Arenadata DB

Хранилище данных - основа растущего бизнеса

ARENADATA

Содержание

- 2 слайда про Arenadata
- Arenadata Enterprise Data Platform
- Общее описание
- Ключевые преимущества вендорского продукта
- Ключевые технические преимущества Arenadata DB
- Референсные кейсы

ARENADATA

2 слайда про Arenadata

ARENADATA

Arenadata

лидер в области платформ больших данных для цифровой трансформации

Greenplum · hadoop · ClickHouse · PostgreSQL · TARANTOOL · kafka · nifi · Open Metadata

СТАБИЛЬНОСТЬ

- **6 лет** на рынке
- **270** человек
- **2 млрд**. ₽ выручка ‘22

60+ КЛИЕНТОВ

- ГПН, Росатом, ММК, Норникель, Сибур;
- X5, Магнит, Ашан;
- ВТБ, ПСБ, ДОМРФ;
- ФНС, ДИТ МСК, ФСС;

NO VENDOR LOCK

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

ONPREM & CLOUD

- Onpremise и ПАКи;
- PaaS: VK Cloud; Cloud.ru
- IaaS: MTS Cloud, ТЛ 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

- **50+** технологических партнеров и интеграторов;
- **2.000+** обученных специалистов и экспертов.

COMMUNITY

- **#1** в мире в сообществе контрибьютеров в Greenplum;
- Профессиональная документация.

РЕГУЛЯТОРЫ

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

© Arenadata 2023

Наш успех – доверие клиентов

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

Retail & FMCG

Банки, страховые и телеком

Госсектор

Пром, Энерго, Добыча

ФНС

ДИТ МОСКВЫ

Национальная система управления данными

Социальный фонд

Счетная палата

Казначейство

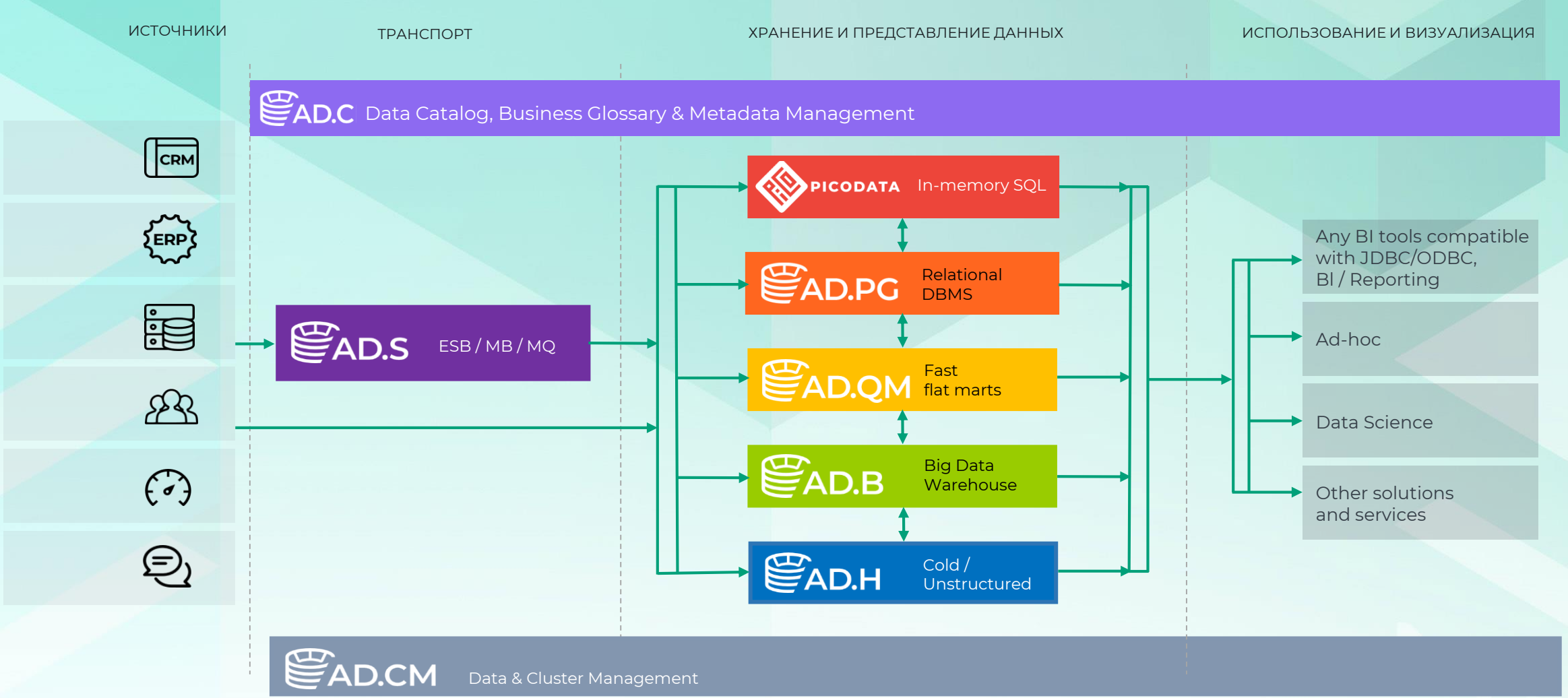
Росреестр

Роструд

Arenadata EDP

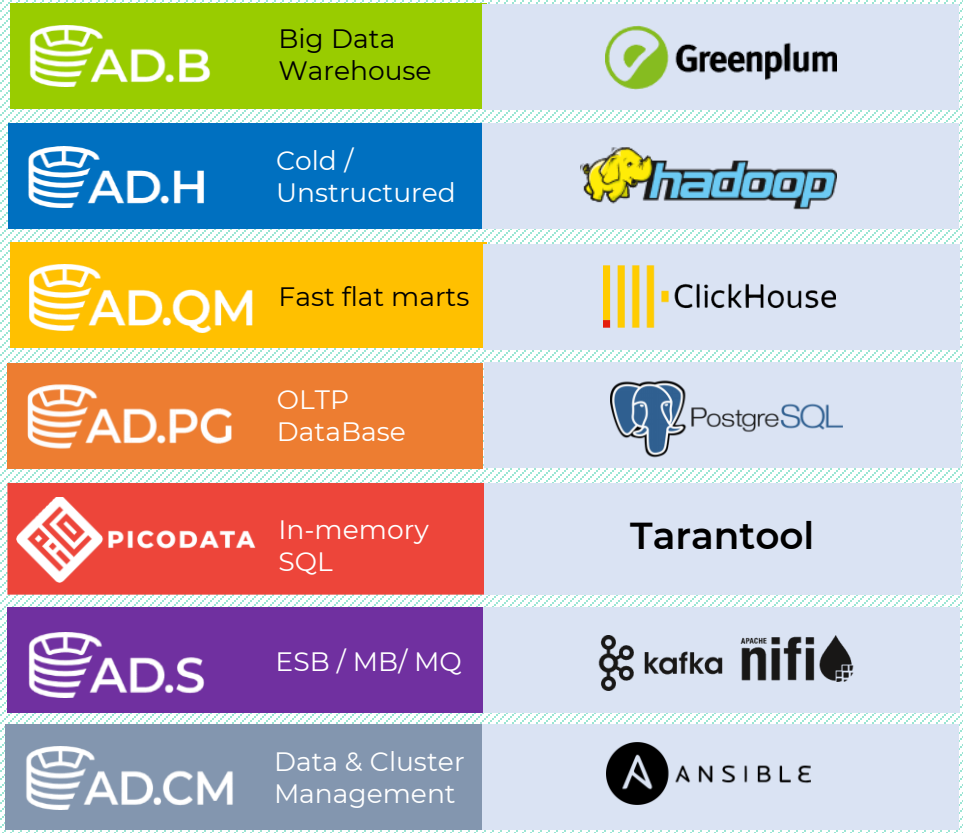
ARENADATA

Arenadata Enterprise Data Platform



© Arenadata 2023

Что внутри?



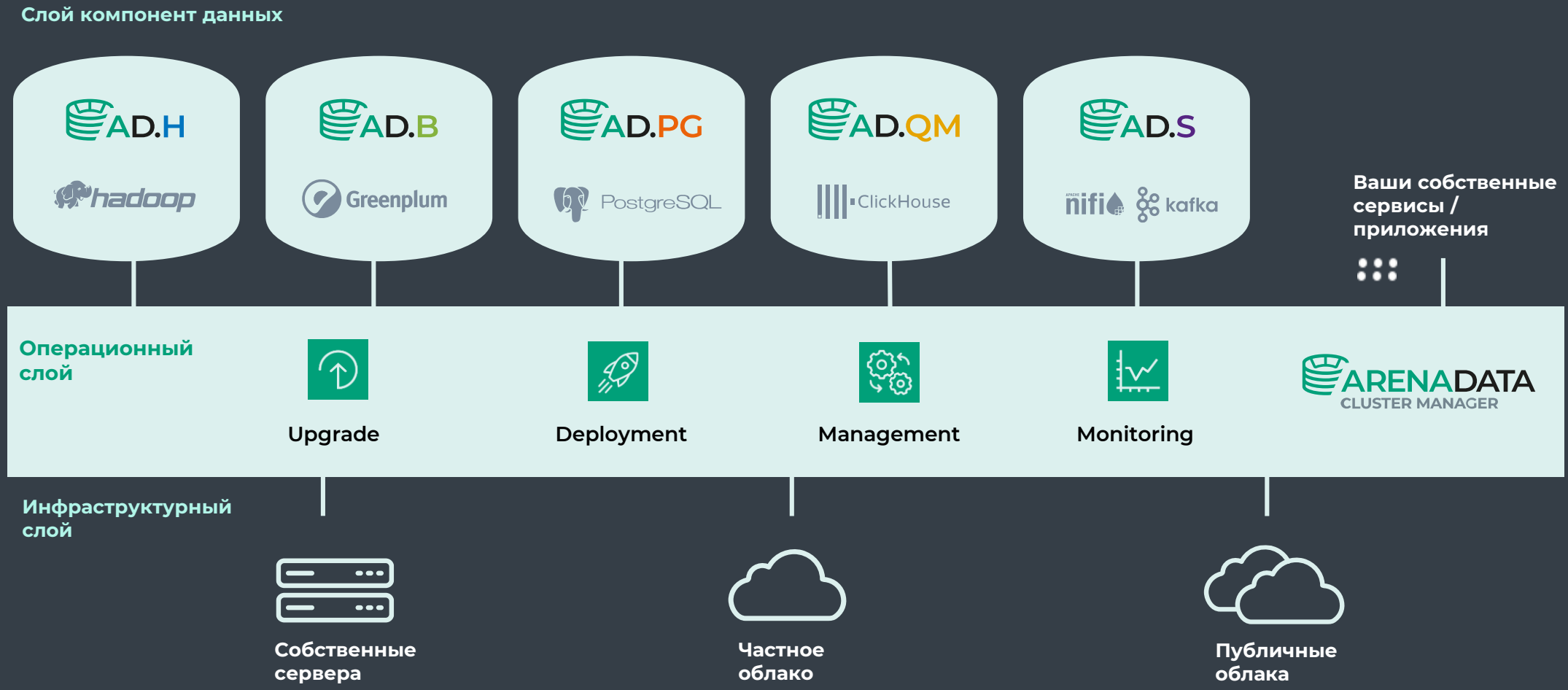
Каждый добавляемый компонент платформы доработан, чтобы стать её частью

