



Быстрая колоночная СУБД для эффективной аналитики

Arenadata QuickMarts

Yandex ClickHouse для Enterprise



Arenadata QuickMarts (ADQM) — кластерная колоночная СУБД, созданная на основе Yandex 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 Yandex

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



Строковые и колоночные СУБД



Рисунок 1. Принцип работы строковой СУБД



Рисунок 2. Принцип работы колоночной СУБД



Архитектура Yandex ClickHouse



ClickHouse – колоночная СУБД

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

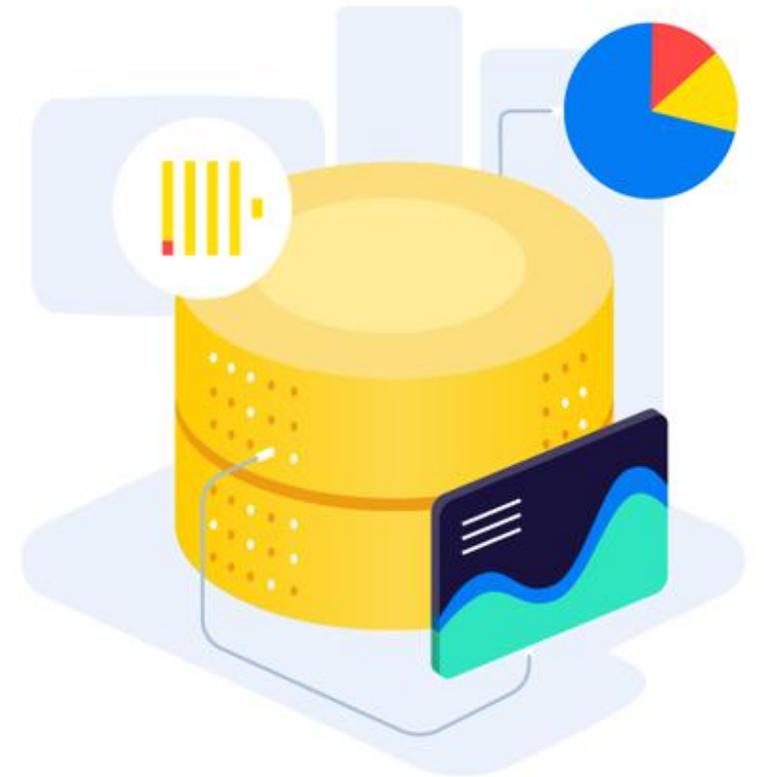


Обработка по векторам

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

Преимущества ClickHouse

- Максимально быстро выполняет OLAP запросы
- Обработка данных в реальном времени
- Возможность хранить петабайты данных
- Использует все ядра CPU, благодаря векторизации и параллельной обработке
- Горизонтальное масштабирование между несколькими хостами
- Отказоустойчивость (кросс-датацентровая репликация)
- Гибкий язык запросов (SQL)
- Современные алгоритмы сжатия (LZ4 и ZSTD)
- Массивы и вложенные типы данных
- Множество интеграций с внешними системами



Скорость на грани фантастики

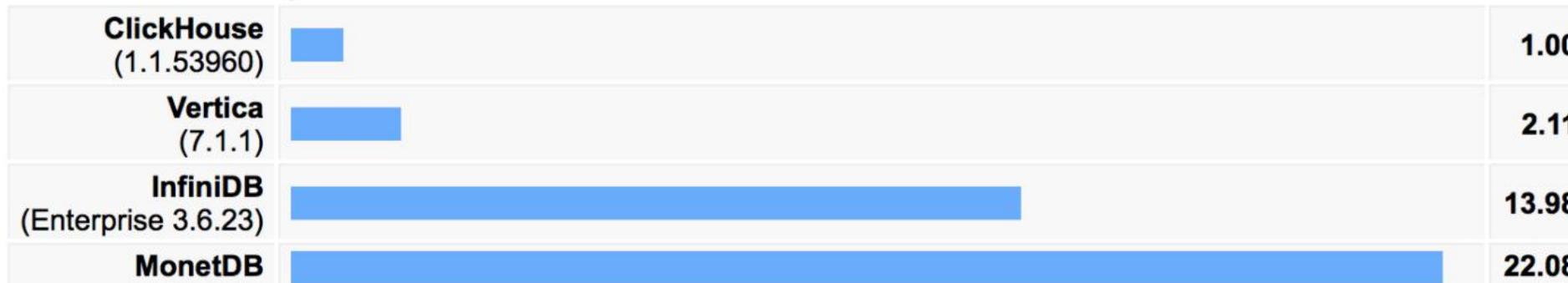
- Типичные запросы быстрее чем за секунду
- > 100 раз быстрее чем Hadoop и обычные СУБД
- 100 млн - 1 миллиард строк в секунду на одной ноде
- До 2 терабайт в секунду для кластера на 400 нод



