

Программное обеспечение
Система управления данными Polyflow
РУКОВОДСТВО АДМИНИСТРАТОРА

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата

Содержание

1	Назначение и условия применения.....	8
1.1	Назначение системы.....	8
1.2	Условия применения.....	8
1.2.1	Серверная часть.....	8
1.2.2	Локальная сеть.....	9
2	Подготовка к работе.....	10
2.1	Состав программного обеспечения	10
2.1.1	Установка и настройка Docker.....	10
2.1.2	Установка компонентов на один сервер	11
2.1.3	Описание утилиты управления сервисом Polyflow manage.py.....	11
2.1.4	Описание файлов и структуры проекта	12
2.1.5	Поддержка локализации	13
2.1.6	Настройки в airflow.cfg.....	14
2.1.7	Требования к СУБД хранилища	14
2.2	Порядок проверки работоспособности	15
2.3	Порядок обновления системы.....	16
3	Описание операций администрирования	17
3.1	Определения и сокращения Polyflow	17
3.2	Настройки компонентов платформы.....	18
3.2.1	Общие настройки компонентов платформы	18
3.2.2	Ограничение ресурсов для контейнера.....	18
3.2.3	Настройка схем авторизации	20
3.3	Операции группы безопасность.....	25
3.3.1	Список пользователей	25
3.3.2	Список ролей	26
3.3.3	Статистика пользователя.....	28
3.3.4	Основные разрешения	29
3.3.5	Отображения/меню	29
3.3.6	Право на просмотр/меню.....	30
3.4	Операции группы Admin	31
3.4.1	Variables	31
3.4.2	Configuration	32
3.4.3	Connections.....	32

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата					

.РЭ

3.4.4 Plugins.....	33
3.4.5 Providers	34
3.4.6 Pools.....	35
3.4.7 XComs.....	35
3.4.8 DAGs Code Editor	36
4 Аварийные ситуации	38
5 Рекомендации по освоению	39

Инв. № подл.					Подпись и дата								
							Взам. инв. №	Инв. № дубл.	Подпись и дата				
											Инв. № дубл.	Подпись и дата	
Изм	Лист	№ докум.	Подп.	Дата	.РЭ		Лист 4						

1 Назначение и условия применения

1.1 Назначение системы

Модуль Polyflow предназначен для решения следующих задач:

- сбор данных из разных источников:
 - файлы
 - внешние системы
 - базы данных
- описание сущностей-получателей и источников;
- мониторинг и управление выполняемыми процессами;
- контроль качества данных;
- трансформация данных.

1.2 Условия применения

Для функционирования АИС необходимо следующее программно-аппаратное обеспечение:

1.2.1 Серверная часть

Минимальные требования к серверному оборудованию следующие:

- CPU 2 vCPU (2.8 ГГц и выше);
- RAM 8 ГБ,
- HDD 10 ГБ

Ориентировочная формула для подсчета конфигурации в зависимости от количества процессов при использовании Local Executor: дополнительно к минимальным системным требованиям необходимо RAM 256-512МБ CPU 0.1 vCPU в среднем на каждый процесс.

Операционная система: Astra Linux Special Edition 1.6 (Смоленск) или аналог.

Права пользователя, разворачивающего приложение: user - non-root with sudo privileges.

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

8

Дополнительные требования к установленным приложениям: Docker версии 20.10.0 и до 25, Docker -compose версия 1.29 и выше.

1.2.2 Локальная сеть

Все компоненты платформы должны находиться в одной подсети или должна обеспечиваться прозрачная маршрутизация. Не рекомендуется использовать NAT. В рамках ознакомления рекомендуется отключить брандмауэры. Внутри локальной сети между всеми компонентами не должно быть ограничений по передаче данных. Для доступа из внешней сети достаточно открыть порт, используемый Polyflow (порт задается при установке). При использовании системы с установленными антивирусами или комплексными системами защиты необходимо обеспечить свободную работу, сетевую активность и взаимодействие компонентов.

