Arenadata Catalog

Высокоуровневая архитектура развертывания Arenadata Catalog

Москва 2024

arenadc.io

Оглавление

Архитектура Arenadata Catalog **3**

Требования к вычислительной среде **4**

PROD среда 4

TEST среда 4

DEV среда 4

Развертывание Arenadata Catalog для среды демонстрации, апробации 4

Установка компонентов на несколько серверов, в том числе используя контейнерную архитектуру 5

Требования к программному обеспечению пользователя **6**

Требования к среде резервного копирования **6**

Программный продукт Arenadata Catalog обеспечивает возможность развертывания на различных платформах, включая контейнеры, физические и виртуальные серверы архитектуры x86. Однако, с использованием контейнерной архитектуры развертывание системы является наиболее эффективным методом для современных организаций. Предоставляемый клиентам пакет дистрибутива Arenadata Catalog включает набор компонентов необходимый для полноценного развертывания решения, таблица 1.

Функциональный блок	Компонент платформы
Сервер приложения Http|API	Arenadata DC Platform
Служебная база данных	PostrgreSQL
Поисковая система	ElasticSearch | OpenSearch
Сервис захвата метаданных	Arenadata DC Ingestion framework Airflow
Сервис управления рабочими процессами Workflow	Camunda 7
Перенаправление логов	Vector
Хранение и поиск логов	Loki, Grafana
Балансировщик нагрузки (опционально)	Nginx
Средство аутентификации (опционально)	KeyCloak

Таблица 1 Компоненты платформы Arenadata Catalog

Архитектура Arenadata Catalog

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

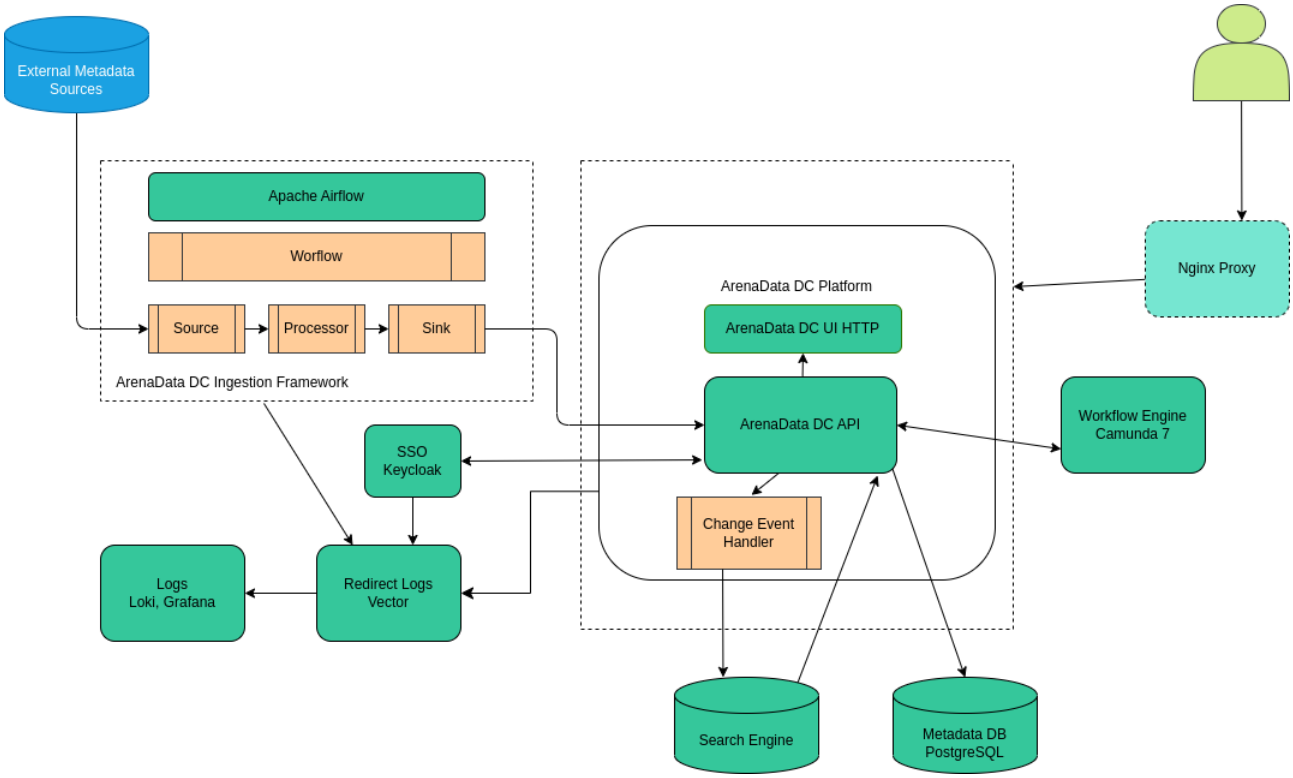

Рисунок 1 Высокоуровневая архитектура платформы Arenadata Catalog

Требования к вычислительной среде

PROD среда

При инсталляции Arenadata Catalog для среды PROD рекомендовано использовать внешнюю БД.
Требования к серверу приложения и компонентам платформы