Производительность ClickHouse



Relative query processing time (lower is better):



[Больше сравнений >>](#)



Возможности по интеграции

- ClickHouse подключается к любой базе данных по JDBC и использует их таблицы как свои
- Может подключаться к другим базам ClickHouse без дополнительной настройки
- Подключается также по REST
- Есть готовые интеграции с другими системами

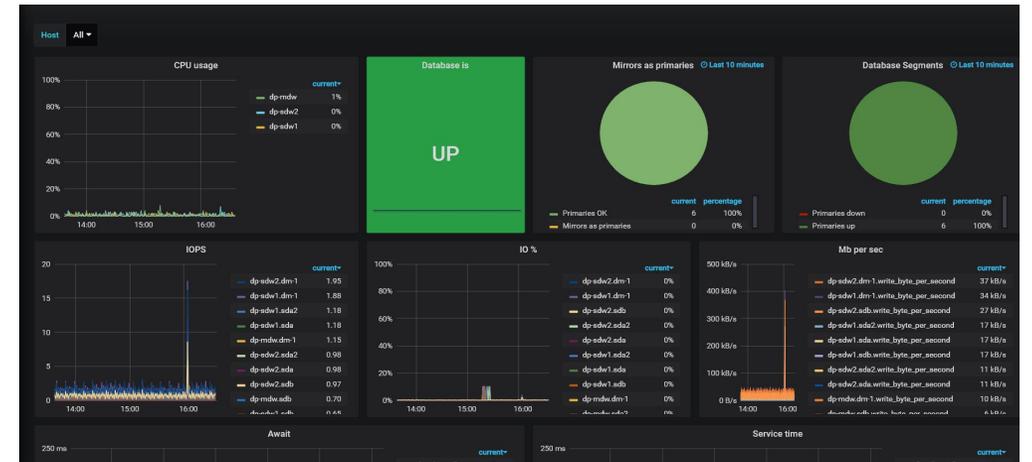
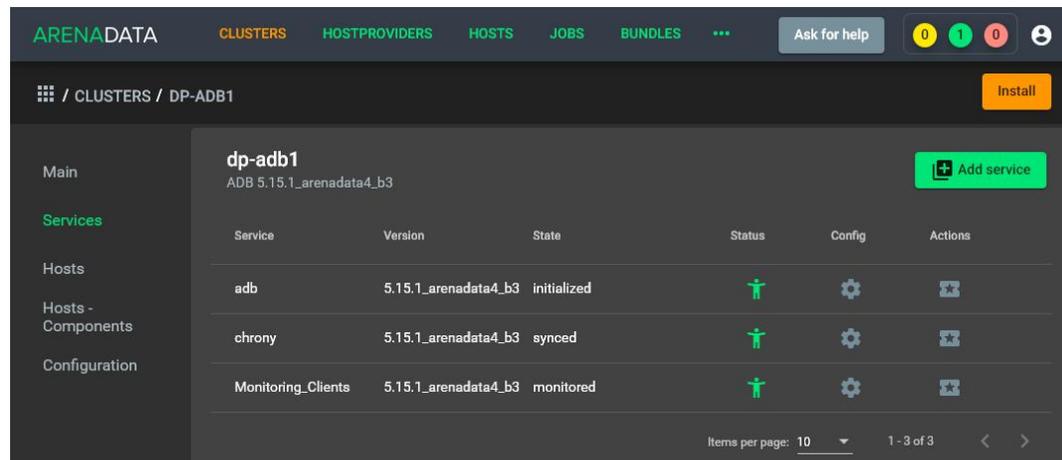
```
SELECT *  
FROM jdbc('mysql://localhost:3306/?user=root&password=root', 'schema', 'table')
```

```
SELECT * FROM remote('ch2', 'db_name', 'table_name', 'user', 'password')
```



Отличие Arenadata QuickMarts от ClickHouse

- Простота установки и настройки из Arenadata Cluster Manager
- Нативная интеграция с Arenadata DB, Hadoop, Streaming в рамках единой платформы Arenadata EDP
- Балансировка нагрузки и распределение доступа к БД с помощью Arenadata Vaje
- Авторизация пользователей, в том числе поддержка интеграции с Kerberos Hadoop