Инв. № подл.	Подпись и дата				Инв. № дубл.	Подпись и дата				Взам. инв. №	Инв. № дубл.				Инв. № подл.	Подпись и дата			
Изм	Лист	№ докум.	Подп.	Дата	.РЭ										Лист				
															9				

2 Подготовка к работе

2.1 Состав программного обеспечения

Модуль Polyflow поставляется в виде нескольких файлов:

- образы Docker, содержащие в себе все компоненты с уже настроенным окружением и всеми внутренними зависимостями;
- python-файл manage.py для быстрого развёртывания и настройки модуля;
- дополнительные файлы, используемые при разворачивании: `__init__.py`, `docker-compose.yml.tpl`, `airflow.cfg.tpl`.

Примечание: Docker — программное обеспечение для автоматизации развёртывания и управления приложениями в среде виртуализации на уровне операционной системы. Суть и одно из предназначений Docker такое же, как и у виртуальных машин — это изоляция работы различных конфликтующих программ внутри одного сервера. Наглядно увидеть отличия между виртуальной машиной и контейнером можно, пройдя по ссылке: <https://www.docker.com/what-docker>.

Таким образом, установка сводится к двум шагам:

1. Установка docker-engine внутри операционной системы;
2. Запуск python-файла manage.py для разворачивания готовых контейнеров с компонентами системы.

Настройка и разработка процессов загрузки и обработки данных описаны в документе «Руководство разработчика». Основные операции с интерфейсом описаны в документе «Руководство пользователя».

2.1.1 Установка и настройка Docker

1. Установить Docker в соответствии с инструкцией:

<https://docs.docker.com/install/linux/docker-ce/ubuntu/>.

Примечание - Рекомендуемая версия 20.10.12.

2. Установить Docker Compose в соответствии с инструкцией:

<https://docs.docker.com/compose/install/>.

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

10

Примечание - Рекомендуемая версия 1.29.2.

2.1.2 Установка компонентов на один сервер

1. Создать директорию для файлов образов

```
cd ~ && mkdir install
```

2. Скопировать из полученного дистрибутива в созданную директорию архивы базовых образов

- df_operator .tar
- df_airflow .tar
- df_postgres13 .tar

3. ВЫПОЛНИТЬ КОМАНДЫ:

```
docker load -i ~/install/df_operator.tar  
docker load -i ~/install/df_airflow.tar  
docker load -i ~/install/df_postgres13.tar
```

4. Создать директории сервиса

```
cd ~ && mkdir dataflow && cd dataflow
```

5. Скопировать из полученного дистрибутива в созданную директорию инициализирующие компоненты

- manage.py
- __init__.py
- docker-compose.yml.tpl
- airflow.cfg.tpl

6. Разместить в созданной директории файлы ключей (key.pem) и сертификатов (cert.pem) (путь до файлов может быть задан при установке).

7. Развернуть сервис

```
python3 manage.py --deploy
```

2.1.3 Описание утилиты управления сервисом Polyflow manage.py

Автоматизирует процесс установки, обновления и управления Polyflow.

Принимает следующие аргументы:

Подпись и дата
Инв. № дубл.
Взам. инв. №
Подпись и дата
Инв. № подл.

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

11

- `-i` или `--interactive` интерактивный режим обновления;
- `--reconfigure` сгенерировать файлы конфигураций;
- `--deploy` развернуть сервисы с нуля;
- `--update` обновить сервисы;
- `--compile` подготовка структуры артефактов;
- `--backup` бэкап бд и конфигурации;
- `--restore` восстановление бд и конфигурацию;
- `--init-project` создание структуры проекта;
- `--no-rebuild` не пересобирать локальные образы;
- `--no-restart` не запускать (перезапускать) сервисы;
- `--no-upgrade` не обновлять базу данных.

Примеры запуска:

- `python3 manage.py --backup --no-restart --no-rebuild --no-upgrade` создать бэкап базы и конфигурации сервиса (при обновлении);
- `python3 manage.py --deploy` развернуть сервис;
- `python3 manage.py --no-restart --no-upgrade` пересобирать образы.

