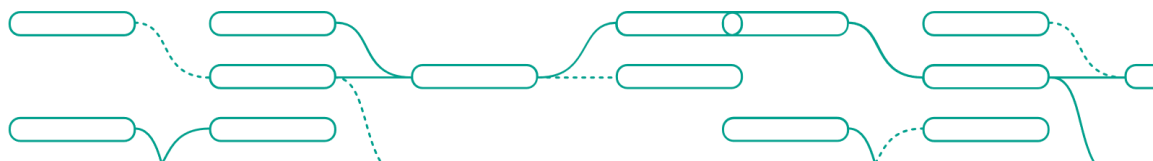


Описание технической архитектуры программного обеспечения
Arena Data Catalog Data Quality Framework
(ADC.DQF)

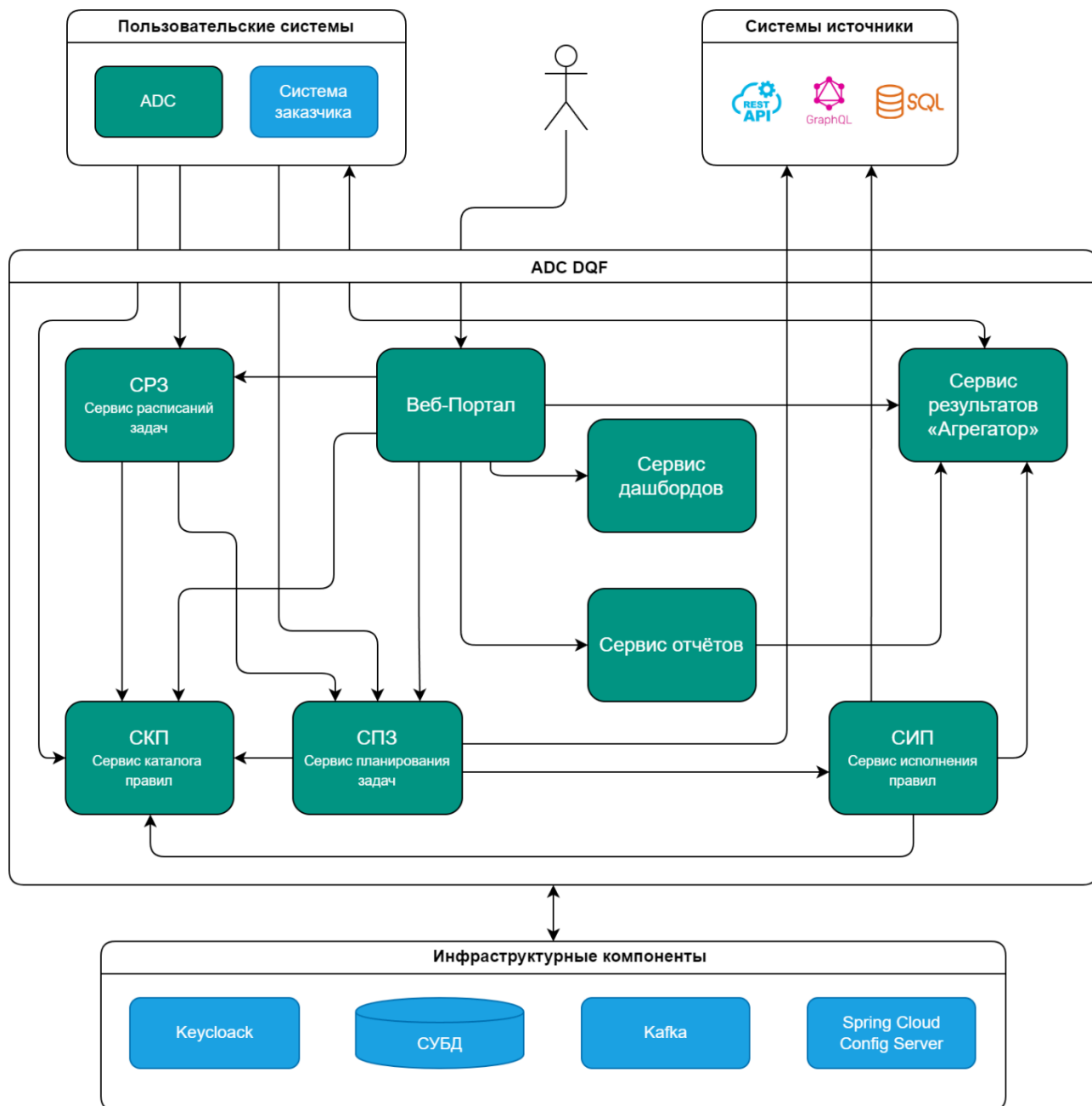
Москва
2025



Оглавление

1	Архитектурная схема.....	3
2	Описание сервисов DQF	4
2.1	Веб-портал	4
2.2	Сервис расписаний задач.....	4
2.3	Сервис каталога правил	5
2.4	Сервис планирования задач	5
2.5	Сервис отчётов	6
2.6	Сервис дашбордов.....	6
2.7	Сервис исполнения правил	6
2.8	Сервис результатов	7
3	Описание инфраструктурных сервисов	8
3.1	Keycloak	8
3.2	СУБД.....	8
3.3	Kafka	8
3.4	Spring Cloud Config Server	8

1 Архитектурная схема



2 Описание сервисов DQF

ADC.DQF представляет собой комплекс программ, выполненный на микросервисной архитектуре. Сервисы DQF представляют собой JAVA-приложения, разработанные с использованием фреймворка Spring. Веб-портал представляет собой браузерное приложение, выполненное на typescript. Сервисы с REST API проверяют права доступа пользователей и систем, используя JWT-токены, полученные от Keycloak.

2.1 Веб-портал

Веб-портал — это модуль ПО DQF, предназначенный для предоставления пользовательского интерфейса пользователю DQF.

Веб-портал реализует следующий набор функций:

1. Предоставление пользовательского интерфейса **редактора правил** для создания и изменения правил.
2. Предоставление пользовательского интерфейса **каталога правил** для отображения правил в таблице с возможностью сортировки и фильтрации, а также создания, редактирования, удаления правил и объединения их в группы по бизнес-потребностям.
3. Предоставление пользовательского интерфейса **планирования задач**, для отображения в таблице задач с возможностью сортировки и фильтрации, а также создания задач по группе или набору правил.
4. Предоставление пользовательского интерфейса **расписаний задач**, для отображения в таблице расписаний с возможностью сортировки и фильтрации, а также создания, редактирования и удаления расписаний выполнения проверок по группе правил.