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

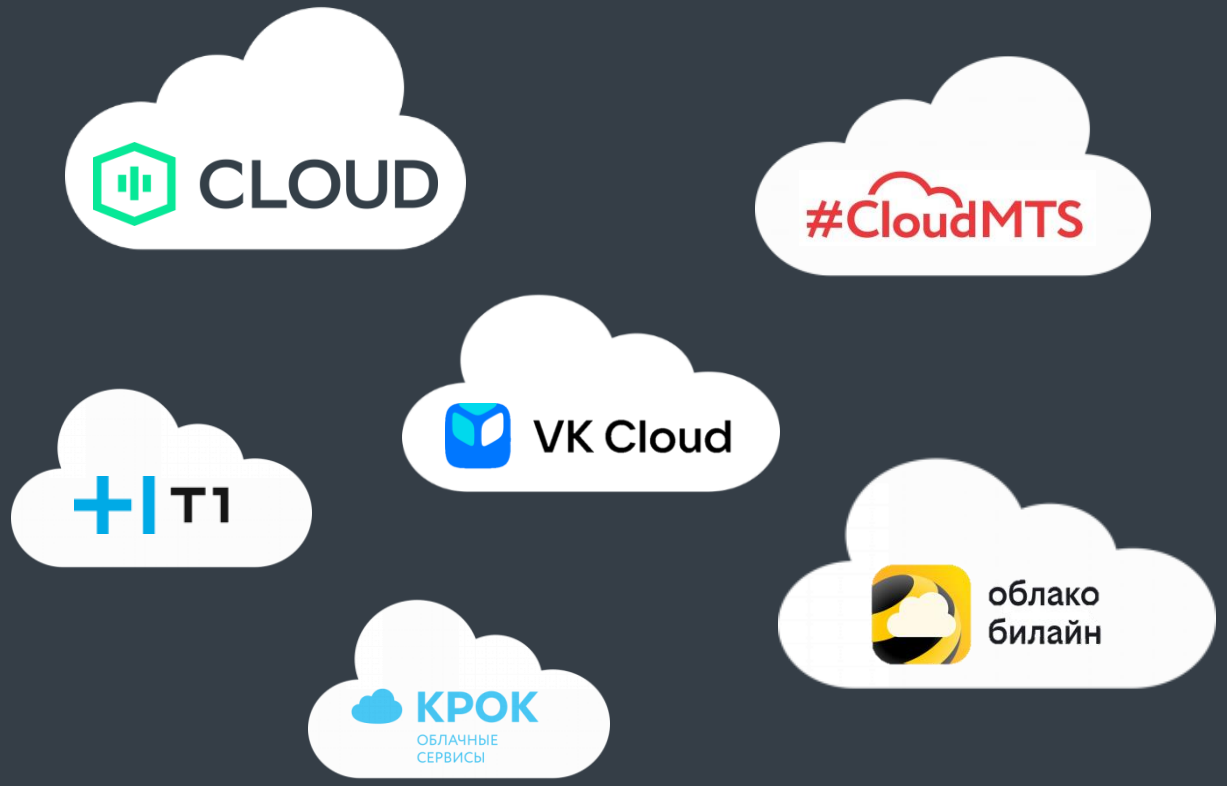
© Arenadata 2023

Гибридный корпоративный ландшафт



© Arenadata 2023

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.

©Arendata 2023

Общее описание

ARENADATA

Почему Greenplum?

Может работать с данными из нескольких источников с минимальной предобработкой

Изолирует процессы разных потребителей друг от друга, гибко разграничивая ресурсы кластера между пользователями

Позволяет анализировать данные с последующим прогнозированием

Поддерживает реляционную модель данных, легко интегрируется с PostgreSQL и другими реляционными решениями

Быстро обрабатывает сложные аналитические запросы благодаря MPP-архитектуре (Massively Parallel Processing)

Предоставляет федеративный доступ к данным в смежных системах через параллельный интерфейс

© Arenadata 2023

Greenplum

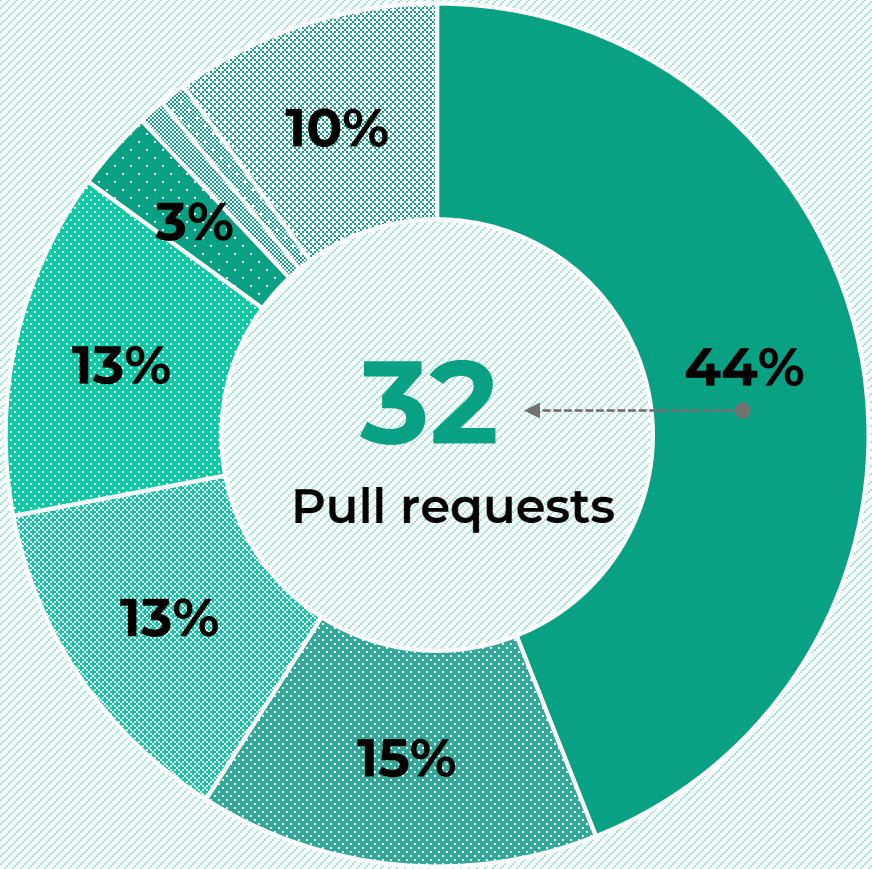
- Greenplum - это массивно-параллельная СУБД с открытым исходным кодом, основанная на PostgreSQL
- Greenplum предоставляет вам больший контроль над развертываемым ПО без vendor lock-in и позволяет открыто влиять на развитие продукта
- Крупнейшие компании по всему миру вносят свой вклад в Greenplum

Alibaba.com
ARENADATA
Avito
LEROY MERLIN
Morgan Stanley
X5 Group
TINKOFF
SBERBANK
vmware
Yandex

©Arenadata 2023

Вклад в развитие Greenplum

1-е заняла Arenadata по вкладу в развитие Greenplum среди международного сообщества*



- Arenadata
- Alibaba
- **Китай (компания неизвестна)**
- **Яндекс**
- Beijing University of Posts and Telecommunications
- Neondatabase
- SeeQR
- **Прочие**

ARENADATA

*Исследование Arenadata, 2023

Отличия ADB от Greenplum

Функционал	Open Source	ADB EE
• Core-функционал • PXF • gpbackup • Коннекторы Greenplum <-> Hadoop и Greenplum <-> JDBC-источники	+	+
• Коннекторы Greenplum <-> Kafka и Greenplum -> ClickHouse	-	+
• ADB Contol (мониторинг на уровне запросов)	-	+
• Офлайн установка	-	+
• Управление деплоем и апгрейдом, • Расширение кластера, • Мониторинг & alerting	-	+
• Инструментарий управление бэкапами (ADBM) • Бинарные инкрементальные бэкапы (WAL binary Backup / Restore)	-	+
• Документация	English only	+
• Поддержка • Обучение по продуктам	-	+
• Операционная система	Ubuntu 18.04 / Redhat 7 / Redhat 6	CentOS 7 / Redhat 7 / Альт 8 СП / Astra Linux 1.7 SE "Орел"
• Доп.консалтинговые услуги (DBAaS, Smart Start, TAM, Аудит)	-	+