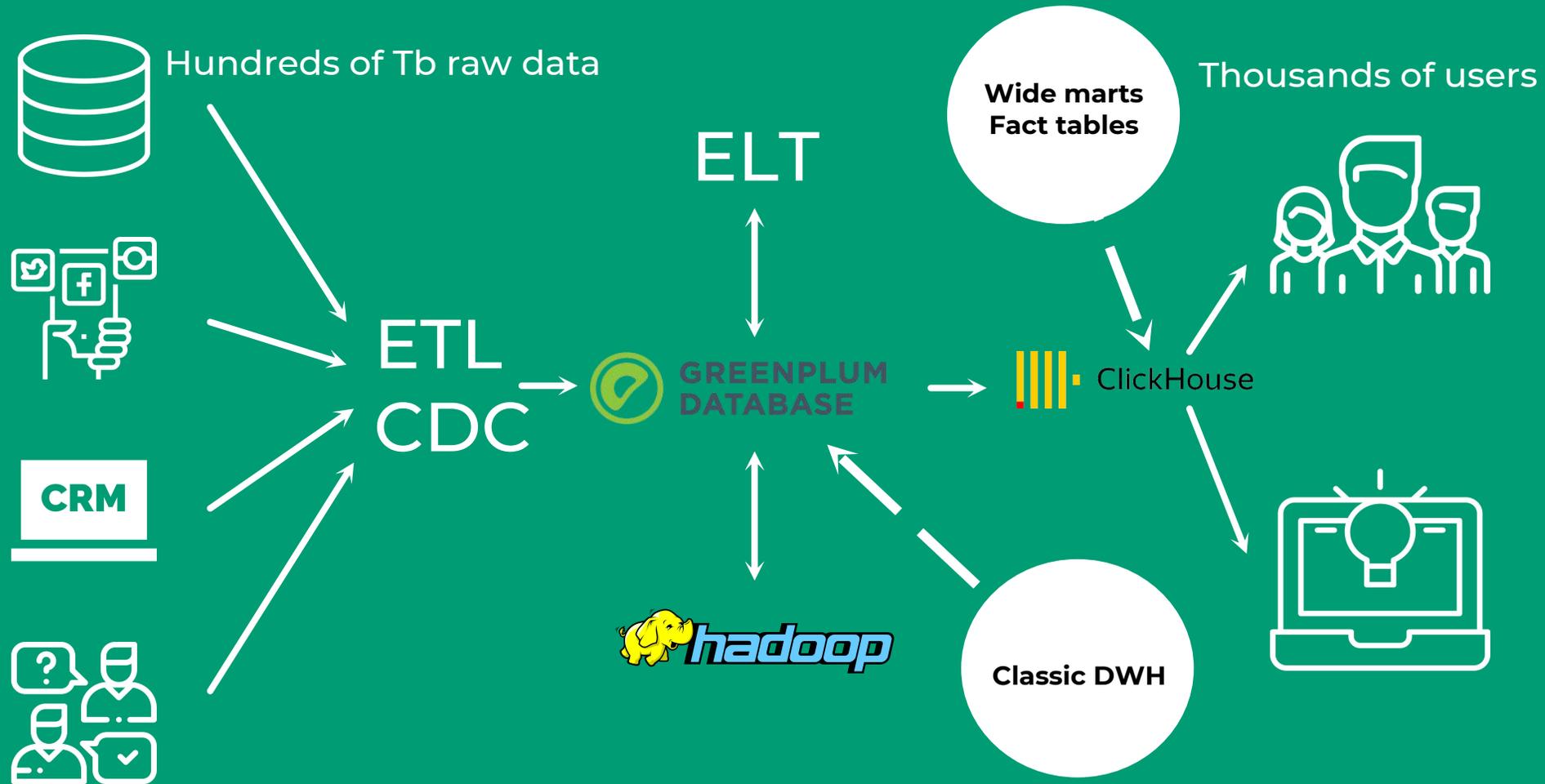


- Мониторинг на базе Graphite и Grafana
- Дополнительное тестирование и контроль качества релизов
- Наличие российской поддержки
- Оперативная доработка и развитие функционала под требования заказчика