- 32 CPU x86, не менее 2 Ггц для сервера ADC и его компонентов;
- 64 Гб оперативной памяти для сервера ADC и его компонентов;
- 300Gb полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек) для сервера ADC;

Требования к серверу БД

- 6 CPU x86, не менее 2 Ггц для сервера БД Postgres;
- 16 Гб оперативной памяти для внешней БД Postgres;
- 500 Гб полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек) для сервера ADC для внешней БД;
- Рекомендован Postgres v14.6 и выше

TEST среда

При инсталляции Arenadata Catalog для среды TEST допускается развертывание всех компонентов на одном сервере. Требования к серверной инфраструктуре для тестового контура:

- 24 CPU x86, не менее 2 Ггц для сервера ADC, включая внутреннюю контейнеризованную БД Postgres;
- 24 Гб оперативной памяти;
- 500Гб полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек);

DEV среда

Минимальные технические требования к вычислительной инфраструктуре для контура разработки:

- 12 CPU x86, не менее 2 Ггц, для сервера ADC, включая внутреннюю контейнеризованную БД Postgres;
- 24 Гб оперативной памяти;
- 200 Гб полезного дискового пространства HDD/SDD (с минимальной скоростью чтения 150МБ/сек);

Развертывание Arenadata Catalog для среды демонстрации, апробации

Возможна установка на один сервер, соответствующий минимальным техническим требованиям:

- 8 CPU x86, не менее 2 Ггц, для сервера ADC, включая внутреннюю контейнеризованную БД Postgres;
- 16 Гб оперативной памяти;
- 200 Гб полезного дискового пространства HDD/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 промышленной среды PROD каждого компонента

Arenadata DC Platform
- 8 ядер, не менее 2 Ггц;
- 16 Гб оперативной памяти;
- 200 Гб полезного дискового пространства (с минимальной скоростью чтения 150МБ/сек);

Camunda 7
- 4 ядра, не менее 2 Ггц;
- 4 Гб оперативной памяти;
- 100 Gb HDD/SSD;

Elastic search | Open Search
- 8 ядер, не менее 2 Ггц;
- 32 Гб оперативной памяти;
- 300 Gb HDD/SSD;

Airflow ArenaData DC Ingestion Framework
- 4 ядра, не менее 2 Ггц;
- 8 Гб оперативной памяти;
- 100 Gb HDD/SSD;

PostgreSQL
- 6 ядер, не менее 2 Ггц;
- 16 Гб оперативной памяти;
- 500 Gb HDD/SSD;
- Рекомендован Postgres v14.6 и выше

Loki
- 2 ядра, не менее 2 Ггц;
- 4 Гб оперативной памяти;
- 100 Gb HDD/SSD;

Vector
- 4 ядра, не менее 2 Ггц;
- 8 Гб оперативной памяти;
- 50 Gb HDD/SSD;

Grafana
- 4 ядра, не менее 2 Ггц;
- 8 Гб оперативной памяти;
- 100 Gb HDD/SSD;

Nginx (опционально)
- 2 ядра, не менее 2 Ггц;
- 4 Гб оперативной памяти;
- 10 Gb HDD/SSD;

KeyCloak (опционально)
- 2 ядра, не менее 2 Ггц;
- 4 Гб оперативной памяти;
- 20 Gb HDD/SSD;

Требования к программному обеспечению пользователя

Весь предлагаемый функционал приложения доступен пользователю через Web-браузер. Поддерживаемые дистрибутивы и версии браузеров приведены в таблице 2.

Браузер	Версия
Yandex Browser	20.3.0.1223 или выше
Opera	56.0.3051.99 или выше
Google Chrome	70.0.3538.67 или выше
Microsoft Edge	80.0 или выше

Таблица 2 Поддерживаемые браузеры

Требования к среде резервного копирования

Требования к необходимым вычислительному 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. В дистрибутиве Arenadata Catalog по умолчанию предоставляется скрипт, который выполняет копирование базы данных и сохраняет полные копии базы данных в течение последних трех дней. Следующая спецификация ресурсов хранения будет достаточна для обеспечения необходимой глубины хранения РК, таблица 3.

Компонент	Регламент РК	Спецификация ресурсов для хранения бэкапов
БД (дамп)	глубина хранения 3 дня, три полных бэкапа	объем дискового пространства сервера БД x3
Приложение ADC для среды PROD (снимок диска)	глубиной хранения 2 недели (14 точек восстановления), 1 раз в неделю выполняется полный бэкап, остальные 6 дней инкрементальный	объем дискового пространства сервера ADC x5
БД для среды PROD (снимок диска)	глубина хранения 2 недели (14 точек восстановления), 1 раз в неделю выполняется полный бэкап, остальные 6 дней инкрементальный	объем дискового пространства сервера БД x5
Приложение и БД ADC для сред TEST/DEV (снимок диска)	глубина хранения - 3 дня (пн, ср, пт), в 1 день выполняется полное бэкапирование, в остальные 2 инкрементальное	объем дискового пространства сервера ADC x3

Таблица 3 Политики бэкапирования

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.