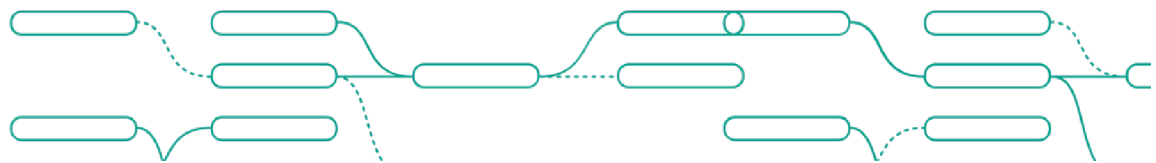




Высокоуровневая архитектура
развертывания

Arenadata Catalog Data Quality
Framework (ADC DQF)



Arenadata Catalog Data Quality Framework (ADC DQF) развёртывается с использованием контейнерной архитектуры. Пакет дистрибутива ADC DQF, предоставляемый заказчику, содержит комплект компонентов, необходимых для полного развертывания решения. Состав пакета приведён в таблице 1.

Функциональный блок	Компоненты продукта
Сервис исполнения правил (СИП)	DQF Engine Standalone
Сервис проверки правил (СПП)	DQF Engine Standalone (профиль probe)
Сервис расписаний задач (СРЗ)	DQF Scheduler
Сервис планирования задач (СПЗ)	DQF Planner
Сервис каталога правил (СКП)	DQF Complex Algorithms Storage Service
Сервис результатов	DQF Aggregator
Сервис дашбордов	DQF Dashboards Storage
Сервис отчетов	DQF Reporter
Веб-портал	DQF Web App
Сервис авторизации	DQF Auth Service
Адаптер результатов для модуля сохранения	DQF Saver Adapter

Таблица 1 Компоненты продукта DQF

Модули экосистемы Arenadata Catalog, которые дополняют ADC DQF и обеспечивают дополнительную функциональность, описаны в таблице 2.

Функциональный блок	Компоненты продукта
Модуль сохранения подробных результатов*	Result Saver
*обеспечивает функцию сохранения результатов на пользовательское s3 совместимое хранилище.	

Таблица 2 Компоненты экосистемы ADC

Инфраструктурные компоненты, необходимые для эксплуатации ADC DQF, описаны в таблице 3.

Функциональный блок	Компонент среды функционирования
Служебная база данных	Postgres 14 или любая совместимая с ней СУБД
Брокер сообщений	Kafka (предпочтительнее), RabbitMQ
Средство аутентификации	KeyCloak - 23.0
Перенаправление логов	Vector - 0.28.0
Хранение и поиск логов	Loki - 2.6.0, Grafana - 9.5.5
Сбор метрик	Prometheus - 2.41.0, Prometheus Graphite Exporter - 0.13.1
Единая точка входа	Nginx
Служба координации**	Zookeeper - 3.9
** используется для распространения и проверки лицензий	

Таблица 3 Дополнительные компоненты для среды функционирования DQF

Архитектура DQF

Возможна установка используя docker compose для контейнерной архитектуры или установка каждого компонента по отдельности.