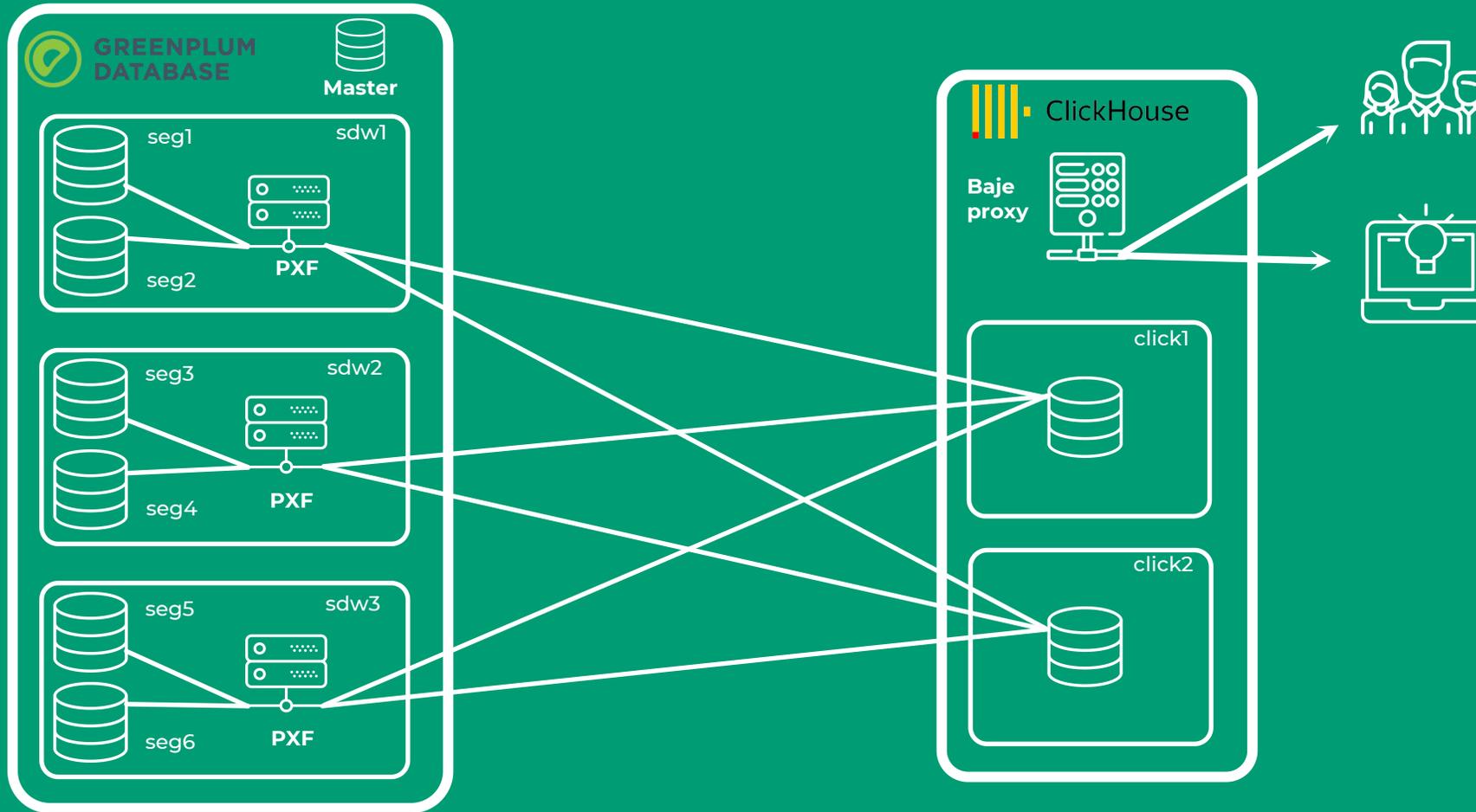
Архитектура Arenadata QuickMarts



Использование ClickHouse и Greenplum



Tkhemali connector



Tkhemali connector для ADB/Greenplum

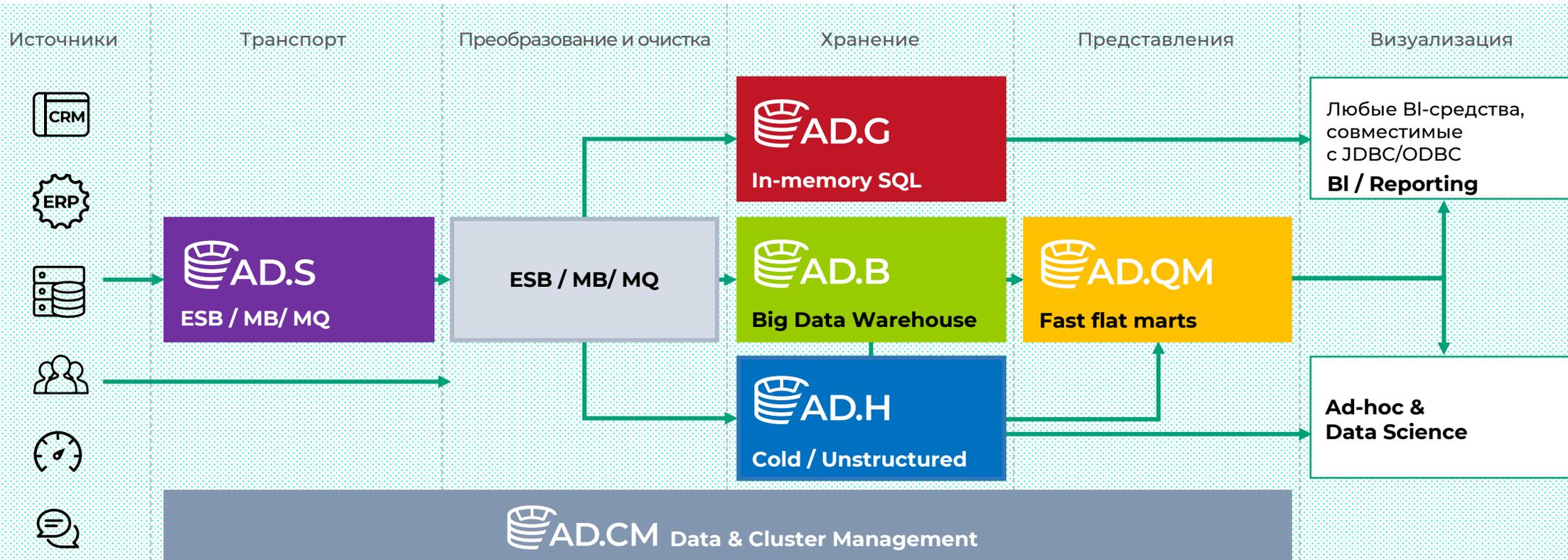
- Параллельная интеграция кластеров - нет узких мест
- Меньше потоков с помощью PXF batch union - количество потоков не зависит от количества сегментов
- Можно уменьшить количество вставок динамически в зависимости от нагрузки ClickHouse
- Транзакционная вставка из Greenplum в ClickHouse

```
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 - часть платформы данных Arenadata EDP



On-premise

