



Arenadata QuickMarts

Быстрая колоночная СУБД
для эффективной аналитики

ClickHouse для Enterprise



Arenadata QuickMarts (ADQM) — кластерная колоночная СУБД, созданная на основе ClickHouse. Она позволяет выполнять аналитические запросы в режиме реального времени на структурированных больших данных.

С помощью ADQM вы можете в режиме реального времени генерировать аналитические отчёты разного плана, используя большие объёмы информации, хранящейся в плоских витринах.

ADQM многократно быстрее традиционных СУБД.



Open Source проект ClickHouse

- Open Source (ASF 2)
- Старт проекта в 2016 году
- Инициатор разработки – Yandex
- Самая быстрая колоночная OLAP СУБД для своих типов запросов
- Векторные CPU-инструкции
- Эксплуатируются кластеры в сотни узлов
- Используется по всему миру



Строковые и колоночные СУБД

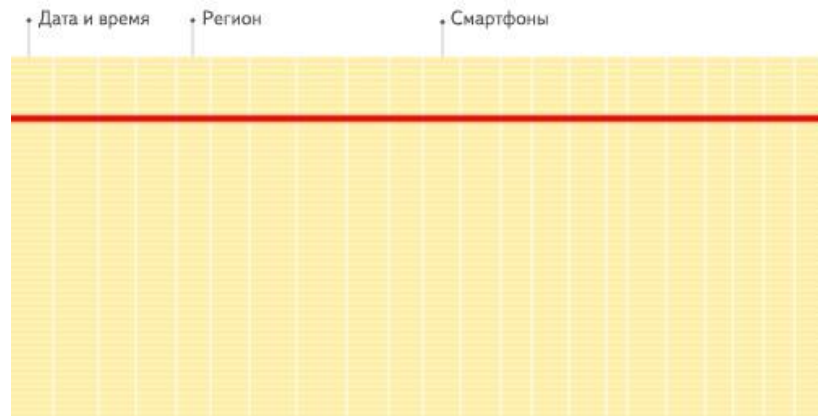


Рисунок 1. Принцип работы строковой СУБД

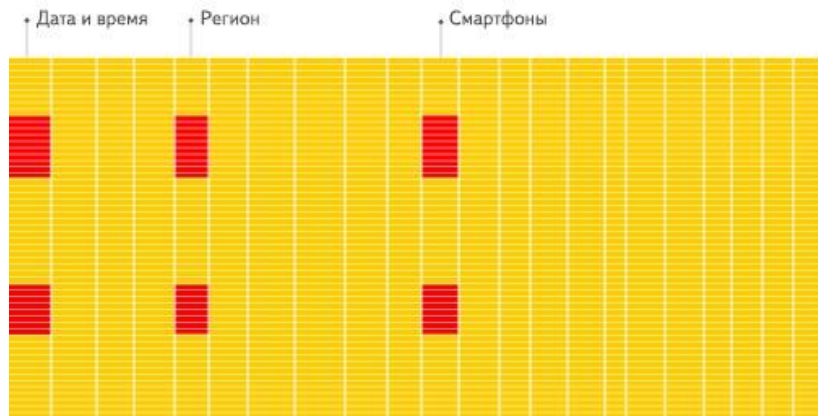


Рисунок 2. Принцип работы колоночной СУБД

Архитектура ClickHouse



ClickHouse – колоночная СУБД

Рядом со значениями не хранятся никаких лишних данных, что позволяет быстро выполнять аналитические запросы (вычисление агрегатных значений по сырым данным).



Обработка по векторам

Данные не только хранятся в столбцах, но и обрабатываются по векторам - кусочкам столбцов. За счёт этого достигается высокая эффективность использования ресурсов процессора.

Преимущества ClickHouse

- Быстрые вставки и хранение данных в отсортированном виде
- Максимально быстро выполняет OLAP запросы
- Возможность хранить петабайты данных
- Максимально эффективное использование оборудования за счёт поддержки векторизации, параллелизма и сжатия данных (LZ4 и ZSTD)
- Горизонтальное масштабирование между несколькими хостами
- Поддержка отказоустойчивости (DR и HA) во множестве дата-центров
- Гибкий язык запросов (SQL)
- Массивы и вложенные типы данных
- Удобная интеграция с внешними системами



Производительность на грани фантастики

- Типичные запросы выполняются быстрее чем за секунду
- > 100 раз быстрее чем Hadoop и обычные СУБД
- 100 млн - 1 миллиард строк в секунду на одном узле кластера
- До 2 терабайт в секунду для кластера на 400 узлов