© Arenadata 2023

ARENADATA DB

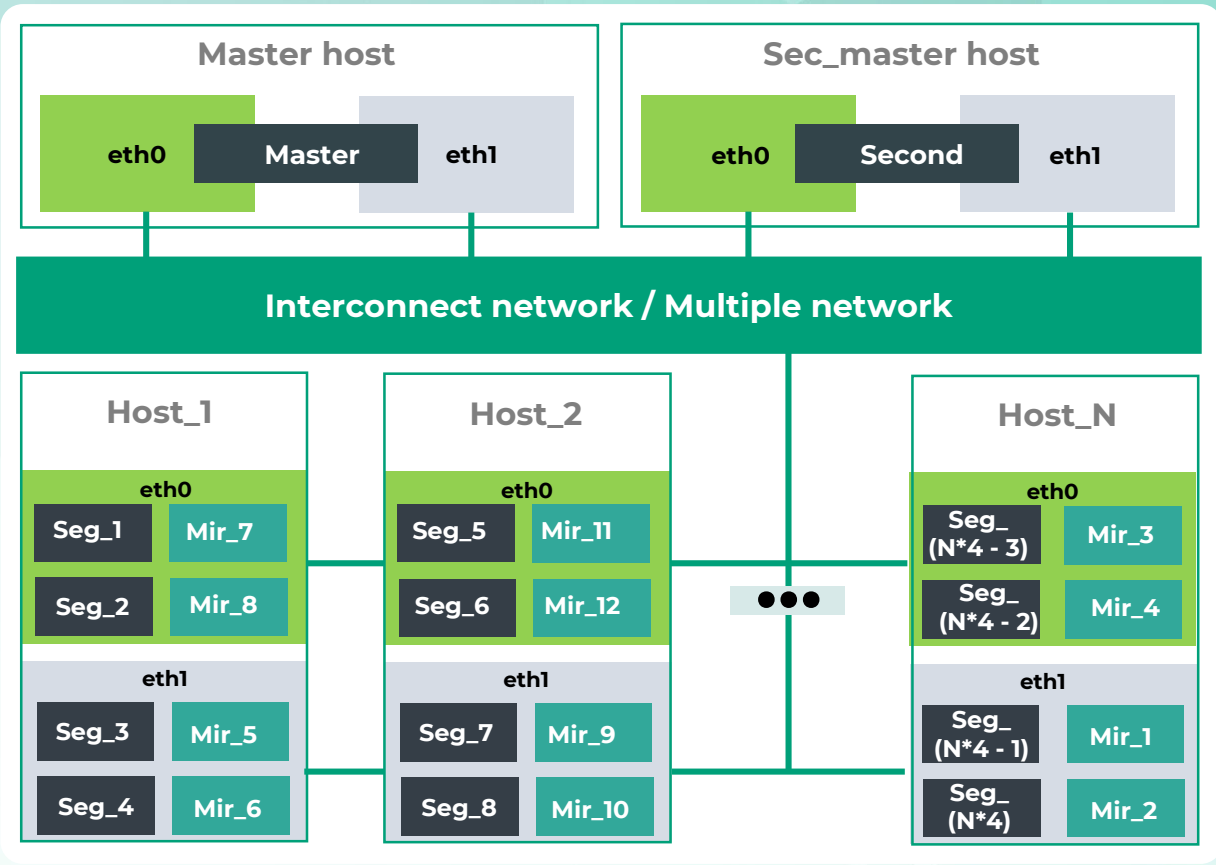
powered by Greenplum

Arenadata DB (ADB) — массивно-параллельная реляционная аналитическая СУБД, построенная на основе технологии с открытым исходным кодом Greenplum.

Применяется в критически важных бизнес-системах и предназначена для хранения и обработки больших объемов данных – от десятков терабайт до десятков петабайт.

Arenadata DB лучше всего справляется с:

- Реализацией аналитических хранилищ размеров в сотни ТБ
- Сложными запросами, обрабатывающими большие объёмы, в том числе содержащими сложные аналитические функции
- ETL/ELT
- Эффективным соединением больших таблиц
- Data Science
- Аналитическими функциями на процедурных языках
- Ad-hoc аналитикой



© Arenadata 2023

Распределённая СУБД для больших данных

Arenadata DB использует кластере из множества серверов и предназначена для хранения и обработки больших объёмов данных — до десятков петабайт.

По мере роста объёма данных вы просто добавляете новые сервера в кластер — обо всём остальном ADB позаботится сама.

Высочайшая скорость обработки, возможность горизонтального масштабирования, отсутствие привязки к аппаратному обеспечению, отказоустойчивость – всё это сделало Arenadata DB оптимальной СУБД для решения целого ряда аналитических задач:

- клиентская аналитика
- формирование отчётности
- анализ оттока клиентов
- предиктивный анализ
- формирование маркетинговых кампаний

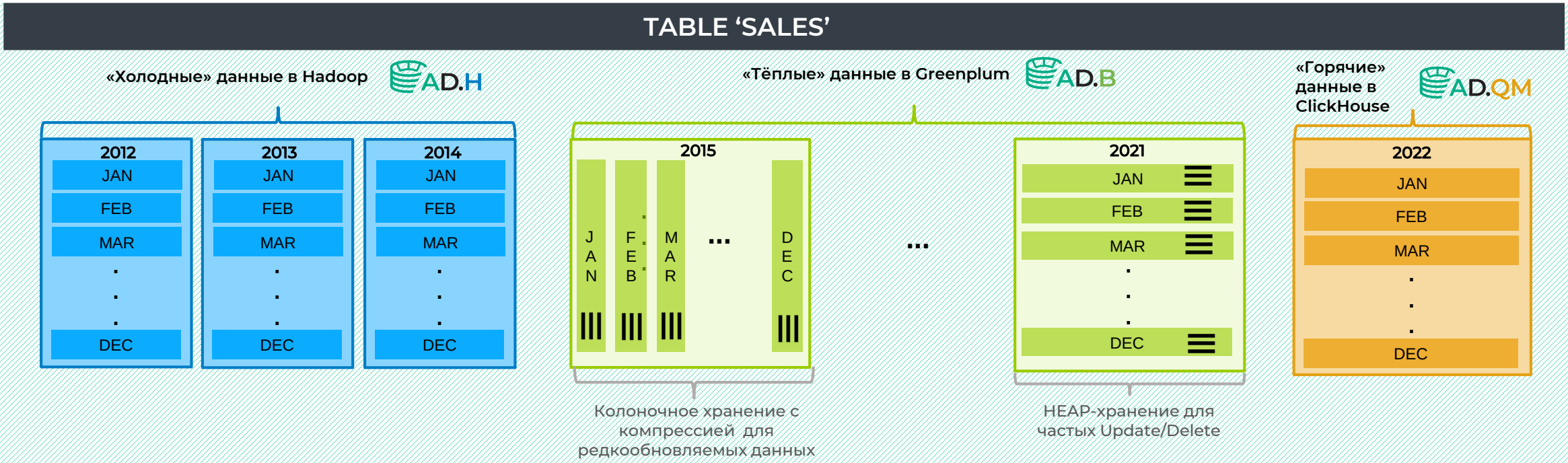
AD.B

© Arenadata 2023

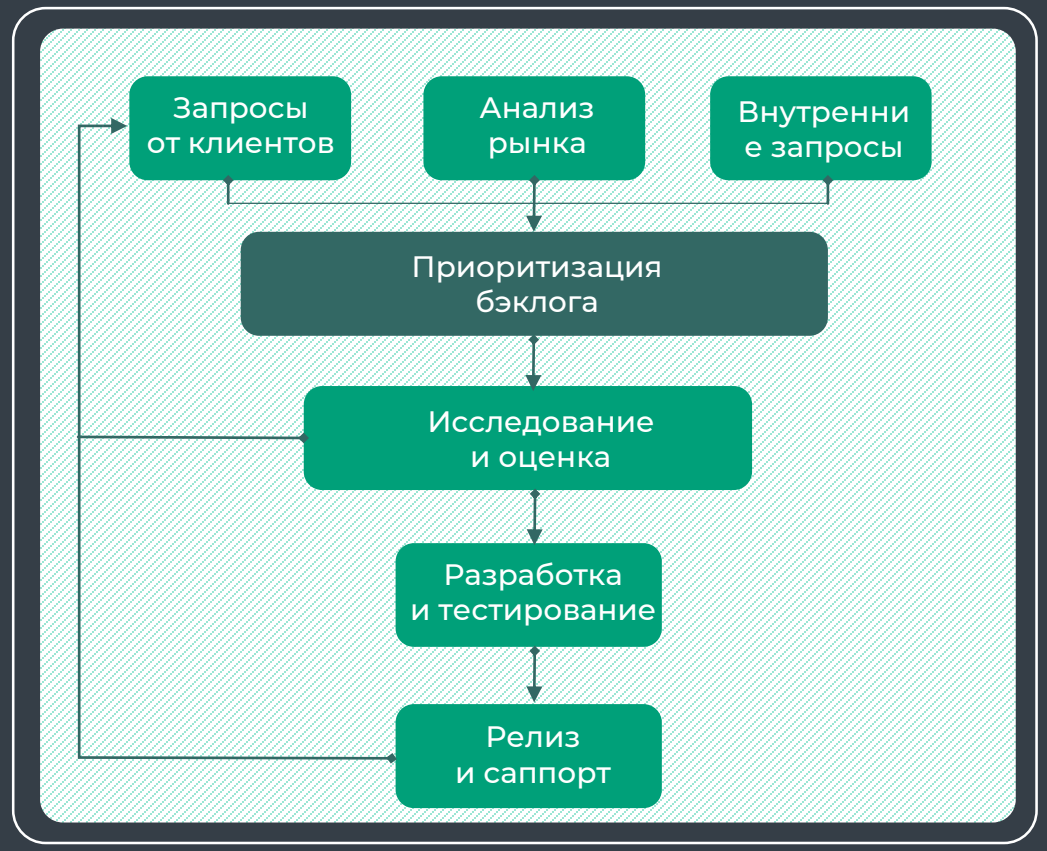
Полиморфное хранение

- Данные, которые устарели и к которым редко обращаются (холодные данные), обычно архивируются и помещаются в долгосрочные хранилища, такие как озеро данных на базе Arenadata Hadoop или хранилище объектов S3.
- Данные, которые являются свежими и подвержены частому доступу (горячие данные), хранятся в операционных базах данных, таких как Arenadata QuickMarts или Arenadata Postgres.
- А данные, требующие небольшой оперативной и большей аналитической нагрузки, сохраняются в базе данных Arenadata DB (теплые данные).

Возможности полиморфного хранения Arenadata DB при использование PXF делают запросы к данным в разных температурных слоях простыми и эффективными.



Процесс развития продуктов Arenadata



За 2021-2022 года

390+
запросов
от клиентов

230+
реализовано

© Arenadata 2023

Ключевые преимущества вендорского продукта

ARENADATA

Российское программное обеспечение

Регистрация в Реестре Минцифры РФ

Отсутствие валютных рисков - цены в рублях

Сертификат ФСТЭК

Центры разработки в РФ: Москва, Санкт-Петербург, Ростов-на-Дону, Хабаровск

Поддержка, документация и учебные курсы на русском языке

Минимизируйте санкционные и валютные риски

Arenadata DB высоко зарекомендовала себя на российском рынке. Она эффективно замещают аналитические системы таких иностранных производителей, как Oracle Exadata, Teradata, Exasol, Vertica и других.

Наши программные решения стали частью проектов миграции с зарубежного ПО в десятках крупнейших организаций, включая ФНС РФ, Счётную палату России, ПАО «Газпром нефть», банк ВТБ, Ак Барс Банк, X5 Retail Group и «Халык Банк» (Казахстан).

© Arenadata 2023

Преимущества вендорского решения

Площадь «зоны» соответствует дополнительным затратам на имплементацию

© Arenadata 2023

Вендорский консалтинг

Технический аккаунт-менеджмент
Обеспечиваем постоянный доступ заказчика к вендорской экспертизе и получению необходимых ресурсов архитекторов, инженеров и аналитиков Arenadata через включение наших экспертов в команду заказчика.