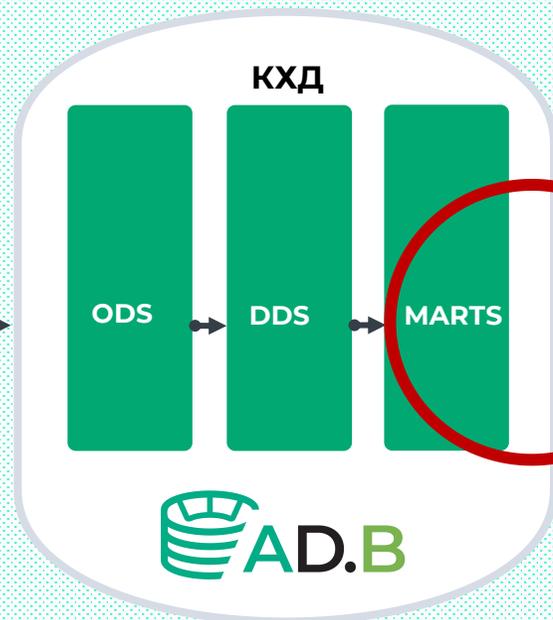
PAAS

Public Cloud

Private Cloud

Кейс №1 использования ADQM

Источники



Проблема: аналитические СУБД не способны обеспечить время отклика в миллисекунды

100s QPS
WHERE / GROUP BY

BI / Ad-hoc



POLY//ATICA

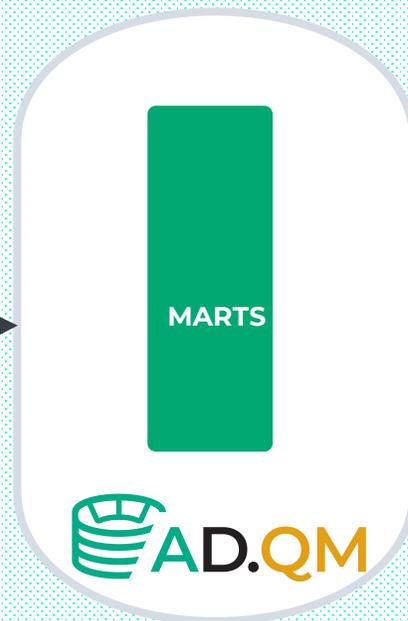
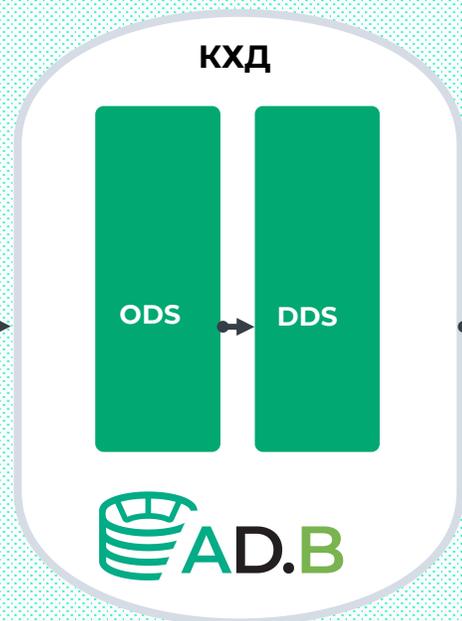
LUXMS



Любые BI-средства,
совместимые с JDBC/ODBC

Кейс №1 использования ADQM

Источники



100s QPS
WHERE / GROUP BY

BI / Ad-hoc



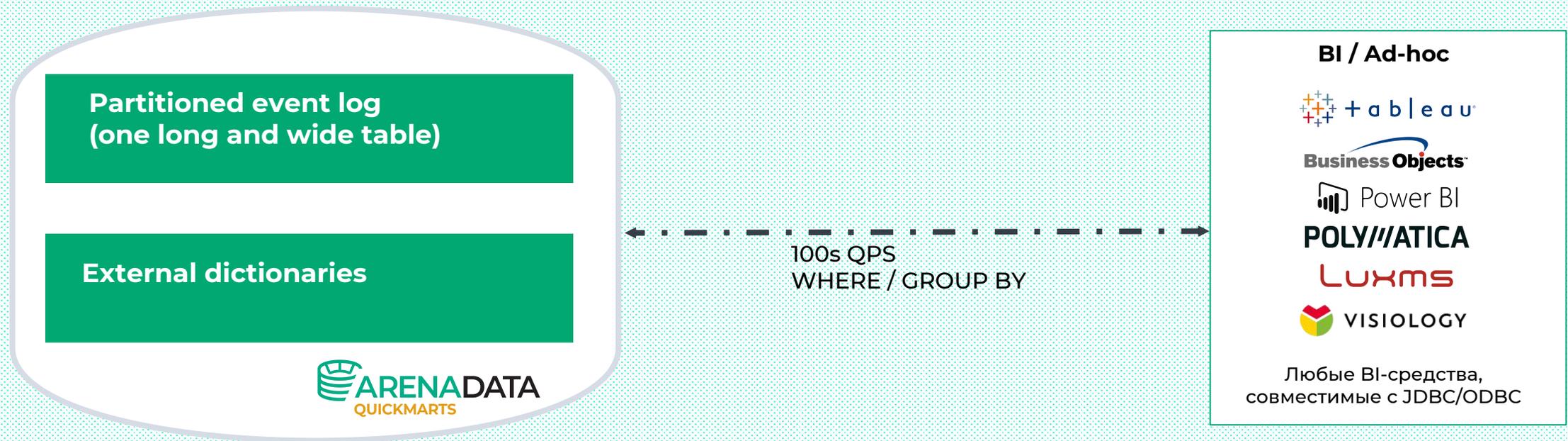
POLY//ATICA

LUXMS



Любые BI-средства,
совместимые с JDBC/ODBC

Кейс №2 использования ADQM