2.1.4 Описание файлов и структуры проекта

Вся работа происходит в папке `projects`, содержимое которого линкуется с папкой `workspace`.

Для инициализации структуры проекта необходимо выполнить команду `python3 manage.py --init-project <Имя проекта>`. Если имя проекта не задано, то будет использовано значение `default`. После выполнения команды будут созданы следующие папки:

- `dataflow/volumes/projects/<Имя проекта>/cache`
- `dataflow/volumes/projects/<Имя проекта>/dags`
- `dataflow/volumes/projects/<Имя проекта>/metadata`
- `dataflow/volumes/projects/<Имя проекта>/plugins`
- `dataflow/volumes/projects/<Имя проекта>/share`

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата

.РЭ

Лист

12

- dataflow/volumes/projects/<Имя проекта>/source

Дополнительно после установки сервиса доступны следующие конфигурационные файлы:

- airflow.env - переменные окружения, используемые airflow;
- operator.env - переменные окружения, используемые в operator (например, схемы БД);
- airflow.cfg - конфигурация планировщика и веб-интерфейса.

Тома:

- /home/osuser/docker/dataflow/volumes/projects - том df_projects;
- /home/osuser/docker/dataflow/volumes/workspace - том df_workspace;
- /home/osuser/docker/dataflow/volumes/data - том df_data;
- /home/osuser/docker/dataflow/volumes/logs - том df_logs.

Прикладная разработка и оператор работают в df_projects, airflow работает с df_workspace.

2.1.5 Поддержка локализации

В приложении реализована интернационализация и поддержан русский язык. Для переключения языка в airflow.cfg для настройки babel_default_locale необходимо указать значение ru.

Поддерживаются пользовательские словари перевода. Для их подключения в локальном Dockerfile airflow необходимо скопировать папку с переводами в соответствующую папку сервиса. Пример команды: `COPY --chown=airflow:airflow translations home/airflow/.local/lib/python3.9/site-packages/airflow/www/translations.`

Папка с переводами должна иметь следующую структуру translations/<locale>/LC_MESSAGES/messages.po, где:

- <locale> - код локали, например: ru, zh, en;

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

- messages.po - файл перевода, который необходимо создать на основе шаблона (можно получить командой `docker cp df-webserver:/home/airflow/.local/lib/python3.9/site-packages/airflow/www/translations/messages.pot ./messages.pot`).

2.1.6 Настройки в airflow.cfg

При разворачивании приложения на тестовых серверах допустимо использовать http. В этом случае настройку `cookie_secure` можно установить в `False`.

2.1.7 Требования к СУБД хранилища

Поддерживаются СУБД PostgreSQL 9.6+ и MS SQL Server 2008 R2 SP3 10.50.6560.0+.

Для корректной работы модулей системные таблицы рекомендуется создавать в отдельной схеме `dwh`, установленной по умолчанию для пользователя `dwh`.

Создание и заполнение системных таблиц выполняется последовательным запуском DAG'ов `df_upgrade_datadb` и `df_populate_datadb`.

Для работы необходимо добавить объект `Connection`:

```
Conn Id: df
Conn Type: <Необходимый тип базы данных хранилища>
Host: <Хост базы данных>
Login: <Логин>
Password: <Пароль>
Extra: {"schema": "<Схема приложения>"}
```

Переопределить наименование схемы `df` возможно через переменную `'DATAFLOW_SCHEMA'`.