Технический аудит
Проводим всесторонний анализ кластеров Arenadata DB, выдаём рекомендации по устранению узких мест, оптимизации распределения ресурсов, анализируем наиболее тяжёлые запросы.

Smart Start
Установка и настройка компонентов платформы Arenadata EDP, интеграция продуктов в инфраструктуру заказчика, инструктаж администраторов и другие работы, необходимые быстрого старта.

ARENADATA

Профессиональные сервисы Arenadata

Администрирование ADB по подписке

Вторая линия поддержки, обеспечивающая работоспособность кластера с возможностью выбора пакета услуг под требования заказчика. Решает следующие задачи:

- Мониторинг;
- Контроль состояния кластера;
- Восстановление после аварий;
- Контроль Backup;
- Управление пользователями.

Техподдержка

Служба поддержки клиентов в режиме24/7 для высоконагруженных критичных систем и 8/5 для менее критичных систем согласно строгим SLA:

- Оперативная диагностика, локализация проблем;
- Устранение сбоев;
- Консультации по использованию функционала;
- Предоставление обновлений и помощь в установке;

Расширенная техподдержка

Расширение службы техподдержки силами экспертов вендора для решения задач в режиме 8/5 или 24/7:

- Обновление минорных версий;
- Обновление мажорных версий;
- Расширение кластера;
- Сбор диагностической информации и применение рекомендаций вендора;
- Для ADB настройка подключения к внешним источникам
- Сопровождение Заказчика в рамках задач базовой техподдержки

© Arenadata 2023

Сервис и поддержка

1 Три линии поддержки, в режиме 24x7 через персонализированную систему. Сотрудники распределены по часовым поясам

2 Одна из лучших в мире экспертиз по ядру Greenplum (2й коммитер по объему кода)

3 Расширение и доработка функционала ядра

24/7
Обрабатываем заявки

2
Языка службы поддержки (RU, EN)

~2300
Успешно разрешенных обращений за 2022 год

95,7%
Средний показатель SLA

190+
Инсталляций на поддержке в год

4 Уникальный опыт и экспертиза команды. Мы поддерживаем десятки клиентов и сталкивались с большинством возможных проблем

5 Проактивная диагностика - собственная система мониторинга и эскалация при сбоях

© Arenadata 2023

Документация Arenadata

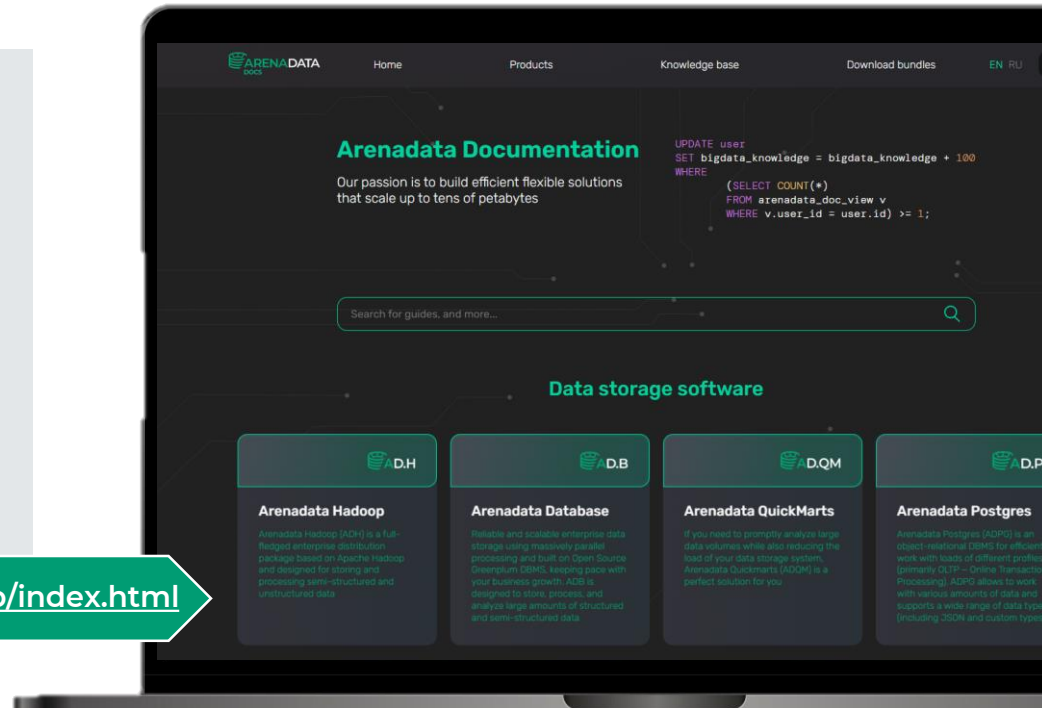
Не копируем, а создаем

Сами разрабатываем примеры

Разрабатываем дизайн архитектурных диаграмм

Не переводим, а пишем на русском и английском языках

- Проверка документации на соответствие функциональности на тестовом стенде
- Проверяем лексику, грамматику
- Проверяем соответствие руководству по стилю
- Проверяет архитектор и системный аналитик
- Качество и удобство
- Поддержка разных устройств: мобильная версия



Перейти на сайт с документацией https://docs.arenadata.io/ru/landing-adb/index.html

Курсы по Arenadata DB

Знания из первых рук

60% практики

Сертификация

Для аналитиков

поможет научиться использовать СУБД Arenadata DB (ADB) для задач обмена и анализа данных

Для разработчиков

поможет разобраться во всех тонкостях Arenadata DB — аналитической, распределённой СУБД, построенной на MPP-системе с открытым исходным кодом Greenplum

Эксплуатация ADB

поможет изучить особенности работы с распределённой СУБД Arenadata DB (ADB), основанной на MPP-системе с открытым исходным кодом Greenplum

Обеспечение требований безопасности

Ролевая модель RBAC сообразно PostgreSQL

Базовые возможности

- Группы и роли (Role Based Access Control, RBAC)
- Права на объекты
- Наследование
- Авторизации по LDAP
- Шифрование соединений и всего внутреннего взаимодействия компонентов СУБД с помощью SSL
- Шифрование данных на дисках кластера с помощью PGP (на уровне таблиц или колонок в таблицах)
- COLUMN-level security

Дополнительно

- Сертификация ФСТЭК по 6УД
- Работа над сертификацией ФСТЭК по 4УД
- Маскирование пароля в PXF
- Поддержка SSL в ADB Kafka коннекторе