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

Использование ADQM



В каких отраслях?

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



Как использовать?

- В связке с Arenadata DB / Greenplum
- Аналитика над структурированными логами и событиями
- Быстрые витрины
- Распределённые вычисления для бизнеса и науки
- Анализ серверных логов

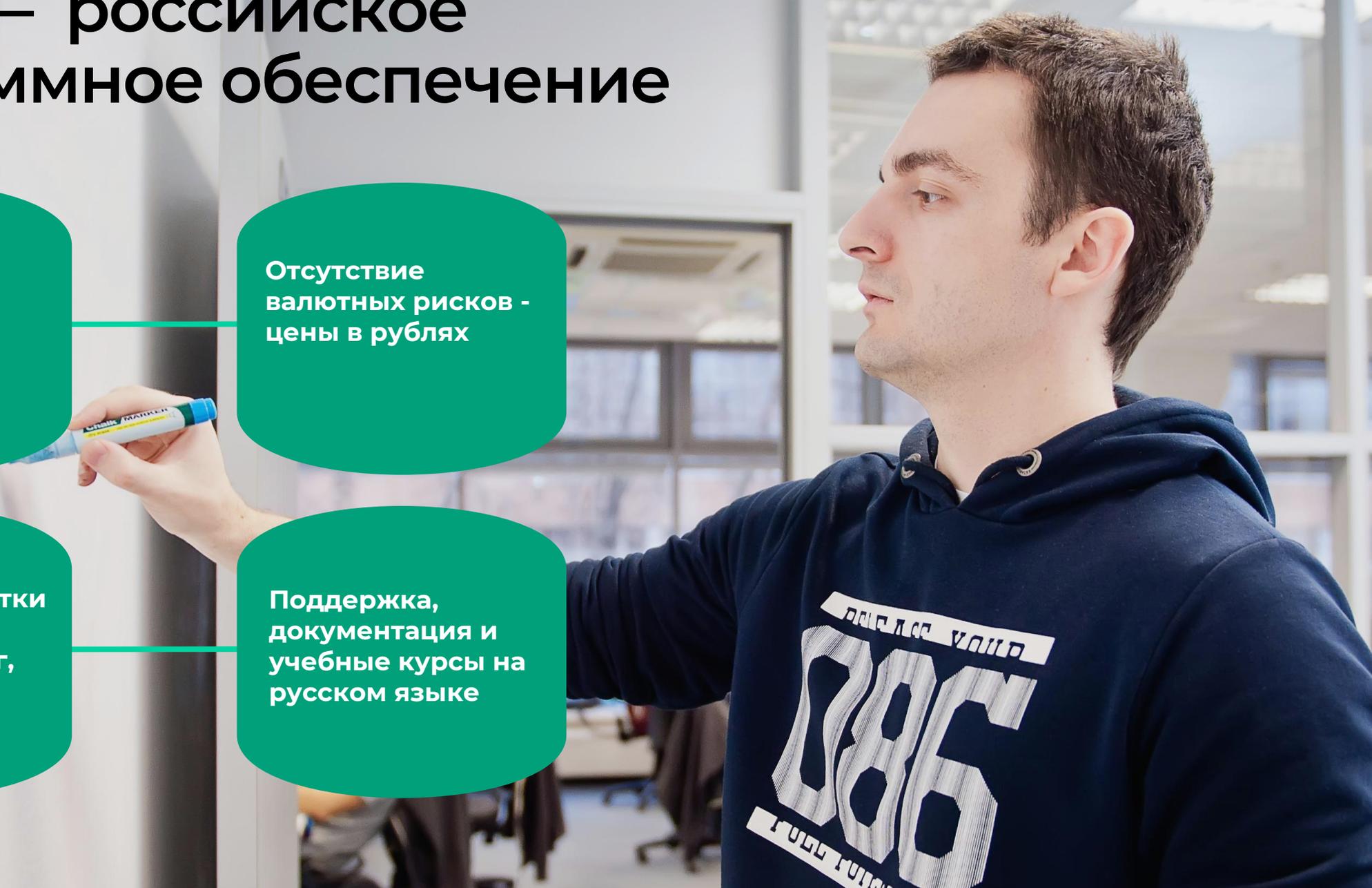
ADQM — российское программное обеспечение

Регистрация в
Реестре
Министерства
Связи

Отсутствие
валютных рисков -
цены в рублях

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

Поддержка,
документация и
учебные курсы на
русском языке