Отличия ADQM от ClickHouse

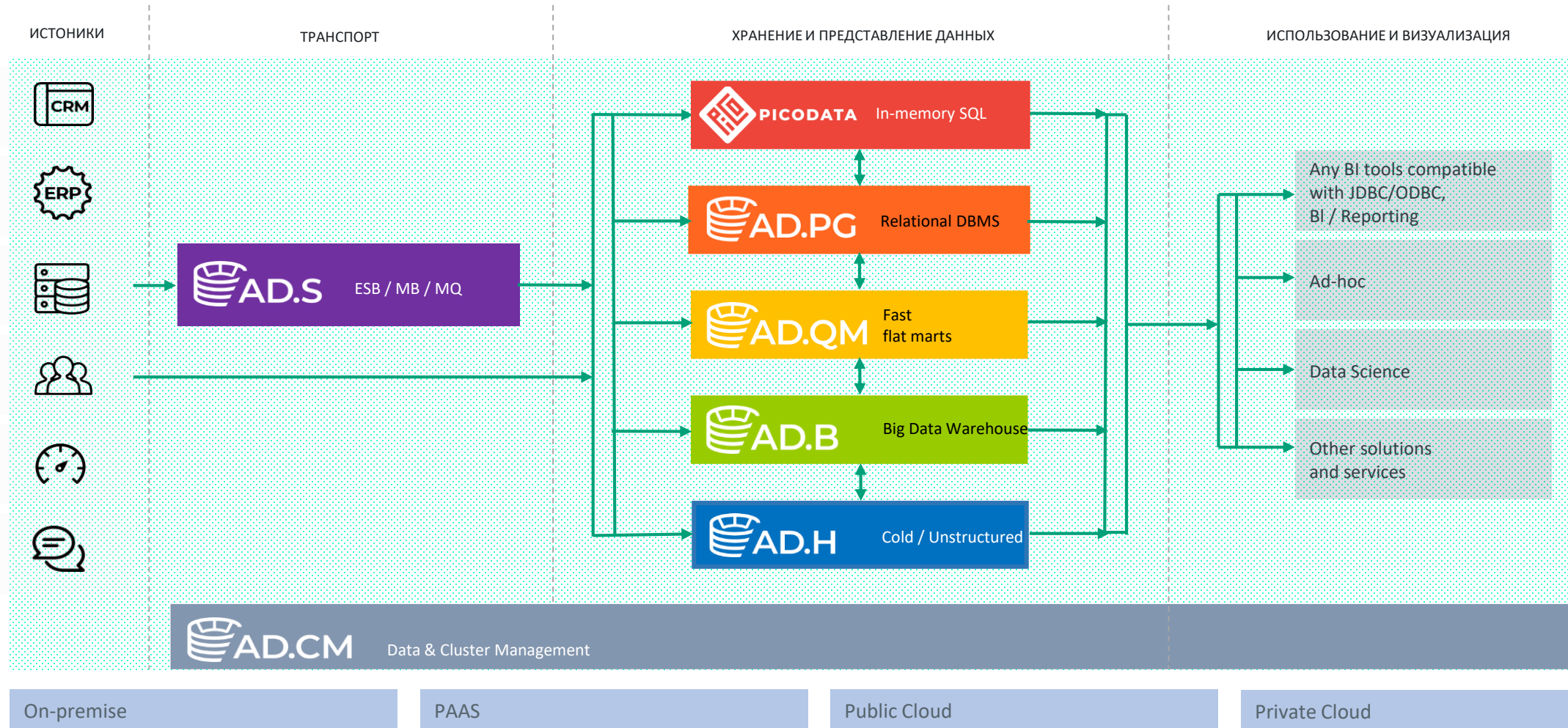
Функционал	Open Source дистрибутив	ADQM EE
Core-функционал	+	+
- Управление деплоем и апгрейдом, - Расширение кластера, - Мониторинг & alerting	-	+
Документация	Без версионирования	+
Производительные коннекторы Spark <-> ClickHouse, Kafka <-> ClickHouse	-	+
Оффлайн установка	-	+
Включение в состав поставки и интеграция с Chproxy	-	+
Поддержка отечественных операционных систем (Альт 8 СП / Astra Linux 1.7 SE «Орёл»)	-	+
Инструментарий управление бэкапами*	-	+
ADQM Control (анализ состояния кластера и alert-менеджер)*	-	+

* - в разработке

Отличия ADQM от ClickHouse

Функционал	Open Source дистрибутив	ADQM EE
Включение в Реестр отечественного ПО	-	+
Доработка и развитие функционала (core) под требования заказчиков	-	+
- Поддержка (8*5 / 24*7) - Обучение по продуктам	-	+
Проведение пилотных проектов и PoC	-	+
Доп.консалтинговые услуги (DBaaS, Smart Start, TAM, Аудит)	-	+

Arenadata Enterprise Data Platform



Что внутри?

 AD.B Big Data Warehouse	 Greenplum
 AD.H Cold / Unstructured	 hadoop
 AD.QM Fast flat marts	 ClickHouse
 AD.PG OLTP DataBase	 PostgreSQL
 PICODATA In-memory SQL	Tarantool
 AD.S ESB / MB/ MQ	 kafka  nifi
 AD.CM Data & Cluster Management	 ANSIBLE

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 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:

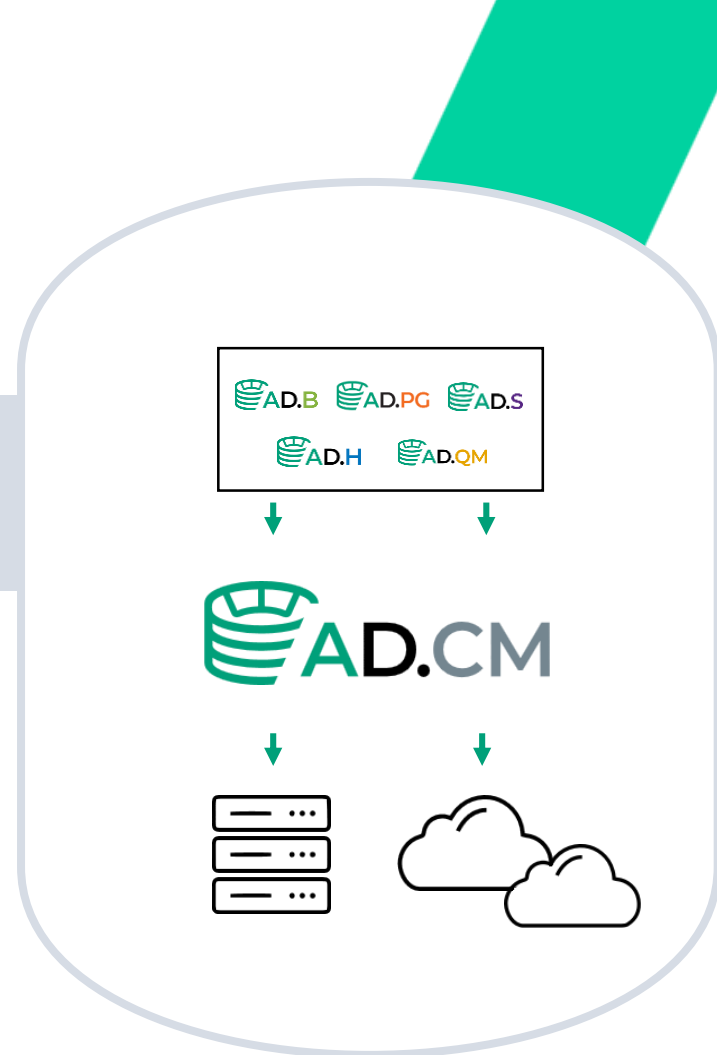
- Установка и настройка
- Обновление
- Управление
- Мониторинг
- Настройка прав доступа
- Интеграция с другими сервисами
- Контейнеризация

Возможности работы с инфраструктурой:

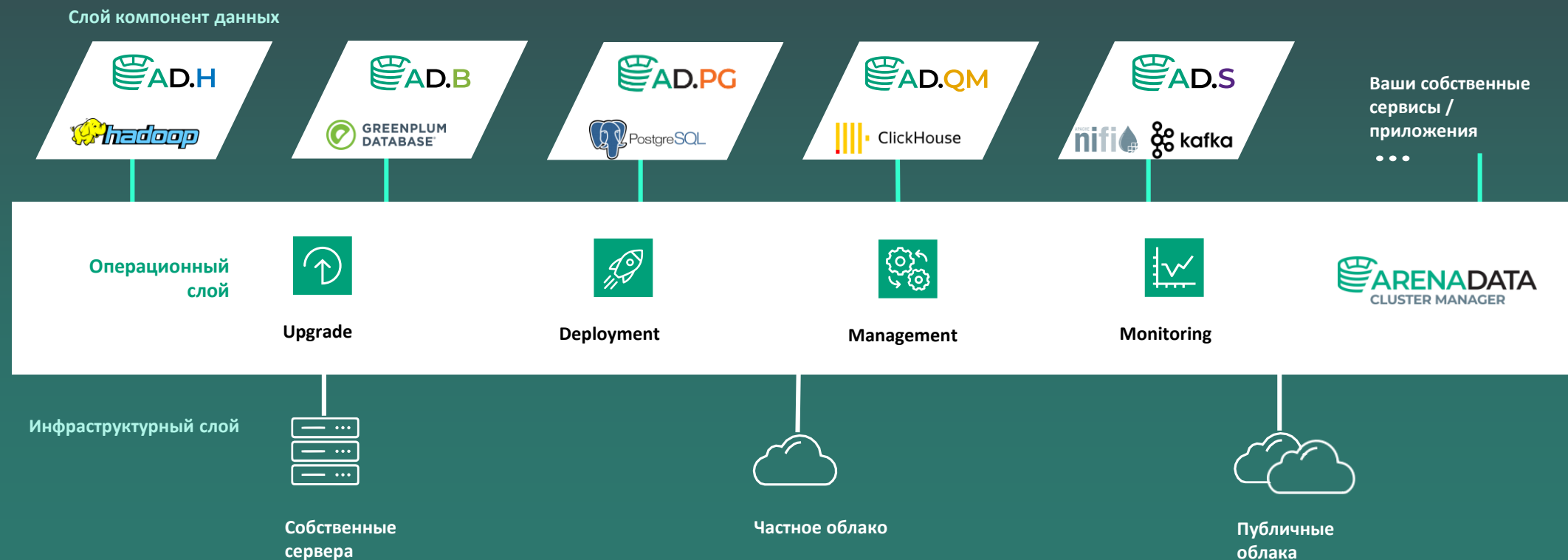
- Создание и удаление виртуальных машин
- Конфигурирование ОС
- Мониторинг
- Управление пользователями
- Настройка прав доступа

Дополнительные возможности:

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

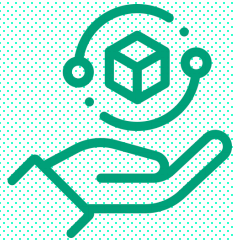


Гибридный корпоративный ландшафт



Удобный мониторинг

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

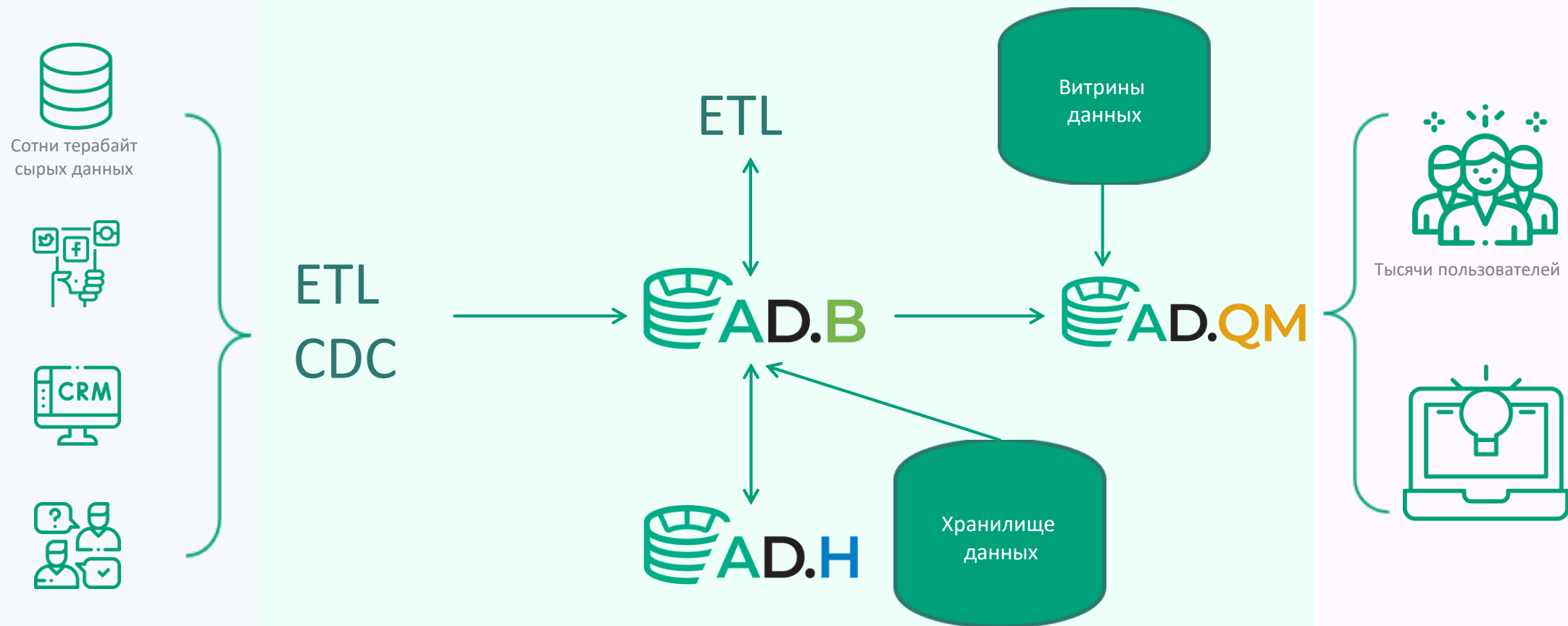


Мониторинг входит в комплект Community & Enterprise Edition

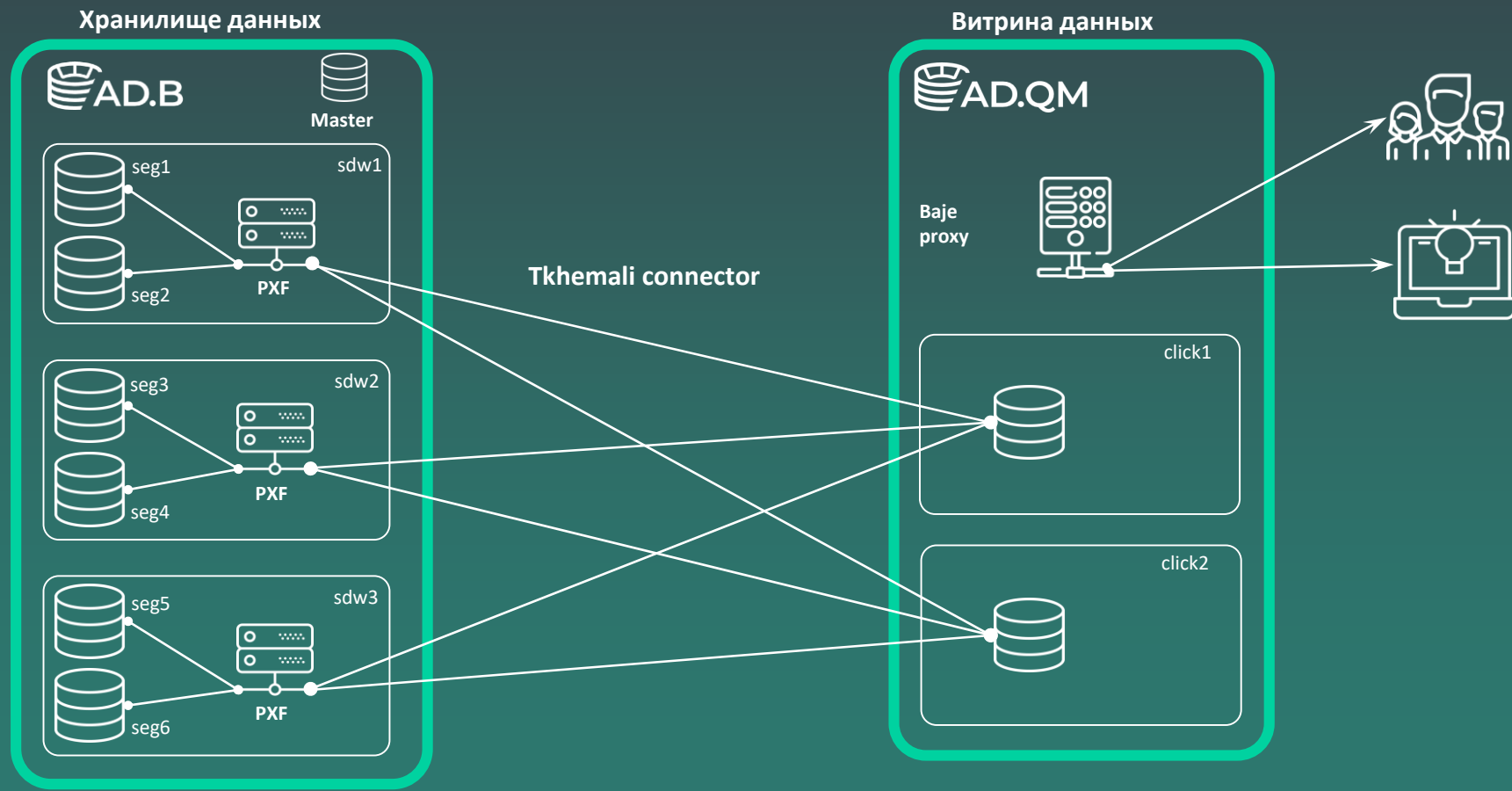
Архитектура Arenadata QuickMarts



Синергия Arenadata QuickMarts и Arenadata DB



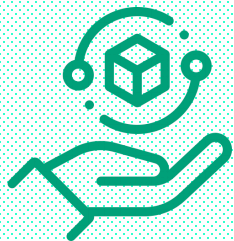
Коннектор для нативной передачи данных из ADB в ADQM



Производительность и надёжность Tkhemali connector

- Параллельная интеграция кластеров - нет узких мест
- Меньше потоков с помощью PXF batch union - количество потоков не зависит от количества сегментов
- Можно уменьшить количество вставок динамически в зависимости от нагрузки ADQM (ClickHouse)