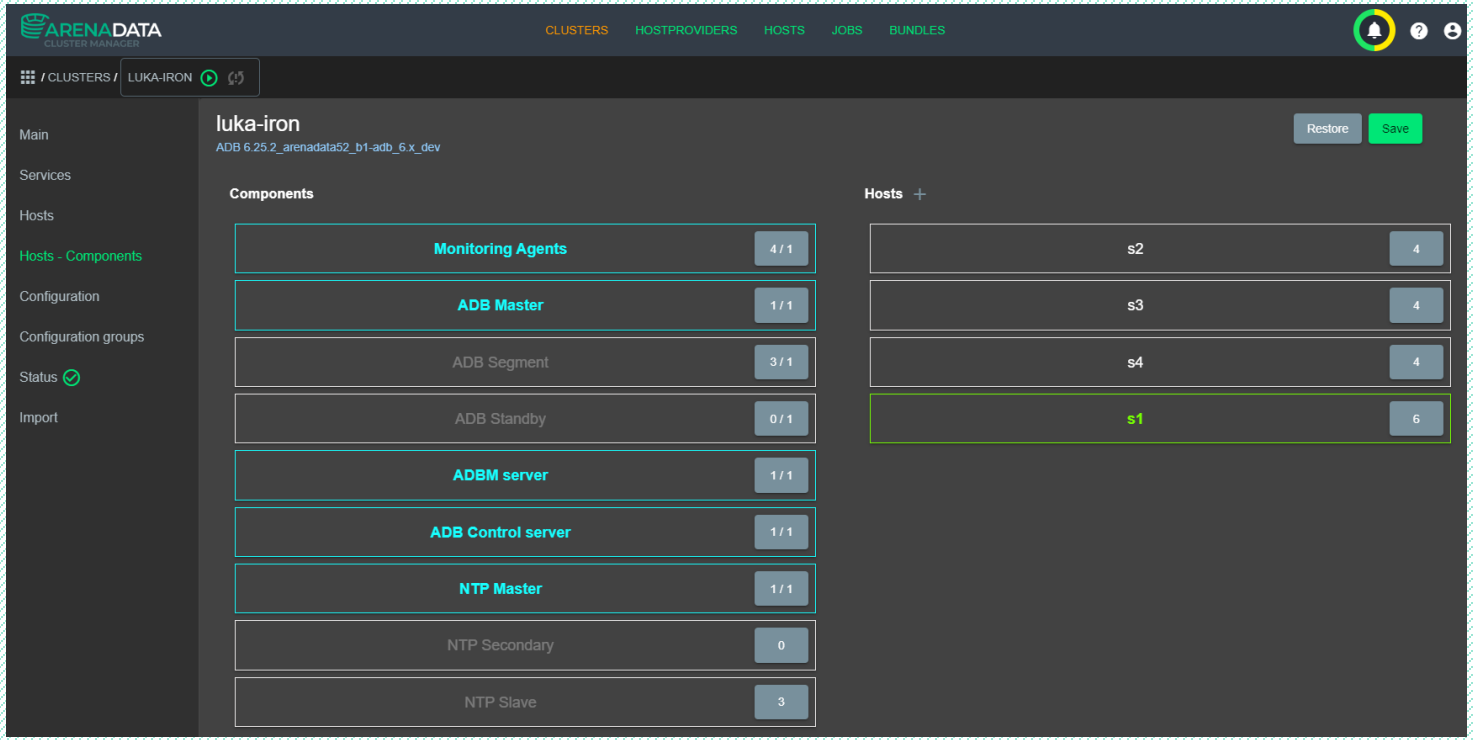
© Arenadata 2023

Ключевые технические преимущества Arenadata DB

ARENADATA

Отличительные особенности ADB

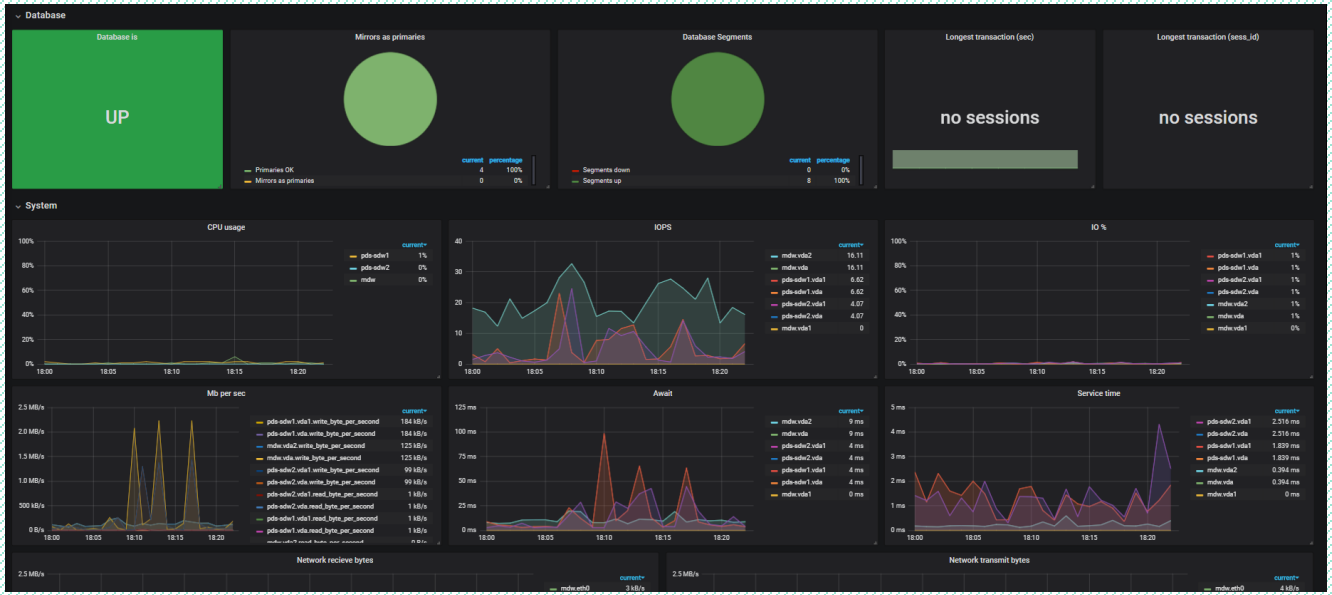
- Простота установки и настройки из Arenadata Cluster Manager
- Нативная интеграция с Arenadata Hadoop, QuickMarts, Streaming в рамках единой платформы Arenadata EDP
- Продвинутая консоль управления поможет выполнять текущие задачи быстрее и более предсказуемо, а значит, безопасней
- Мониторинг на основе Graphite и Grafana
- Установка в закрытых ЦОДах без доступа в интернет
- Оперативная доработка и развитие функционала под требование заказчика
- Наличие поддержки, документации и учебных курсов на русском языке



© Arenadata 2023

Удобный мониторинг

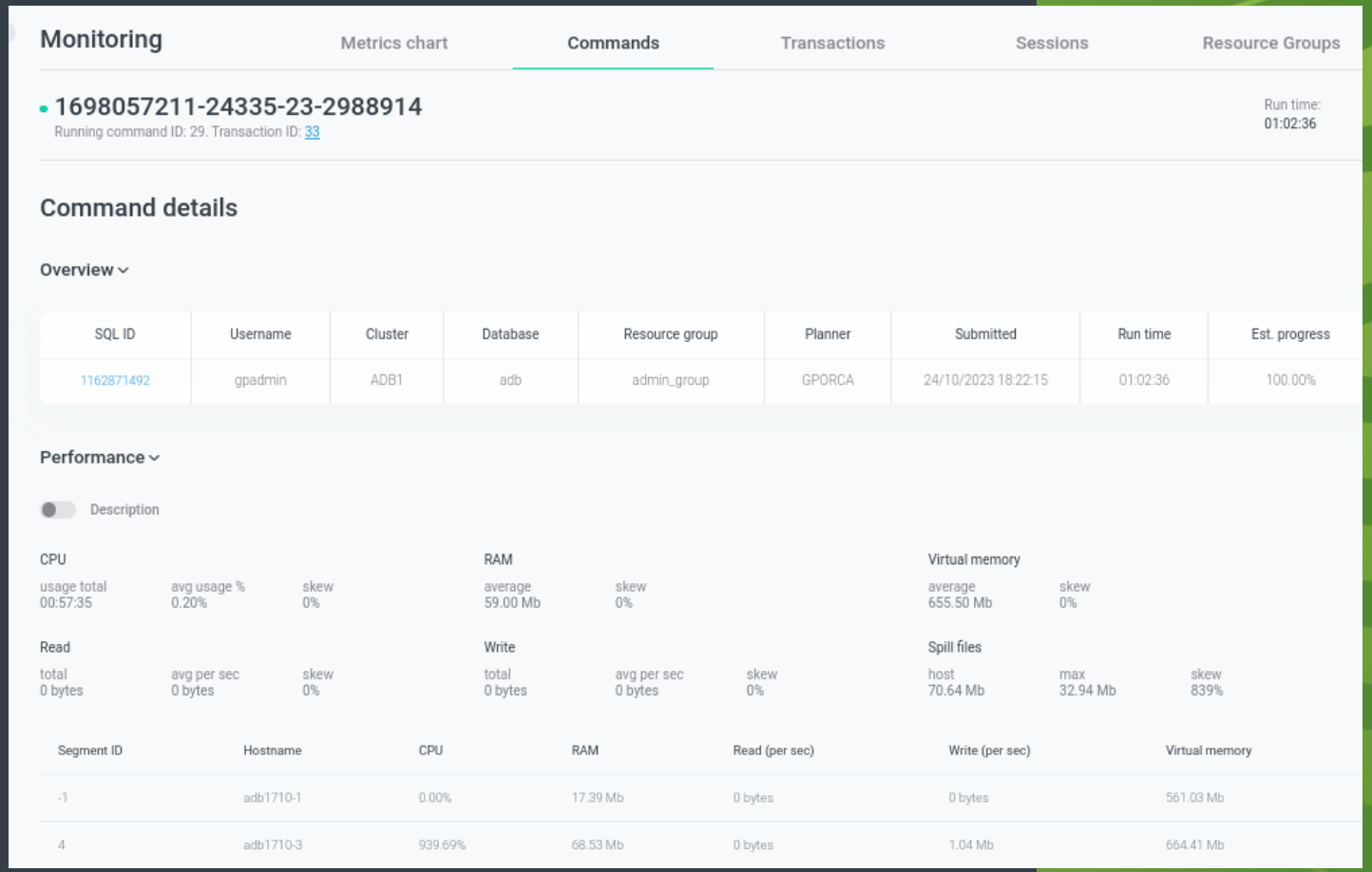
- Инструмент администратора
- Отвечает на вопрос: «что происходит с кластером?»
- Детализация на уровне сегмент-сервера
- Open-source (Grafana + Graphite)



Мониторинг входит в комплект Community & Enterprise Edition

Аналитика запросов с ADB Control

- Отслеживание статуса и прогресса выполнения запроса
- Возможность отмены запроса
- Поиск тяжелых запросов по CPU/RAM/Spill
- Отслеживание в запросах перекосов (skew) по сегментам в части количества данных/CPU/RAM/Spill/Read/Write
- Возможность настройки лимитов ресурсных групп и их мониторинга
- Визуализация изменения различных метрик в подключенных кластерах ADB