Если пользователь базы хранилища отличается от пользователя `df`, ему требуется явная выдача разрешений на системные таблицы, пример запроса:

```
grant select, insert, update, delete on all tables in schema df to dwh;
grant select, usage on all sequences in schema df to dwh;
grant execute on all functions in schema df to dwh;
```

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

14

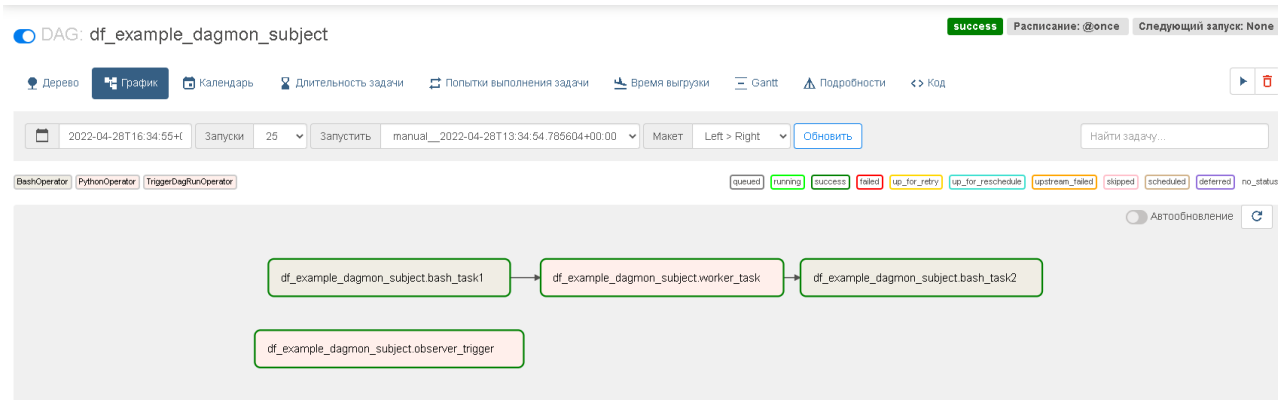


Рисунок 2. Форма просмотра информации о процессе

2.3 Порядок обновления системы

Ниже указан общий порядок обновления:

1. запустить обновление

```
python3 manage.py --update --backup --no-upgrade
```

2. перезапустить сервис

```
docker-compose down && docker-compose up -d
```

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата	.РЭ					Лист
										16
Изм	Лист	№ докум.	Подп.	Дата						

3 Описание операций администрирования

3.1 Определения и сокращения Polyflow

Определения и сокращения Polyflow представлены в таблице 1.

Таблица 1. Определения Polyflow

Термин/Сокращение	Определение
Аутентификация	Проверка принадлежности пользователю указанного им пароля.
Пользователь	Авторизованный пользователь, учетная запись которого позволяет просматривать данные на портале.
Веб-интерфейс	Сайт в компьютерной сети, который предоставляет пользователю интерактивный интернет-сервис, который работает в рамках этого сайта.
Планировщик (Airflow Scheduler)	Компонент, который отслеживает состояние DAG и запускает задачи, зависимости которых были удовлетворены. После запуска системы планировщик работает непрерывно, чтобы отслеживать и синхронизировать папку, содержащую объекты DAG.
Хранилище данных (англ. Content Repository, Data Warehouse, DWH)	Предметно-ориентированная информационная база данных, сочетающая в себе функции системы управления версиями, поисковой машины и СУБД.
Система управления данными Polyflow	Сервис оркестровки сбора и обработки разнородных данных хранилища произвольной архитектуры.
Polyflow	Краткое наименование программного обеспечения «Система управления данными Polyflow»
DAG (Directed Acyclic Graph)	Смысловое объединение задач, которые необходимо выполнить в строго определенной последовательности согласно указанному расписанию.
Task	Операции, применяемые к данным, например: загрузка данных из различных источников, их агрегирование, индексирование, очистка от

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

17

	дубликатов, сохранение полученных результатов и прочие ETL-процессы.
OpenID	Открытый стандарт децентрализованной системы аутентификации, предоставляющей пользователю возможность создать единую учётную запись для аутентификации на множестве не связанных друг с другом интернет-ресурсов, используя услуги третьих лиц.
Operator	Сущность, на основе которой создаются экземпляры заданий, где описывается, что будет происходить во время исполнения экземпляра задания.
Sensor	Тип оператора, позволяющий описывать реакцию на определенное событие.
ETL	(от англ. Extract, Transform, Load – дословно «извлечение, преобразование, загрузка») – один из основных процессов в управлении хранилищами данных.
SLA	(от англ. Service Level Agreement) – соглашение об уровне сервиса.