5. Предоставление пользовательского интерфейса **дашбордов** для отображения и редактирования виджетов с результатами выполнения задач по исследуемым параметрам в форме линейных графиков, столбцов или карточек с числом, соответствующим исследуемому численному показателю задач
6. Предоставление пользовательского интерфейса **отчётов**, для отображения в таблице отчётов по задачам с возможностью сортировки и фильтрации по типам для их последующей загрузки в формате эксель, а также создания отчётов по типам: детальный отчёт, сводный отчёт,

2.2 Сервис расписаний задач

Сервис расписаний задач — это сервис ADC.DQF, предназначенный для управления расписанием задач.

Сервис расписаний задач реализует следующий набор функций:

1. Создание расписаний для автоматического формирования задач на проверку данных с использованием групп правил.

2. Просмотр свойств расписания по его идентификатору.
3. Просмотр связанных задач на проверку данных, созданных согласно установленному расписанию.
4. Просмотр списка всех расписаний.
5. Редактирование свойств расписания.
6. Деактивация расписаний.

2.3 Сервис каталога правил

Сервис каталога правил (СКП) – это сервис ADC.DQF, предназначенный для хранения и управления правилами проверок и их группами.

Сервис каталога правил (СКП) реализует следующий набор функций:

1. Управление реестром правил: создание новых правил для проверки данных, удаление, редактирование и их просмотр;
2. Управление версиями правил, обеспечивающая хранение предыдущие версии правил;
3. Управления реестром групп: создание, удаление, редактирование и просмотр новой группы для объединения правил, которые проверяют данные одного типа или основаны на выполнении одного бизнес-правила (и другие причины объединения),;
4. Управление версиями групп, позволяющая изменять состав группы без потери истории изменений.

