

Arenadata Catalog (ADC)TM

Инструкция по установке программного обеспечения

Москва
2023

Журнал изменений

Дата	Версия	Комментарий
25.10.2022	0.1	Начальная версия документа
16.01.2023	0.2	Актуализация с точки зрения инфраструктуры и способа поставки
25.01.2023	0.3	Добавление способа развертывания с учетом внешней БД
01.02.2023	0.4	Актуализация способа развертывания с учетом релиза v 0.2

Содержание

Журнал изменений	2
1 Введение	4
1.1 Термины и определения	4
1.2 Сокращения и обозначения	4
1.3 Общие положения	4
2 Назначение ПО	5
3 Требования к установке	6
4 Подготовка к установке	6
5 Установка ПО	7
6 Проверка доступности сервиса	8
7 Контакты технических специалистов	8

1 Введение

1.1 Термины и определения

Термин	Значение
База данных	Совокупность данных, хранимых в соответствии со схемой, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных
Бизнес-глоссарий	Словарь для бизнес-пользователей. Словарь состоит из бизнес-терминов, которые могут быть связаны друг с другом, и позволяет распределять их по предметным областям, чтобы их можно было понимать в разных контекстах
SQL	Декларативный язык программирования, применяемый для создания, модификации и управления данными в реляционной базе данных, управляемой соответствующей СУБД
Data Lineage	Информация, которая описывает движение данных от источника их происхождения по точкам обработки и применения
Keycloak	Инструмент с открытым исходным кодом, используемый для аутентификации пользователей в организации, с возможностью настройки single sign-on и управления доступом

1.2 Сокращения и обозначения

Сокращение	Наименование
ПО	Программное обеспечение
СУБД	Система управления базами данных
БД	База данных
AD.C	Arenadata Catalog
УЗ	Учетная запись
ТП	Техническая поддержка

1.3 Общие положения

Инструкция предназначена для должностных лиц, осуществляющих установку программного обеспечения (ПО) Arenadata Catalog (ADC). Документ содержит пошаговое описание действий для установки специального программного обеспечения системы.

2 Назначение ПО

Основным предназначением ПО ADC является загрузка метаданных из различных систем обработки и анализа данных по всей компании с ведением и тесной интеграцией корпоративного Бизнес-Глоссария.

Программное обеспечение Arenadata Catalog обладает следующими характеристиками:

- Автоматический сбор метаданных из различных источников, включая возможность профилирования и получения примеров данных;
- Установка критериев качества данных и назначение проверок;
- Обогащение метаданных описанием и указанием владельцев;
- Визуализация происхождения данных — data lineage;
- Создания гибко настраиваемого Бизнес-глоссария;
- Управление объектами каталога метаданных и Бизнес-глоссария посредством настраиваемых рабочих процессов;
- Связывание объектов метаданных и Бизнес-глоссария;
- Полнотекстовый поиск информации в ADC.

3 Требования к установке

Для установки ADC по данной инструкции необходимо выполнение следующих требований:

1. Операционная система– Ubuntu 20.04.4 LTS, CentOS 7.9.2009.
2. ПО– Docker version 20.10.17, build 100с701.
3. ПО Docker compose version 1.29.2.
4. Пользователь с правами sudo.

Следует проверить версию установленного Docker Compose на соответствие требованиям.

В неподходящей версии Docker compose сборка не произойдет, так как используется версия контейнера 3.9.

Можно воспользоваться командой:

```
docker-compose -v
```

В случае отсутствия необходимо произвести установку согласно инструкции по [ссылке](#).

4 Подготовка к установке

Необходимо выполнить команду аутентификации на сервере репозитория, используя реквизиты полученные в технической поддержке:

```
docker login repo.arenadc.io -u login -p password
```

Далее необходимо загрузить файл Docker compose и environment файлы по [ссылке предоставленной технической поддержкой](#). В environment файлах:

1. data-sec.json – отвечает за наполнение realm хранилища, Keycloak модуля аутентификации(заполнять при использовании docker-compose-keycloak.yml), необходимо заполнить поля связанные с host(искать - 10.80.0.46/postgresql) ,
2. adc.env – отвечает за параметризацию переменных docker-compose-keycloak.yml необходимо заменить поля связанные с host (искать - 10.80.0.46/postgresql),
3. adc_noauth.env – отвечает за параметризацию переменных docker-compose.yml необходимо заменить поля связанные с host (искать - 10.80.0.46/postgresql);

Указанные выше файлы необходимы для параметризации Docker compose и интеграции Keycloak с Data Catalog. Есть альтернативный способ настройки интеграции через web-интерфейс Keycloak, который находится на порту 8081. Также есть облегченная сборка compose без использования Keycloak. А параметризация нужна в первую очередь для порядка в переменных среды и развертывания и избежания хардкода в Docker compose файлах.

5 Установка ПО

Для работы приложения рекомендуется использовать директорию “/opt/adc/”.

Загруженную конфигурацию для docker compose поместить в директорию “/opt/adc/” и перейти в эту директорию.

```
cd /opt/adc/
```

Чтобы запустить ПО выполните в консоли одну из следующих команд:

```
sudo docker compose -f docker-compose.yml --env-file adc_noauth.env up -d #basic authentication
#or
sudo docker compose -f docker-compose-keycloak.yml --env-file adc.env up -d #keycloak authentication
```

Необходимо ожидание инициализации базы данных в пределах 10 минут.

В случае использования внешней БД PostgreSQL в выбранном из вышеперечисленных environment – файлов надо переопределить значения блока подключения к БД, а именно:

```
#Database configuration for server container
SERVER_DB_DRIVER_CLASS=org.postgresql.Driver
SERVER_DB_SCHEME=postgresql
SERVER_DB_USE_SSL=false
SERVER_DB_USER=openmetadata_user
SERVER_DB_USER_PASSWORD=openmetadata_password
SERVER_DB_HOST=postgresql
SERVER_DB_PORT:-5432
SERVER_DB_DATABASE=openmetadata_db
```

Перед установкой системы необходимо создать сущности в БД, выполнив int.sql скрипт, что поставляется в комплекте с docker compose environment. Использовать скрипты необходимо только при условии выделенной PostgreSQL, в случае использования встроенного контейнера БД, миграции проводить необходимости нет.

Примечание: в случае необходимости выполнения миграций по БД нужно учитывать факт наличия volume, и при наличии использовать команды: docker cp, docker exec, psql >.

Для контейнера БД также предусмотрено автоматизированное бекапирование с ротацией дампов:

```
cron_backup.sh #add/delete job with backup_pg.sh to cron
backup_pg.sh #create pg_dump with timestamp and rotate
```

6 Проверка доступности сервиса

Понять о состоянии сервисов можно посредством команд:

```
docker ps -a
docker compose ps -a
docker compose logs
```

После инициализации БД запускаем Arenadata Catalog в web-браузере по адресу:

```
http://\[server address\]:8585
```

В случае технической необходимости остановить сервисы и/или переустановить сервисы используем следующие команды:

```
cd /opt/adc #go to working dir
docker compose "chosen env files" stop #stop all containers
docker compose "chosen env files" rm #stop all containers
docker volume prune #clear all volumes
docker container prune #clear all stopped containers
```

7 Контакты технических специалистов

В случае возникновения трудностей при установке программного обеспечения, свяжитесь с технической поддержкой, используя электронный адрес info@arenadc.io.