3.2 Настройки компонентов платформы

3.2.1 Общие настройки компонентов платформы

Перед установкой необходимо проверить наличие файлов ключей и сертификатов `key.pem` и `cert.pem`, пути до которых необходимо будет указать в ходе установки.

3.2.2 Ограничение ресурсов для контейнера

Для ограничения ресурсов контейнера необходимо внести изменения в файл `docker-compose.yml`:

- `resources`: настраивает ограничения физических ресурсов для запуска контейнера на платформе. Эти ограничения могут быть настроены как:

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

18

- `limits`: платформа должна запрещать контейнеру выделять больше.
- `reservations`: платформа должна гарантировать, что контейнер может выделить как минимум заданный объем.

```

services:
  frontend:
    image: awesome/webapp
    deploy:
      resources:
        limits:
          cpus: '0.50'
          memory: 50M
          pids: 1
        reservations:
          cpus: '0.25'
          memory: 20M

```

- `cpus` настраивает ограничение или резервирование того, сколько доступных ресурсов ЦП (как количество ядер) может использовать контейнер.
- `memory` настраивает ограничение или резервирование объема памяти, который может выделять контейнер, в виде строки, выражающей значение байта .
- `pids` настраивает ограничение PID контейнера, заданное как целое число.

Име. № подл.	Подпись и дата
Взам. инв. №	Име. № дубл.
Подпись и дата	Подпись и дата

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

19

3.2.3 Настройка схем авторизации

3.2.3.1 Аутентификация через сервис, поддерживающий OpenID Connect

Приведен пример настройки oauth2 sso polyflow через keycloak (настройка на промышленном решении должна выполняться с учетом особенностей целевой среды):

1. развернуть keycloak (<18 или --spi-login-protocol-openid-connect-legacy-logout-redirect-uri=true, пока не вышла 19 с logout clien_id);
2. создать клиент в выбранном пространстве (Realm -> Configure -> Clients -> Create):
 - o id: dataflow
 - o protocol: openid-connect
3. настроить клиент (Clients -> Dataflow -> Settings):
 - o access type: confidential
 - o authorization enabled: on
 - o valid redirect urls: https://dataflow_host:dataflow_port/*
4. создать роли (Clients -> Dataflow -> Roles -> Add Role):
 - o dataflow_admin
 - o dataflow_op
 - o dataflow_user
 - o dataflow_view
 - o dataflow_public
5. создать группы (Realm -> Manage -> Groups):
 - o dataflow_admin
 - o dataflow_op
 - o dataflow_user
 - o dataflow_view
 - o dataflow_public
6. связать роли и группы (Group -> Edit -> Role Mappings);

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата

.РЭ

Лист

20

- для каждой созданной группы Client Roles -> dataflow -> Available Roles -> select role -> Add selected;
7. при создании пользователя включить его в соответствующую группу;
 8. тестовый пример запроса токена:

```
curl -s -X POST https://keycloak.local:7070/realms/testrealm/protocol/openid-connect/token \
  --header 'Content-Type: application/x-www-form-urlencoded' \
  --data-urlencode 'grant_type=password' \
  --data-urlencode 'client_id=dataflow' \
  --data-urlencode 'client_secret=Bc4tBp05HoUXZxOTigXhY9os4sXSNxCZ' \
  --data-urlencode 'username=admin' \
  --data-urlencode 'password=somepass' | jq
```

9. создать файл webserver_config.py в директории dataflow. Пример webserver_config.py:

```
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements.  See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership.  The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License.  You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing,
# software distributed under the License is distributed on an
# "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY
# KIND, either express or implied.  See the License for the
# specific language governing permissions and limitations
# under the License.
"""
webserver_config
Referencies
```