- Транзакционная вставка из ADB в ADQM

```
create writable external table tkh_ext (  
    id int,  
    text1 text,  
    text2 text  
)  
LOCATION ('pxf://tkh_test?PROFILE=TKH&URL=click-1:8123,click-  
2:8123,click-3:8123')  
FORMAT 'TEXT'  
ENCODING 'UTF8'  
distributed by (id);  
  
insert into tkh_ext select * from internal table;
```



Tkhemali connector входит в состав Arenadata DB Enterprise Edition

ADQM Spark Connector

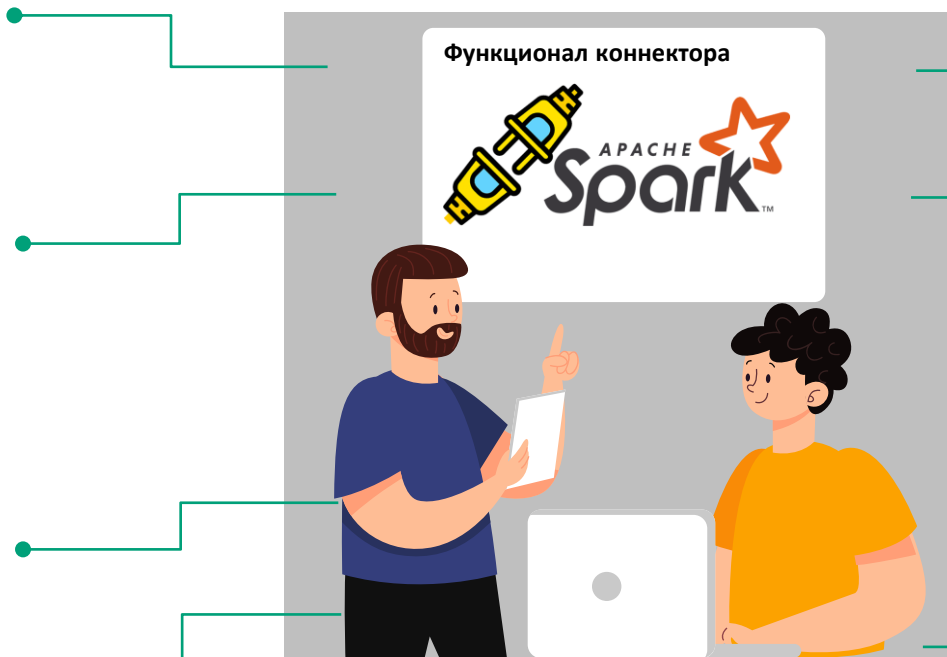
Чтение данных из ADQM

Запись данных в ADQM
используя различные
режимы:

- overwrite
- append
- errorIfExists

Поддержка
структурированных данных

Автоматическое
формирование схемы данных



Настраиваемое партиционирование

- целочисленные типы
- даты

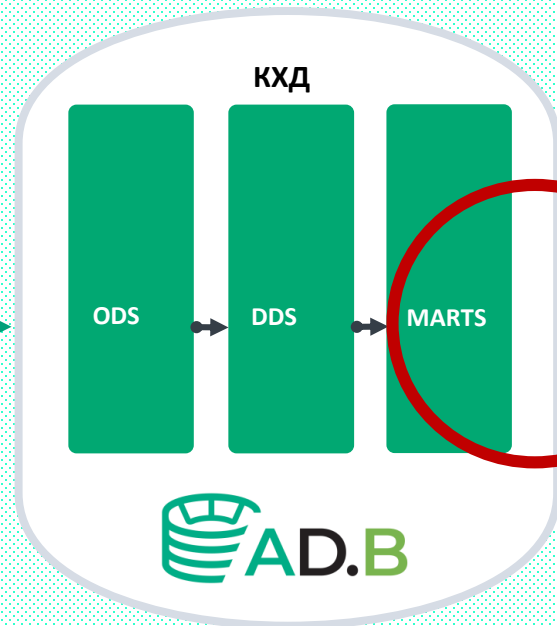
Push-down операторов:

- исключение колонок
- push-down фильтров

Выполнение произвольного SQL

Сценарий использования №1

Источники



Проблема: аналитические СУБД не способны обеспечить время отклика в миллисекунды

100s QPS
WHERE / GROUP BY

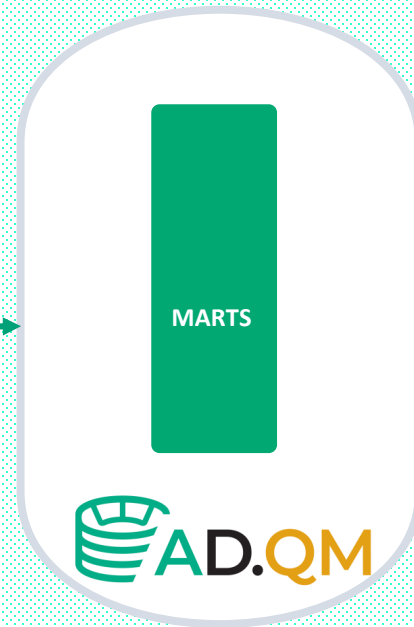
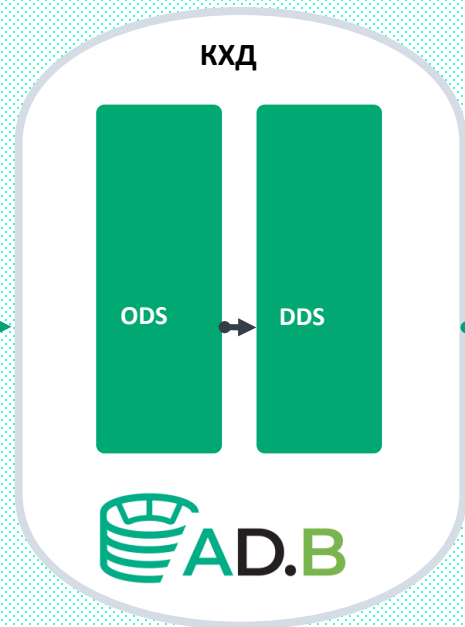
BI / Ad-hoc



Любые BI-средства, совместимые с
JDBC/ODBC

Сценарий использования №1

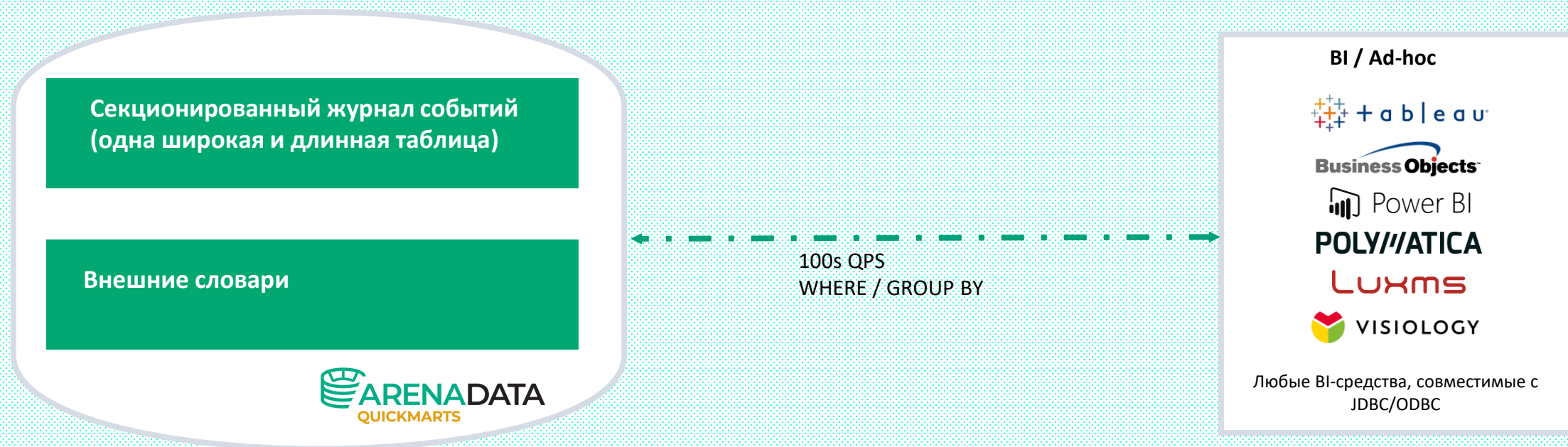
Источники



100s QPS
WHERE / GROUP BY



Сценарий использования №2



- Широкая и очень длинная таблица данных (событий)
- Все остальные таблицы в виде словарей в памяти

Использование ADQM



В каких отраслях?

- Ритейл и e-commerce
- Банки и финансовые операции
- Телекоммуникация
- Мониторинг технических и бизнес-метрик
- Информационная безопасность
- Веб-аналитика и аналитика мобильных приложений
- Реклама и RTB
- Онлайн-игры
- Интернет вещей



Как использовать?