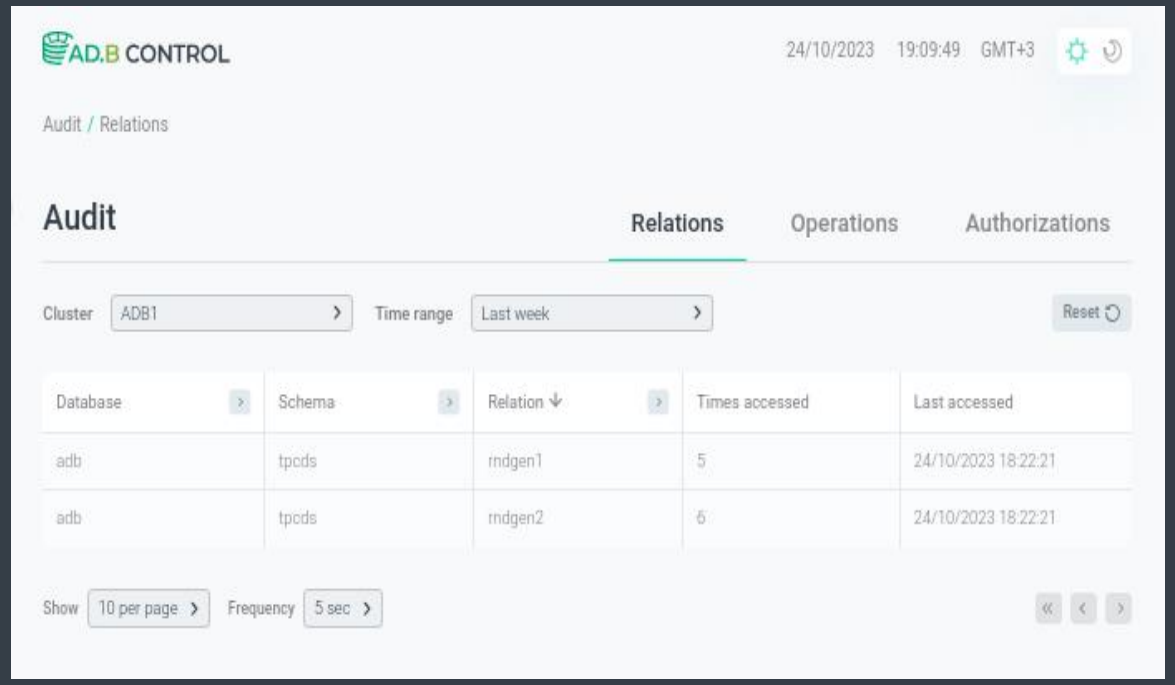


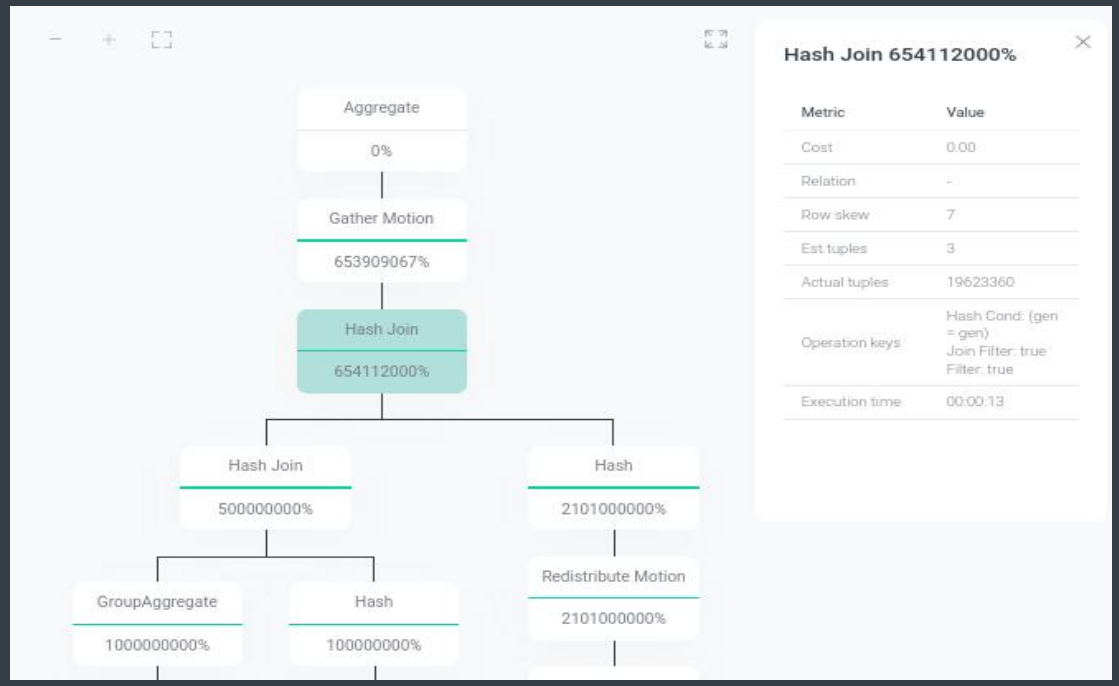
© Arenadata 2023

ADB Control (ADBC)

Аудит доступа к чувствительным данным

Визуализация этапов (шагов) и прогресс выполнения



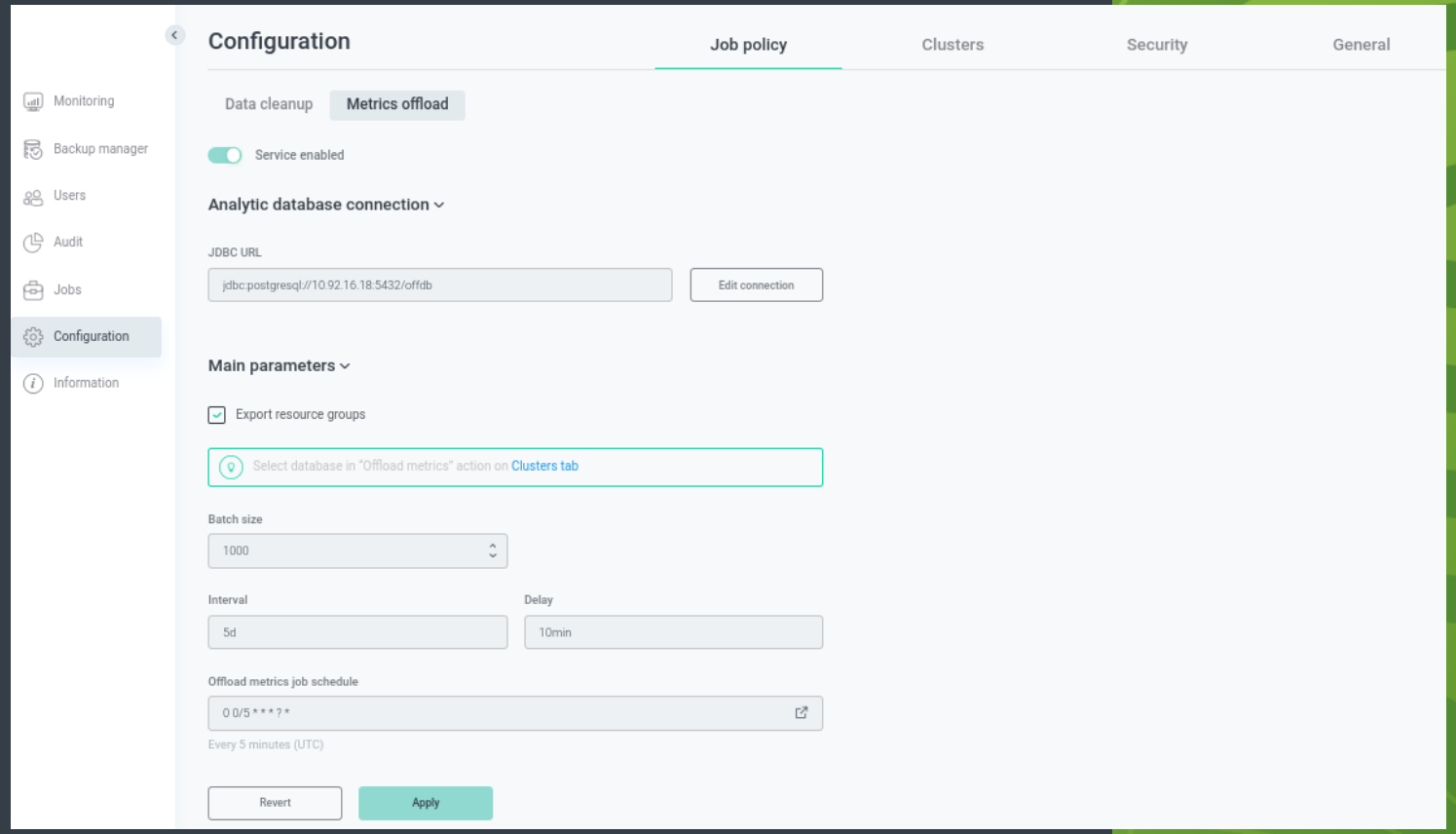


© Arenadata 2023

ADB Control (ADBC)

Гибкие настройки

- Настройка выгрузки данных мониторинга во внешнюю аналитическую БД
- Управление политиками безопасности
- Настройка вариантов авторизации и ролевых доступов пользователей
- Управление кластерами
- Гибкая настройка внутренних политик очистки данных
- Возможность настройки удобного для пользователя часового пояса
- Полная информация о состоянии приложения



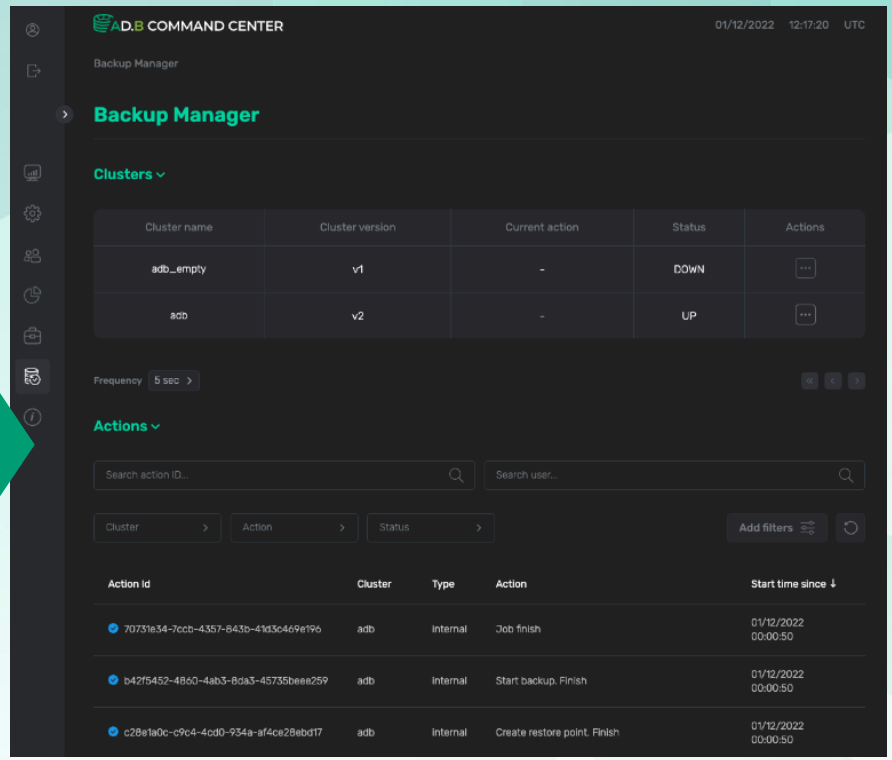
© Arenadata 2023

Arenadata DB Backup Manager (ADBM)

Arenadata DB Backup Manager - это инструмент для неблокирующего физического резервного копирования и восстановления кластеров Arenadata DB.

- **“неблокирующее РК”** – для снятия или применения резервной копии не требуется даунтайм системы
- **“физическое РК”** – копирование данных на уровне файлов с данными, с использованием write-ahead логов между бэкапами

Пользовательский интерфейс реализован в Arenadata DB Control 4.x



- Собственная разработка Arenadata
- Поставляется в составе Arenadata DB Enterprise Edition, не лицензируется отдельно
- Развертывается в виде отдельного сервиса Arenadata DB EE

© Arenadata 2023

Сравнение технологий бэкапирования

Логический бэкап

- Параллельный
- С настраиваемым сжатием
- CPU-bound
- Использует команды SQL, транзакции и блокировки
- Блокирует DDL
- PITR только в точки выполнения бэкапов (редко)
- Грануляция бэкапа / восстановления до таблицы
- Удобен для переноса данных в кластеры иной топологии или версии ADB