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ


```

class CustomSecurityManager(AirflowSecurityManager):
    authoauthview = CustomAuthRemoteUserView

    def oauth_user_info(self, provider, response=None):
        if provider == PROVIDER:
            token = response["access_token"]
            me = jwt.decode(token, algorithms="RS256", verify=False)
            groups = me["resource_access"]["dataflow"]["roles"] # unsafe

            if len(groups) < 1:
                groups = ["dataflow_public"]
            else:
                groups = [str for str in groups if "dataflow" in str]

            userinfo = {
                "username": me.get("preferred_username"),
                "email": me.get("email"),
                "first_name": me.get("given_name"),
                "last_name": me.get("family_name"),
                "role_keys": groups,
            }
            log.debug("user info: {0}".format(userinfo))
            return userinfo
        else:
            return {}

SECURITY_MANAGER_CLASS = CustomSecurityManager
APP_THEME = "cerulean.css" # default "bootstrap-theme.css"

```

10.подключить файл webserver_config.py как том в docker-compose.yml:

```

volumes:
  - ./webserver_config.py:/home/airflow/webserver_config.py

```

11.в airflow.cfg в секции webserver установить cookie_samesite = lax;

12.перезапустить сервис dataflow.

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

3.3 Операции группы безопасность

Через верхнее меню Безопасность доступны формы, относящиеся к безопасности системы.

3.3.1 Список пользователей

Данная форма позволяет найти, просмотреть, добавить, изменить или удалить пользователей системы (Рисунок 3).

Имя	Фамилия	Логин	Email	Активен?	Роль
admin	super	admin	admin@example.com	True	[Admin]

Рисунок 3. Список пользователей

По кнопке «показать запись» по выбранному в таблице пользователю доступна подробная информация и данные по аудиту (Рисунок 4).

Ваши данные	
Логин	admin
Активен?	True
Роль	[Admin]
Количество входов	785

Частные данные	
Имя	admin
Фамилия	super
Email	admin@example.com

Данные по аудиту	
Последний вход	2022-05-11 15:31:39.556015
Не удалось посетить количество входов	0
Создано	2022-04-20 10:04:09.479935
Кем создано	None
Когда изменено	2022-04-20 10:04:09.479941
Кем изменено	None

Рисунок 4. Информация по пользователю

Подпись и дата
Инв. № дубл.
Взам. инв. №
Подпись и дата
Инв. № подл.

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

В случае необходимости данные по пользователю можно отредактировать (Рисунок 5):

Редктировать пользователя

Имя * admin
Выводить имя пользователя

Фамилия * super
Выводить фамилию

Логин * admin
Имя пользователя для авторизации на БД или LDAP, не используется при OID аутентификации

Активен?
Удалять пользователя плохая практика, просто сделайте его неактивным

Email * admin@example.com
E-mail ользователя. Будет также использоваться для OID аутентификации

Роль *
Роль пользователя, определяющая список его прав

Сохранить ↩

Рисунок 5. Редактирование данных пользователя

Для добавления нового пользователя необходимо нажать кнопку «+» и ввести необходимую информацию (Рисунок 6):

Добавить пользователя

Имя * Имя
Выводить имя пользователя

Фамилия * Фамилия
Выводить фамилию

Логин * Логин
Имя пользователя для авторизации на БД или LDAP, не используется при OID аутентификации

Активен?
Удалять пользователя плохая практика, просто сделайте его неактивным

Email * Email
E-mail ользователя. Будет также использоваться для OID аутентификации

Роль * Select Value
Роль пользователя, определяющая список его прав

Пароль * Пароль
The user's password for authentication

Подтвердите пароль * Подтвердите пароль
Повторите пароль

Сохранить ↩

Рисунок 6. Добавление нового пользователя

3.3.2 Список ролей

Данная форма позволяет найти, просмотреть, добавить, изменить или удалить роли системы. Также через форму осуществляется настройка прав ролей (Рисунок 7).