- В связке с Arenadata DB
- Анализ временных рядов
- Быстрые витрины
- Data Science
- Анализ серверных журналов

Матрица соответствия программных продуктов

Заменяемая технология	Тип нагрузки / задача	На что заменяем
Oracle DB/Exadata, Microsoft SQL Server, Teradata, Vertica	Витрины данных в КХД	Arenadata QuickMarts + Arenadata DB
SAP HANA	In-memory RDBMS	Arenadata QuickMarts + Arenadata DB
Amazon Timestream, HCL Informix, InfluxDB Enterprise, Kdb+, TimescaleDB	СУБД для анализа временных рядов	Arenadata QuickMarts

 **ARENADATA**
QUICKMARTS



teradata.

SAP HANA

ORACLE®
EXADATA

Microsoft®
SQL Server®

 Amazon
TimeStream

 influxdb

ADQM — российское программное обеспечение

Регистрация в Реестре
Минцифры РФ

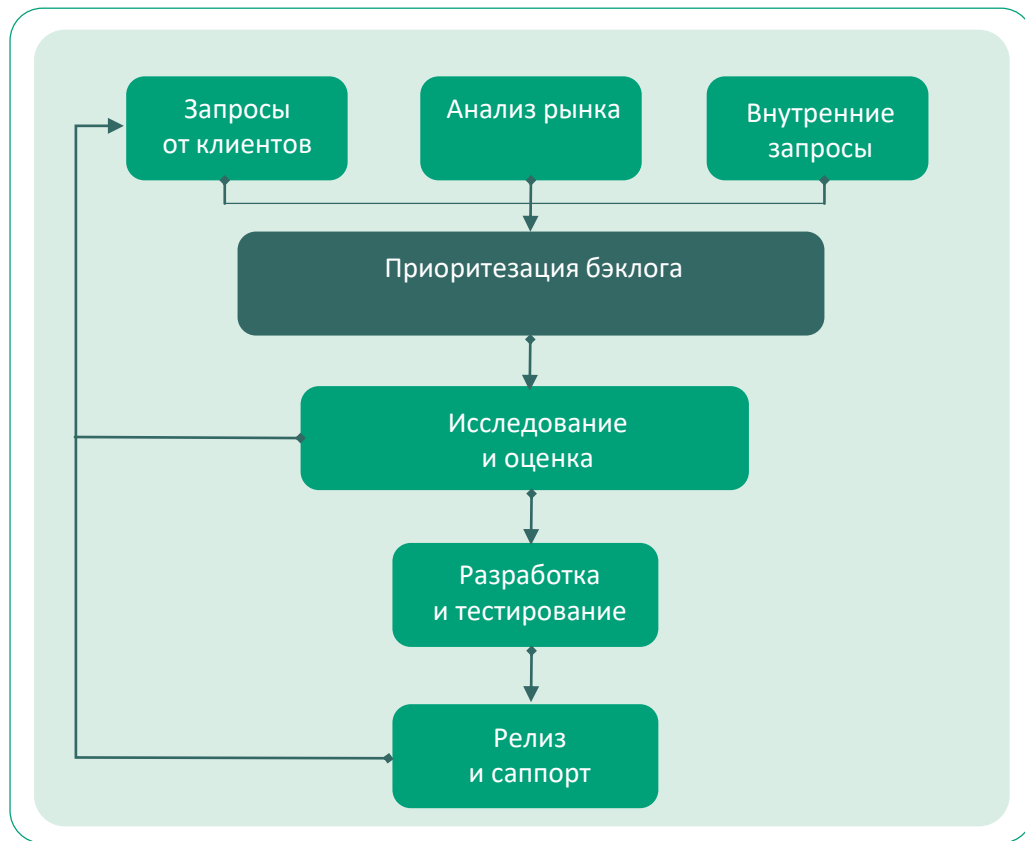
Отсутствие валютных
рисков - цены в рублях