Физический бэкап

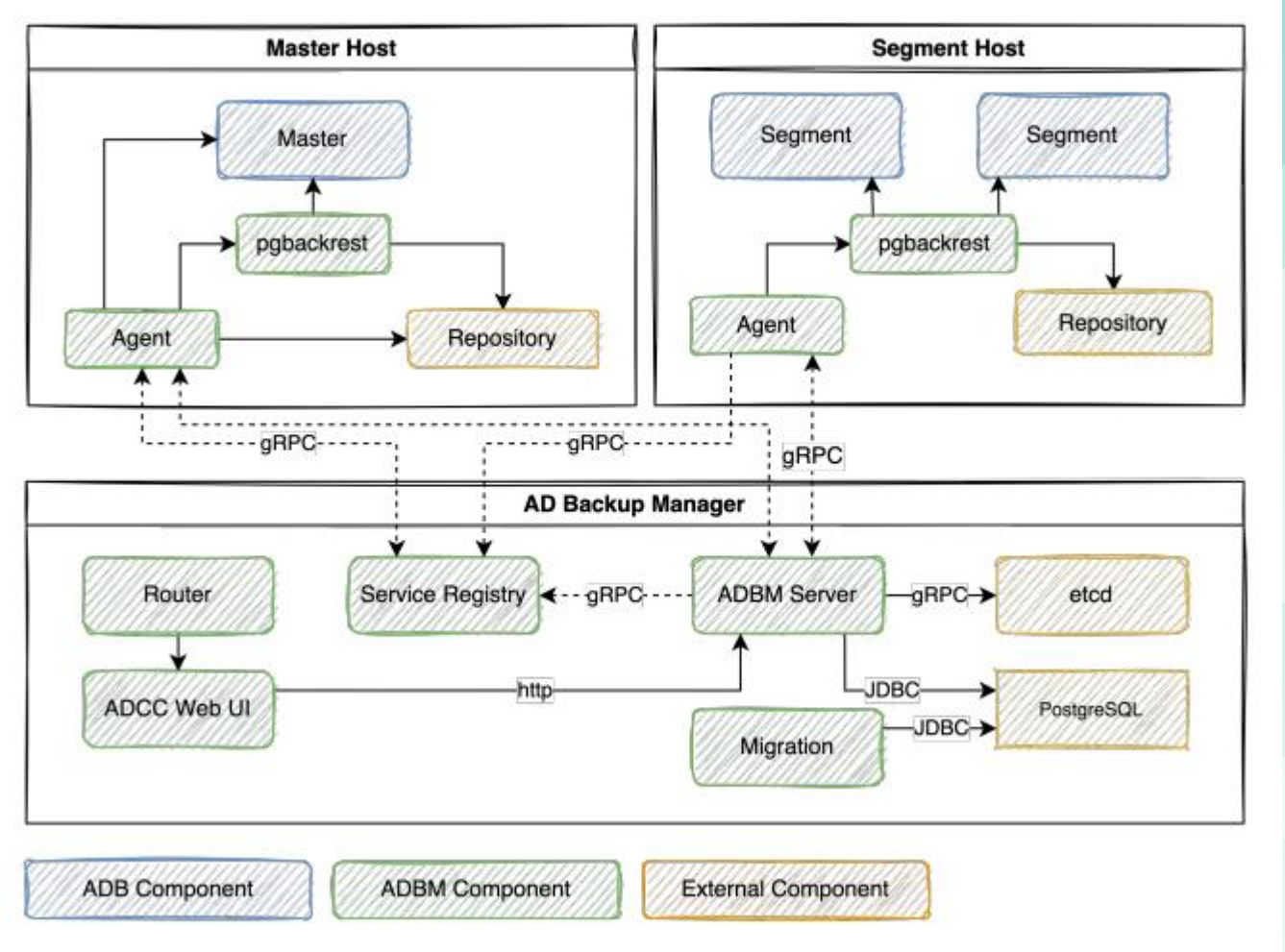
- Параллельный
- С настраиваемым сжатием
- IO-bound
- Использует файлы и непрерывную архивацию Write Ahead журналов
- Не блокирует DDL
- PITR в именованные точки восстановления (часто)
- Грануляция бэкапа и восстановления - до кластера ADB
- Может быть применён для переноса данных только в кластеры топологии и версии ADB, аналогичной исходной

© Arenadata 2023

Архитектура ADBM

ADBM включает в себя следующие компоненты:

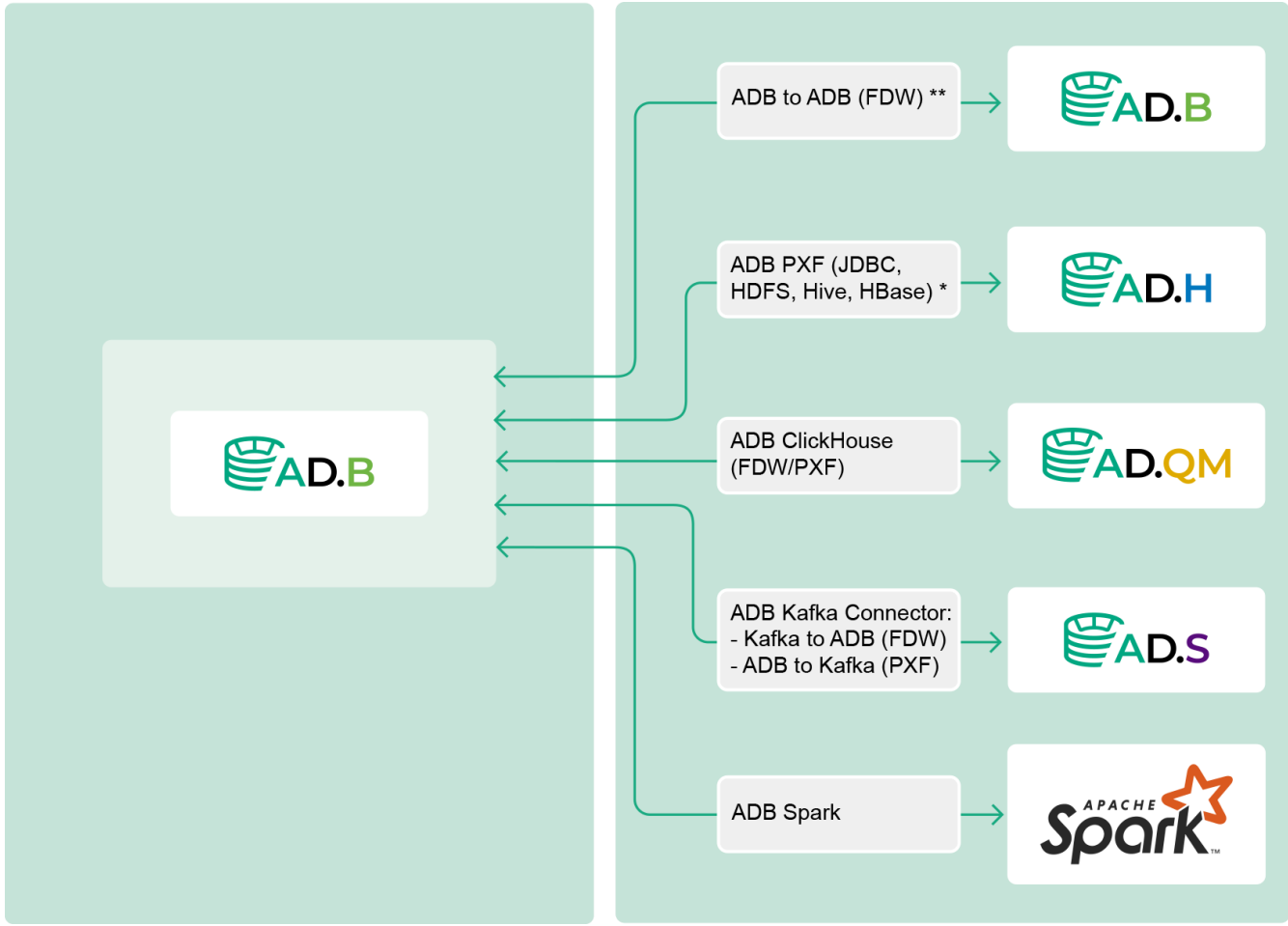
- ADBM Server
- Backup Agents
- etcd (action metadata storage)
- Service Registry
- PostgreSQL
- Backup Manager UI



Коннекторы

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

- Arenadata DB
- Arenadata Hadoop
- Arenadata Streaming
- Arenadata QuickMarts



*** ADB PXF Hive и HBase коннектор работают только на чтение**

**** ADB to ADB (FDW) на вставку в удаленный кластер работает через мастер**

© Arenadata 2023

ADB Spark connector

Обмен данными между Arenadata DB и Apache Spark

Интеграционное решение обеспечивает распределённый двунаправленный перенос данных между Data Lake на базе Hadoop и Spark (spark2 и spark3) и data warehouse на базе Arenadata DB

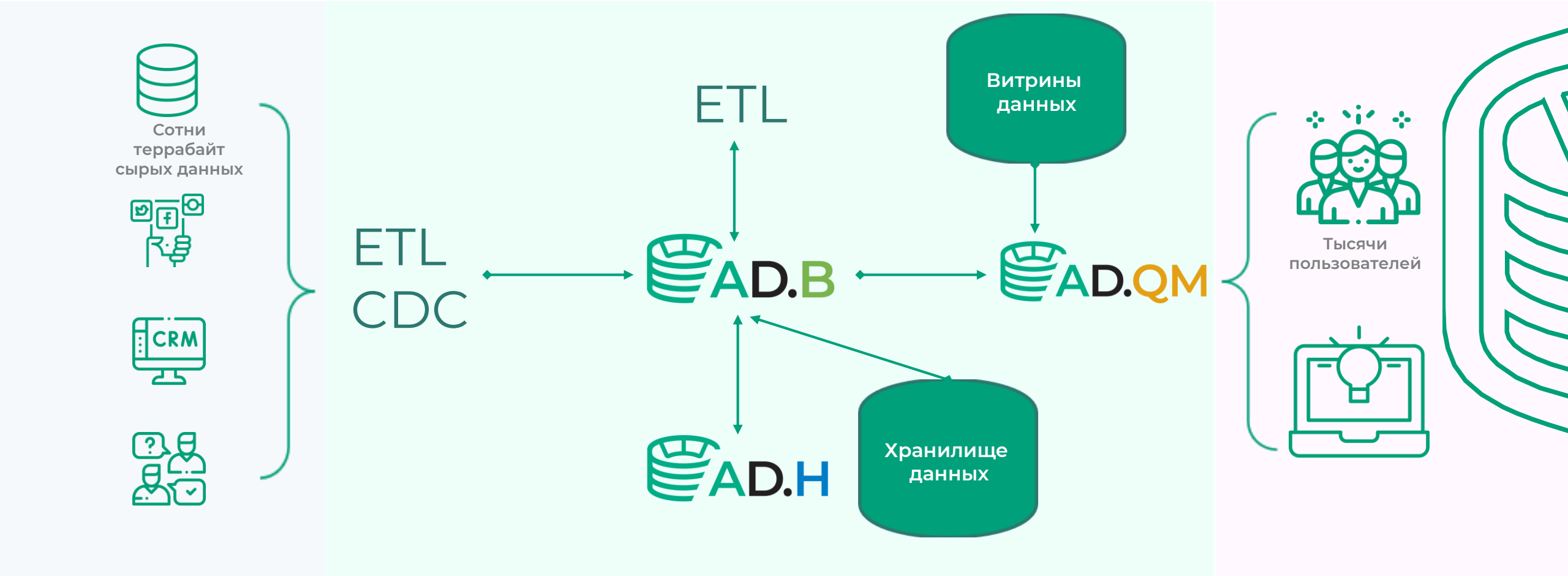
Поставляется в составе Arenadata Hadoop начиная с версии ADH 3.1.2.