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

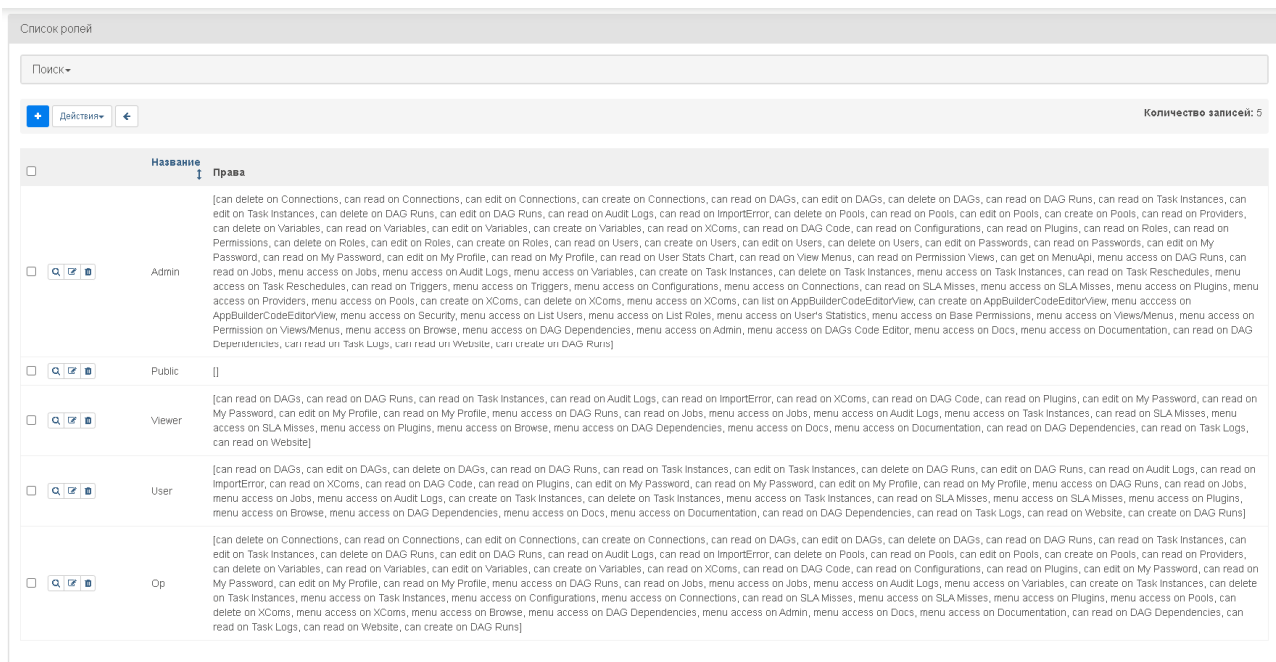


Рисунок 7. Список ролей

Для выбранной роли доступна информация по назначенным правам (Рисунок 8) и пользователям (Рисунок 9), которым роль доступна.

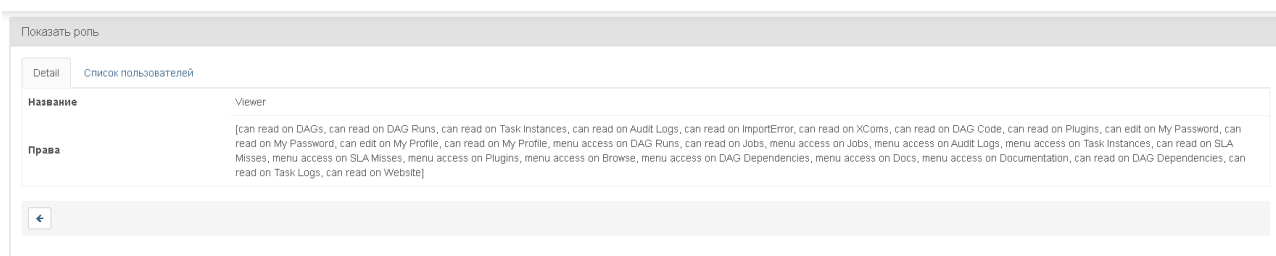


Рисунок 8. Информация по правам роли



Рисунок 9. Информация по пользователям роли

В случае необходимости состав прав роли можно отредактировать (Рисунок 10).