2.4 Сервис планирования задач

Сервис планирования задач (СПЗ) — это сервис ADC.DQF, предназначенный для

- Создания задачи на проверку набора данных
- Отправки запросов на проверку качества данных в СИП
- Подготовки набора данных на проверку, в том числе из внешних источников

Сервис планирования задач (СПЗ) реализует следующий набор функций:

1. Создание задачи на проверку данных с использованием набора правил или указания группы правил.
2. Просмотр статуса задачи в текущий момент времени.
3. Просмотр всех задач на проверку данных.
4. Принудительная остановка выполнения задачи.
5. Формирование и передача запросов на проверку качества данных, содержащих данные проверяемого объекта и идентификатор правила.
6. Настройка проверки либо для всего входного массива исходных данных, либо отдельно для каждого элемента массива.

2.5 Сервис отчётов

Сервис отчётов — это сервис ADC.DQF, предназначенный для создания и передачи пользователю отчётов о результатах задач в формате Excel (xlsx).

Сервис отчётов реализует следующий набор функций:

1. Обращение в сервис результатов для получения результатов выполнения задачи
2. Формирование отчёта в формате excel (xlsx)
3. Отправку отчёта пользователю.

2.6 Сервис дашбордов

Сервис дашбордов — это сервис ADC.DQF, предназначенный для создания, хранения изменения и передачи пользователю конфигурации дашбордов.

2.7 Сервис исполнения правил

Сервис исполнения правил (СИП) – это сервис ADC.DQF, предназначенный для исполнения правил на исходных данных.

Сервис исполнения правил (СИП) реализует следующий набор функций:

1. Получение входного сообщения с данными для проверки и идентификатором правила в соответствии с согласованным программным интерфейсом.
2. Получение декларативного описания правила из сервиса каталога правил
3. Исполнение правил, заданных в виде цикла или последовательности шагов, включающих вызовы базовых алгоритмов и/или механизмов управления потоком.
4. Исполнение базовых алгоритмов, таких как:
 - HTTP/REST запросы.
 - JDBC запросы.
 - Проверка и сравнение атрибутов,
 - Проверка идентификаторов – валидация ИНН, СНИЛС, ОГРН.
 - Нативное подключение к продукту Фактор HFL для проверок ФИО и адресов.
 - Присвоение значений атрибутам, включая математические операции.
 - Сбор и фильтрация данных.
 - Добавление атрибута в список.
5. Исполнение алгоритма управления потоком:
 - Последовательное исполнение.
 - Итерирование по коллекции.

- Выполнение альтернативных сценариев по условию.
 - Выполнение цикла пока выполняется условие.
6. Управление контекстом и результатами
- Формирование и передача между правилами контекста, содержащего данные и промежуточные результаты выполнения шагов.
 - Создание сообщений-ответов с результатами выполнения правила и их асинхронная передача во внутреннюю очередь брокера сообщений.

2.8 Сервис результатов

Сервис результатов (CP) – это сервис ADC.DQF, предназначенный для получения, хранения и предоставления информации о результатах выполнения задач и правил проверок.

Сервис результатов (CP) реализует следующий набор функций:

1. Получение и хранение данных о результатах выполнения правил.
2. Формирование списка результатов всех выполненных правил задачи. Результат содержит статус выполнения правила и дополнительную информацию по исполнению данного правила.
3. Формирование статистики выполнения задачи.
4. Формирование суммарной статистики выполненных задач за период с возможностью фильтрации.
5. Формирование детализированной статистики выполненных задач за период с заданным шагом и возможностью фильтрации.

3 Описание инфраструктурных сервисов

Сервисы, которые требуются для работы ПО

3.1 Keycloak

Система управления идентификацией и доступом (IAM).
Обеспечивает аутентификацию и авторизацию пользователей и прикладных систем.
Используется для выдачи и проверки JWT-токенов при взаимодействии сервисов по протоколу REST.

3.2 СУБД

Реляционная система управления базами данных.
Предназначена для хранения структурированных данных прикладных сервисов.
Обеспечивает транзакционность, целостность и долговременное хранение информации.

3.3 Kafka

Распределённая стриминговая платформа.
Используется в качестве транспорта сообщений между сервисами.
Обеспечивает надёжную доставку, горизонтальное масштабирование и высокую производительность при обмене данными.

3.4 Spring Cloud Config Server

Сервер централизованного управления конфигурацией.
Обеспечивает хранение и распространение параметров настройки прикладных сервисов из единого источника.
Использование необязательно, применяется для упрощения администрирования и обеспечения консистентности конфигураций.