Возможности:

- Перенос данных из ADB в Spark любым из пяти способов партиционирования
- Запись данных из Spark в ADB в режимах Append, Overwrite или ErrorIfExists
- Автоматическое формирование схем данных
- Поддержка отсекания колонок и push-down фильтров
- Извлечение статистики и ключа распределения данных из ADB
- Оптимизация выполнения count-выражений в запросах

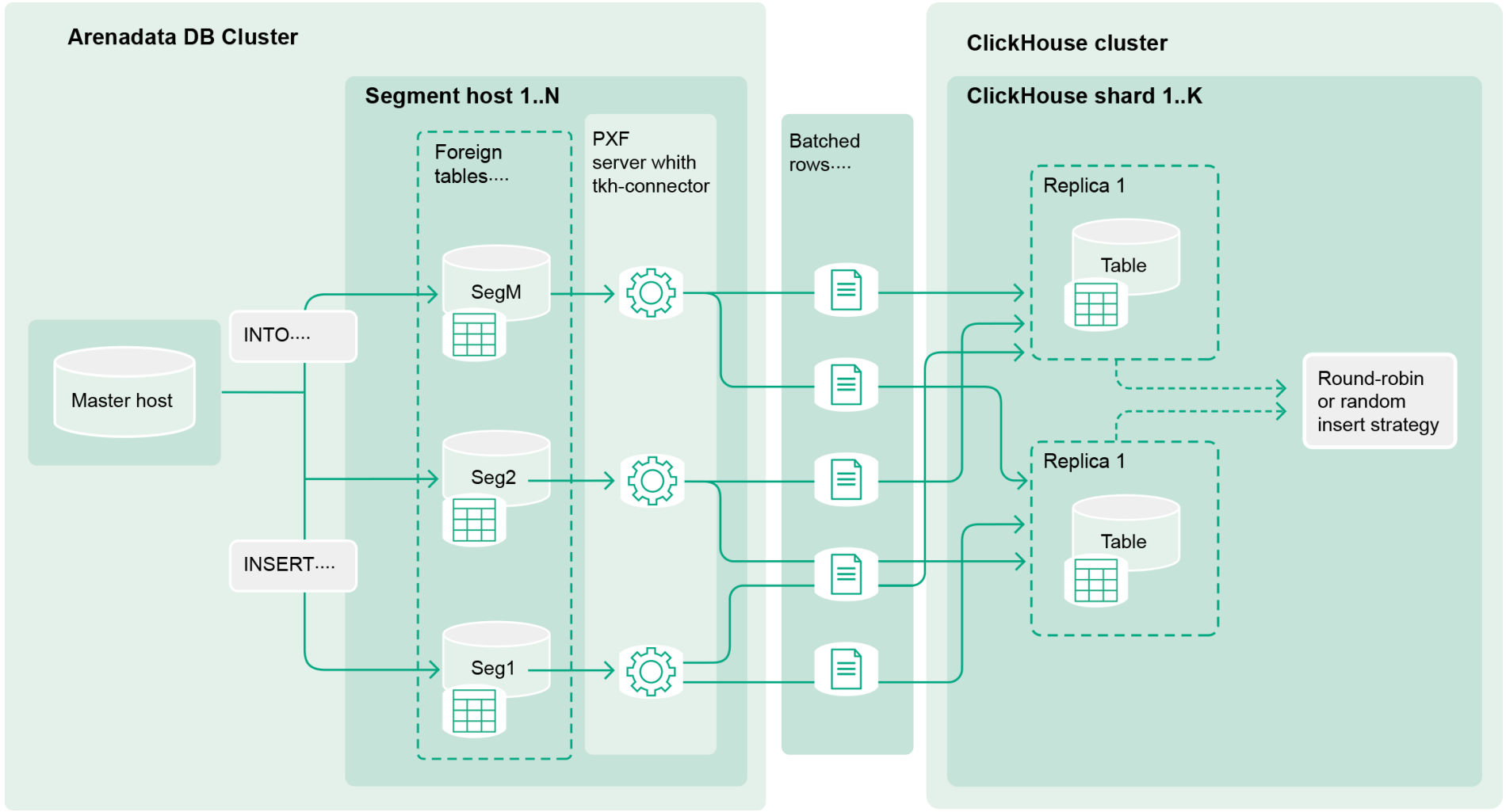
© Arenadata 2023

Синергия Arenadata DB и Arenadata QuickMarts



© Arenadata 2023

Коннектор для нативной передачи данных из ADB в ClickHouse



© Arenadata 2023

Преимущества ADB ClickHouse connector

- Транзакционная вставка из ADB в ClickHouse
- Поддержка запросов из ClickHouse
- Многоуровневые настройки соединения на уровне foreign table
- Механизм распределения равномерной нагрузки по хостам ClickHouse
- Гибкое управление кастомными настройками через команды ALTER SERVER или ALTER FOREIGN TABLE
- Больший контроль над освобождением ресурсов благодаря Foreign Data Wrapper API

```
CREATE SERVER clickhouse_server
FOREIGN DATA WRAPPER tkh_fdw
OPTIONS (
                database 'default',
                hosts 'dev-adqm-01:8123, dev-adqm-02:8123,
dev-adqm-03:8123, dev-adqm-04:8123',
                distribution_type 'random',
                lines_batch_size '100000',
                send_compressed 'true',
                send_delay '300',
                insert_distributed_sync 'true',
                use_staging 'true',
                staging_table_name_format '$_tmp_$');

CREATE FOREIGN TABLE tkhemali_test (id INT,
value_string TEXT)
SERVER clickhouse_server
OPTIONS (resource 'test_distr',
                hosts 'dev-adqm-01:8123',
    cluster 'default_cluster');
```

ADB ClickHouse Connector входит в состав Arenadata DB Enterprise Edition

© Arenadata 2023

Референсные кейсы

ARENADATA

Решаемые задачи с Arenadata DB

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 инженерных систем

- Импортозамещающие ПАК на базе Arenadata DB и Arenadata QuickMarts.
- Возможно добиться снижения сложности эксплуатации озера данных путем миграции с Hadoop на Arenadata DB.
- Прозрачная методика перехода, минимум рисков и сохранение всех преимуществ.

Корпоративная аналитика от BI до AI

- Интеграция с любыми BI-системами, совместимыми с JDBC/ODBC.
- Возможность быстрого развёртывания и сворачивания зон «песочниц» для пилотных проектов и проверки статистических гипотез.
- Работа со всеми аналитическими инструментами в единой среде.

© Arenadata 2023

Кейс №1: увеличение производительности КХД

ЗАДАЧА

Проблема с производительностью хранилища данных и отчетности, построенной на базе классических реляционных СУБД, например Oracle DB, MS SQL Server, PostgreSQL.

РЕШЕНИЕ

Повышения производительности можно достичь за счет использования MPP-СУБД Arenadata DB.

ВТБ

- в 2 раза повысилось качество и доступность данных;
- в 2 раза сократится срок вывода на рынок новых продуктов;

Росбанк

- сократили затраты на лицензии и „железо“,
- стали тратить меньше времени на аналитику,
- все данные у нас теперь аккумулируются в одном хранилище.

ПСБ Банк

- более чем в 10 раз быстрее обработка большинства типовых запросов.

Кейс №2: быстрая аналитика

ЗАДАЧА

Быстро оценивать результаты продаж и маркетинговых кампаний, снижать время на вывод новых продуктов и эффективно управлять рисками на базе аналитики, построенной на качественных данных.

РЕШЕНИЕ

Arenadata DB прекрасно справляется с задачами построения корпоративного хранилища данных (КХД), аналитики с помощью BI-инструментов, AD-HOC запросов и data science уровня предприятия.

ВТБ

- с 6 дней до 1 дня сократятся сроки формирования управленческой отчетности и подготовки витрин данных для бизнеса

X5

- Компания получила надежный инструмент, доступный любому из 2000 аналитиков, для быстрого принятия решений на основе больших данных.
- 2 тыс. аналитиков могут работать параллельно

FMCG компания

- Российский бизнес полностью управляет решением и собственными аналитическими процессами.
- Это позволит существенно сократить время для разработки и внедрения новой функциональности в будущем.

Кейс №3: снижение TCO КХД

ЗАДАЧА

Высокая стоимость владения системами КХД особенно часто возникает в компаниях, которые строят свои КХД на базе узкоспециализированных аналитических решений таких как TERADATA, ORACLE Exadata или SAP HANA.

РЕШЕНИЕ

Снижения стоимости владения КХД за счет оптимизации затрат на лицензирование при сохранении высокой производительности можно добиться путем миграции на Arenadata DB.

ВТБ

- Сокращение стоимости платформы на 6 млрд руб. за 5 лет за счет внедрения Arenadata.

X5

- снижение затрат на хранение данных на 20% за счет использования разно температурного хранилища

Кейс №4: импортозамещение в КХД

ЗАДАЧА

Вопрос перехода на отечественное ПО обусловлен рядом постановлений РФ и затруднительной поддержкой и развитием платформ от иностранных вендоров.

РЕШЕНИЕ

Arenadata DB решает задачу импортозамещения для аналитических систем

ВТБ

- Миграция данных (с Teradata и Oracle Exadata) на стек программных продуктов Arenadata На Arenadata переведены большинство аналитических информационных систем ФНС РФ.
- На стеке Arenadata в промышленной эксплуатации сейчас работает почти 400 серверов

X5

- Миграция с SAP Hana и 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.

Ак Барс Банк

- Миграция корпоративного хранилища данных с IBM Netezza на новую платформу Arenadata DB.

Матрица соответствия программных продуктов

Мигрируемый продукт	Тип нагрузки / задача	Куда мигрируем
Oracle (< 1000 TPS), RAC, Exadata (DWH), SAP BW	OLAP	Arenadata DB
Teradata, IBM Netezza, Vertica, Exasol, Presto, Tanzu/Pivotal Greenplum, SAP Sybase IQ	DWH, Реляционная MPP СУБД	Arenadata DB
Snowflake, RedShift	Реляционная облачная MPP, DWH	Arenadata DB (cloud)
Azure SQL, Google Big Query	Реляционная облачная СУБД	Arenadata DB (cloud), Arenadata Hadoop (cloud)

© Arenadata 2023

Простые шаги к платформе Arenadata

1. Протестируйте community версию наших продуктов и ознакомьтесь с технической документацией:
https://network.arenadata.io/

2. Пройдите обучение по продуктам Arenadata:
https://arenadata.tech/courses/

3. Включите наши продукты в свой scope проектов

4. Будьте в курсе релизов, событий и новостей:
Хабр , ТГ, YouTube

© Arenadata 2023

Оставайтесь на связи

Наши новости
в телеграм-канале

ARENADATA