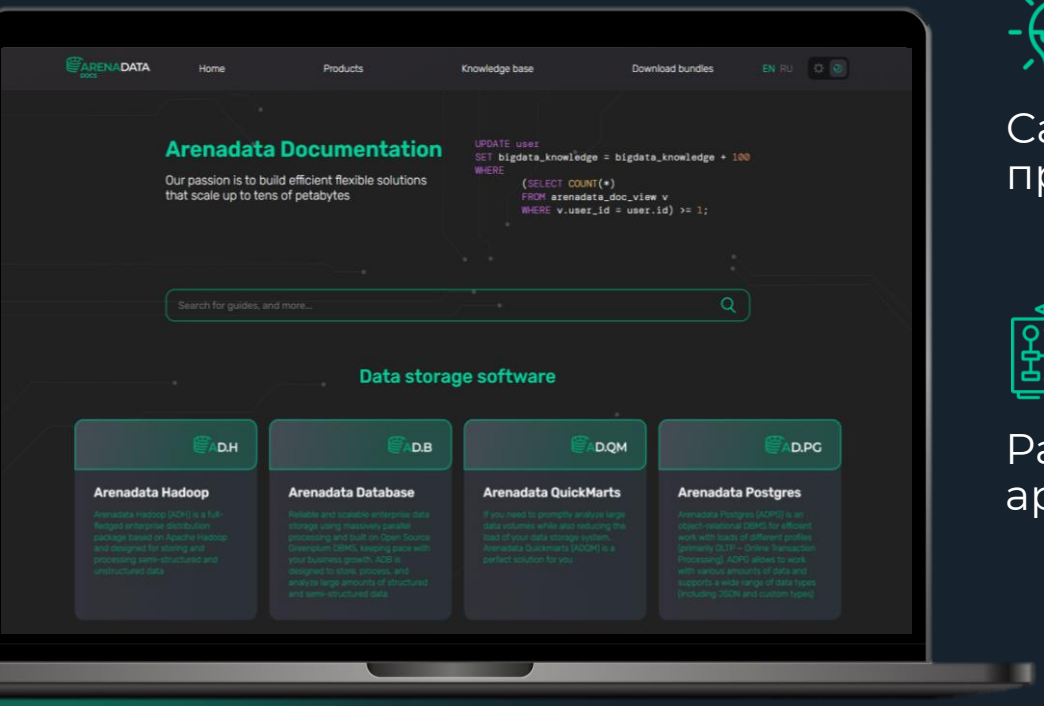
Комплекс тактических услуг по оптимизации использования ресурсов, настроек ресурсных групп, настроек резервного копирования, адаптации сайзинга к изменяющейся нагрузке и оптимизации прикладных запросов

Надёжный старт

Комплекс услуг по развёртыванию ПО Arenadata

Услуги предоставляются для продуктов ADB, ADH, ADQM, ADS, ADPG

Удобная и понятная онлайн-документация мирового уровня



Не копируем, а создаём

Обеспечиваем поддержку разных устройств

Сами разрабатываем примеры

Проводим экспертную проверку архитекторами и аналитиками

Разрабатываем дизайн архитектурных диаграмм

Не переводим, а пишем на русском и английском языках

Перейти на сайт с документацией

ЗНАНИЯ ИЗ ПЕРВЫХ РУК:

С сертификацией и присвоением статуса Arenadata Certified Specialist

Знания из первых рук

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

Узнать подробнее о курсах

Практика от экспертов

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

16 учебных программ по 8 направлениям

- Архитектура платформы данных
- Курсы по Arenadata DB
- Курсы по Arenadata Hyperwave
- Курсы по Arenadata QuickMarts
- Курсы по Arenadata Streaming
- Курсы по Arenadata Postgres
- Курсы по Picodata
- Курсы по Arenadata Catalog

ОСТАВАЙТЕСЬ НА СВЯЗИ

Наши новости
в телеграм-канале

ARENADATA