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

Поддержка,
документация и
учебные курсы на
русском языке



Процесс развития продуктов Arenadata



За 2021 год

250+

запросов
от клиентов

130+

реализовано

Направления развития ADQM



Уровни поддержки Arenadata QuickMarts

24x7

8x5

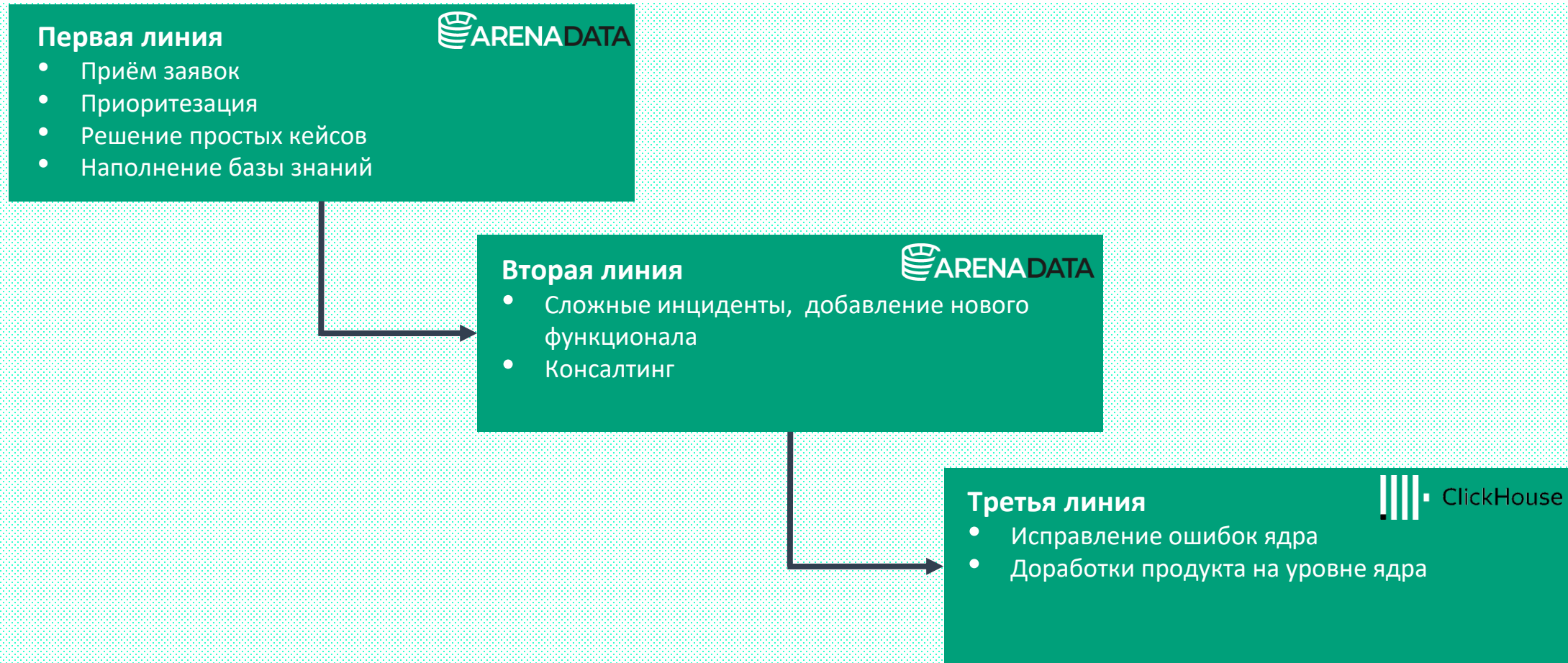


Поддержка разработчиков

Наши доработки, опубликованные в ClickHouse:

- ORC input format
- Доработка jdbc и jdbc-bridge
- Доработка коннектора к HDFS
- Аутентификация в Kerberos для HDFS и Zookeeper (в разработке)
- Поддержка OR в JOIN

Входящий запрос



Услуги по поддержке клиентов

- Vendor support
- Выделенный сервис-менеджер

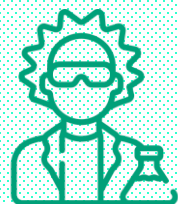
Дополнительные услуги:

- DBA as a Service
- Регулярный аудит
- TAM
- Smart Start



Корпоративная платформа хранения и обработки больших данных

Курсы по Arenadata QuickMarts



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.

Наш успех – доверие клиентов

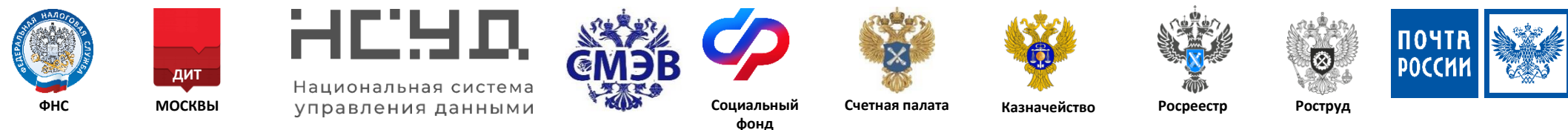
РИТЕЙЛ



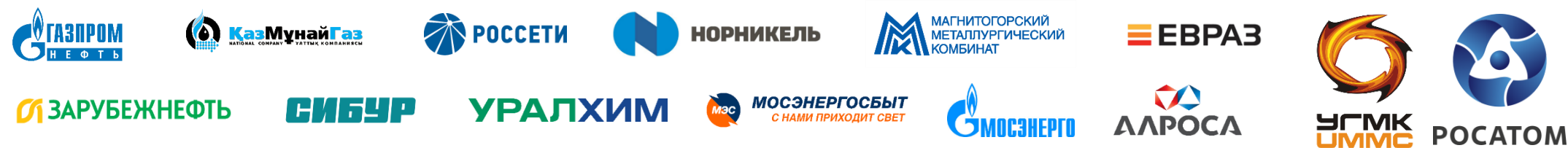
БАНКИ
ТЕЛЕКОМ



ГОСУДАРСТВО



ПРОМ
ЭНЕРГО
ДОБЫЧА
ПРОЧЕЕ



3 шага к Arenadata QuickMarts

1

Расскажите нам о задачах, ИТ-инфраструктуре и получите рекомендации по компонентам платформы и выбору партнёра.

2

Загрузите бесплатно необходимые пакеты для платформы данных с нашего сайта.

3

Разверните нашу платформу в своей инфраструктуре и отправьте нам технические и функциональные требования, чтобы получить индивидуальное предложение.





Корпоративная платформа хранения
и обработки больших данных

**Оцените потенциал Big Data
и Open Source вместе с Arenadata Enterprise Data
Platform**

- Узнайте больше на arenadata.tech
- Скачайте бесплатно на store.arenadata.io
- Получите консультацию по почте info@arenadata.io