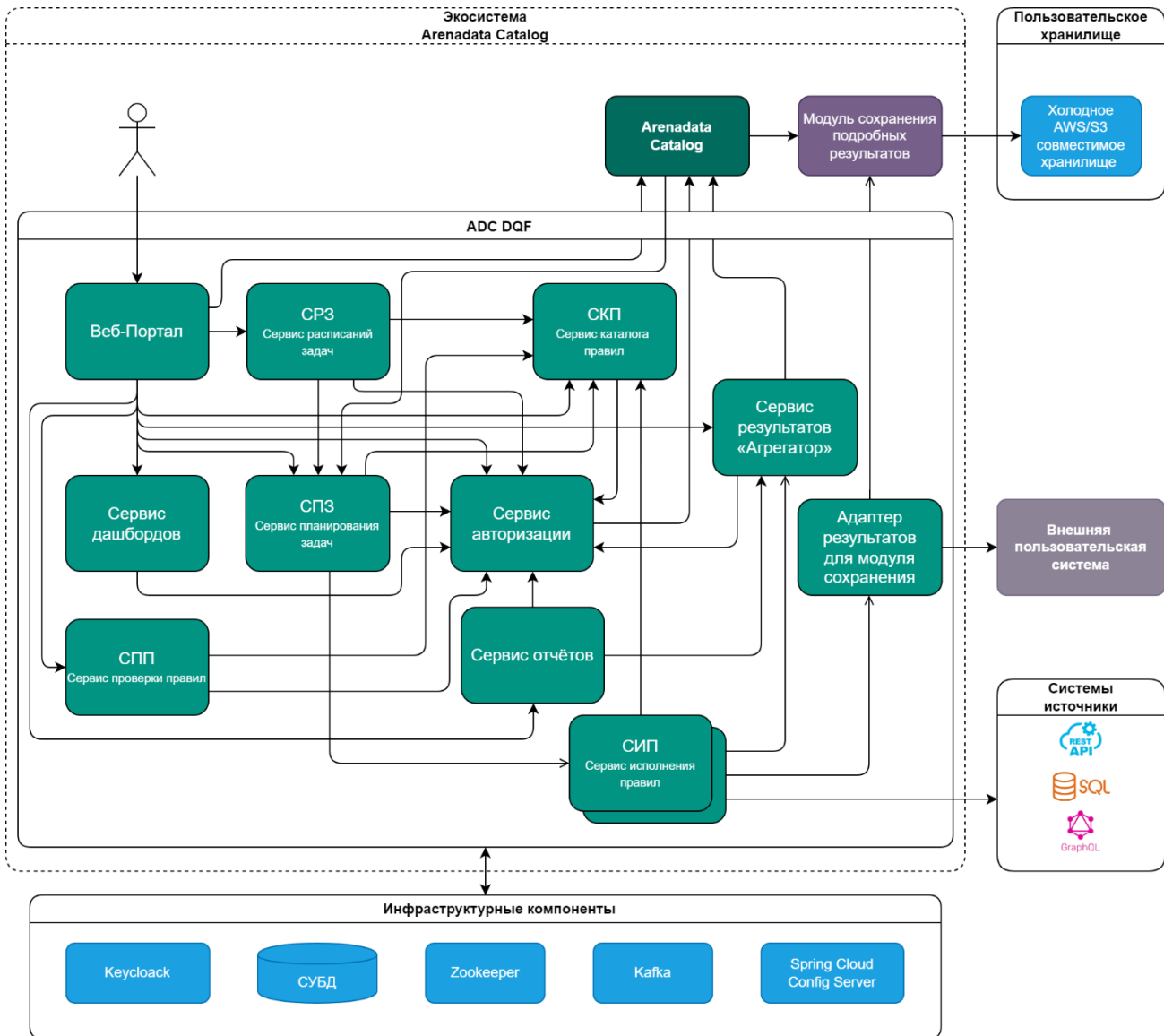


Рисунок 1 Высокоуровневая архитектура продукта DQF

Требования к вычислительной среде

PROD среда

Расчет требований к вычислительной среде для среды PROD готовится индивидуально в зависимости от многих факторов.

Минимальные требования к серверу приложения и компонентам продукта

- 16 ядер CPU x64, не менее 2 ГГц (рекомендуется платформа не старше Intel Cascade Lake);
- 16 Гб оперативной памяти;
- 14 Гб полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек);

Минимальные требования к серверу БД

- 8 ядер CPU x64, не менее 2 ГГц (рекомендуется платформа не старше Intel Cascade Lake);
- 16 Гб оперативной памяти;
- 50 Гб полезного дискового пространства SDD (с минимальной скоростью чтения 150МБ/сек);

TEST среда

Требования к серверу приложения и компонентам продукта

- 13 ядер CPU x64, не менее 2 ГГц (рекомендуется платформа не старше Intel Cascade Lake);
- 13 Гб оперативной памяти;
- 12 Гб полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек);

Требования к серверу БД

- 4 ядер CPU x64, не менее 2 ГГц (рекомендуется платформа не старше Intel Cascade Lake);
- 8 Гб оперативной памяти;
- 50 Гб полезного дискового пространства SDD (с минимальной скоростью чтения 150МБ/сек);

При необходимости проведения нагрузочного тестирования рекомендуется использовать вычислительную среду, идентичную по ресурсам PROD-среде.