Поддержка 24x7

Санкт-Петербург

Москва

Алма-Аты

Хабаровск

Сотрудники
распределены
по часовым
поясам

Входящий запрос



Первая линия



- Приём заявок
- Приоритезация
- Решение простых кейсов
- Наполнение базы знаний



Вторая линия



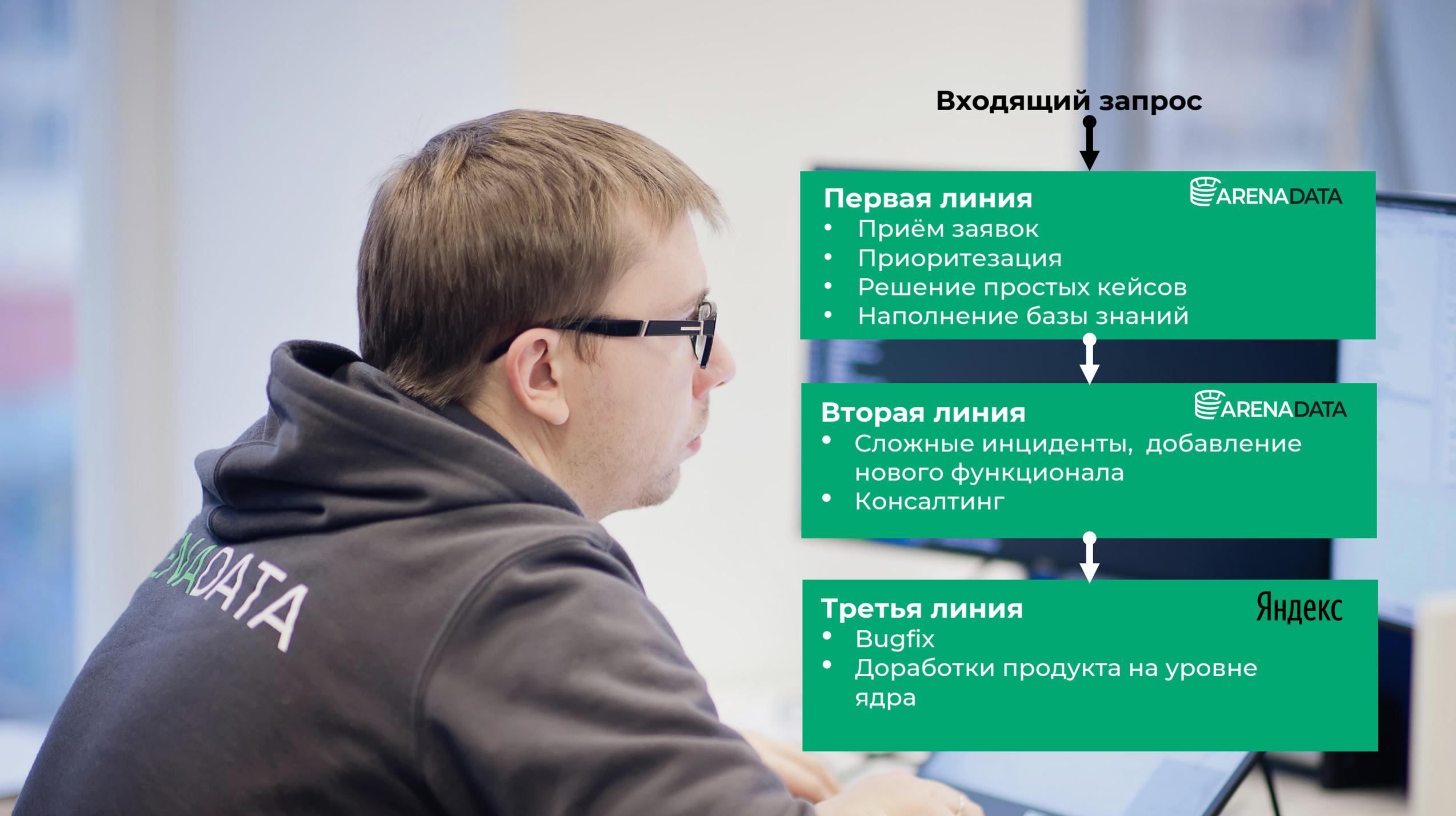
- Сложные инциденты, добавление нового функционала
- Консалтинг



Третья линия



- Bugfix
- Доработки продукта на уровне ядра



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)

Услуги по поддержке клиентов

- Vendor support
- DBA as a service

VIP-пакет

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



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.

Нам доверяют



Наши партнёры

ВНЕДРЕНИЕ

accenture

КРОК

IBS

Jet

DIS
GROUP

АПЛАНА
ЦЕНТР РАЗРАБОТКИ

IBS Platformix

ИТ ФИЛОСОФИЯ

POLYMEDIA

отр

ЛАНИТ

Sapiens
solutions

quadrimum

Rubbles
SBDA Group

LUXMS
группа компаний

ОБЛАЧНЫЕ ПРОВАЙДЕРЫ

MAIL.RU
CLOUD
SOLUTIONS

Yandex Cloud

Ростелеком

КРОК
ОБЛАКО

IBS DataFort

ТЕХНОЛОГИЧЕСКИЕ ПАРТНЁРЫ

accenture digital

@ mail.ru
group

Яндекс

YA
DR

IBM

Informatica

POLY//ATICA

VISIOLOGY

LUXMS
группа компаний

Форсайт
Аналитическая платформа

ARENADATA

3 шага к платформе Arenadata

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