DEV среда

Минимальные технические требования к вычислительной инфраструктуре для контура разработки:

Требования к серверу приложения и компонентам продукта

- 13 ядер CPU x64, не менее 2 Гц (рекомендуется платформа не старше Intel Cascade Lake);
- 13 Гб оперативной памяти;
- 12 Гб полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек);

Требования к серверу БД

- 4 ядер CPU x64, не менее 2 Гц (рекомендуется платформа не старше Intel Cascade Lake);
- 8 Гб оперативной памяти;
- 50 Гб полезного дискового пространства SDD (с минимальной скоростью чтения 150МБ/сек);

Установка компонентов на несколько серверов

Требования к серверам в случае распределённой установки зависят от многих факторов и являются предметом дополнительного обсуждения. В качестве рекомендаций используйте приведенные ниже характеристики для каждого компонента

Сервис исполнения правил

- 4 ядер, не менее 2 Гц;
- 4 Гб оперативной памяти;
- 2 Gb SSD;

Сервис расписаний задач

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Сервис планирования задач

- 2 ядра, не менее 2 Гц;
- 2 Гб оперативной памяти;
- 2 Gb SSD;

Сервис каталога правил

- 2 ядра, не менее 2 Гц;
- 2 Гб оперативной памяти;
- 2 Gb SSD;

Сервис результатов

- 2 ядра, не менее 2 Гц;
- 2 Гб оперативной памяти;
- 2 Gb SSD;

Сервис дашбордов

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Сервис отчетов

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Веб-портал

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Сервис авторизации:

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Адаптер результатов для модуля сохранения:

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Модуль сохранения подробных результатов*:

- 2 ядра, не менее 2 Гц;
- 2 Гб оперативной памяти;
- 2 Gb SSD;

PostgreSQL:

- 8 ядер, не менее 2 Гц;
- 16 Гб оперативной памяти;
- 50 Gb SSD;

Zookeeper:

- 1 ядро, не менее 2 Гц;
- 1 Гб оперативной памяти;
- 1 Gb SSD;

Kafka:

- 4 ядра, не менее 2 Гц;
- 8 Гб оперативной памяти;
- 100 Gb SSD;

Nginx:

- 2 ядра, не менее 2 Гц;
- 4 Гб оперативной памяти;
- 10 Gb SSD;

Требования к программному обеспечению конечного пользователя

Весь предлагаемый функционал продукта доступен пользователю через Web-браузер. Поддерживаемые дистрибутивы и версии браузеров приведены в таблице 4.

Браузер	Версия
Opera	74.0 или выше
Google Chrome	88.0 или выше
Microsoft Edge	88.0 или выше
Firefox	85.0 или выше
Safari	14.0.2 или выше

Таблица 4 Поддерживаемые браузеры

Требования к среде резервного копирования

Требования к необходимым вычислительным ресурсам для резервного копирования и восстановления варьируются в зависимости от различных факторов, таких как политики выполнения резервного копирования, наличие инструментов для выполнения операций, требования к восстановлению данных (RPO, RTO) и других факторов. Требования к среде резервного копирования являются предметом дополнительного обсуждения.

Документ предоставлен исключительно в информационных целях, и его содержание может быть изменено без предварительного уведомления.

Все торговые марки, знаки обслуживания, зарегистрированные торговые знаки или зарегистрированные знаки обслуживания являются собственностью их законных владельцев.