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

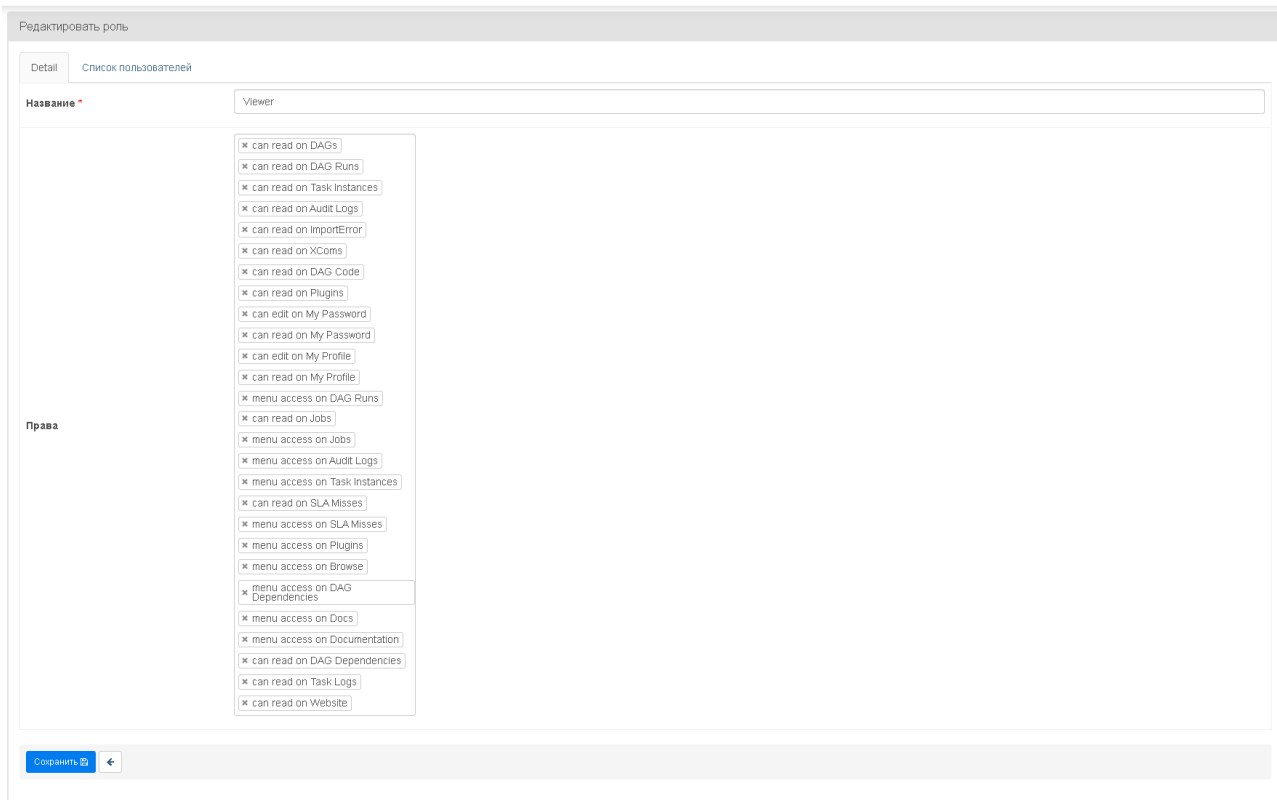


Рисунок 10. Редактирование роли

Также через форму управления ролями доступно добавление новых пользователей той или иной роли.

3.3.3 Статистика пользователя

Форма отображает информацию со статистикой по пользователям (Рисунок 11).

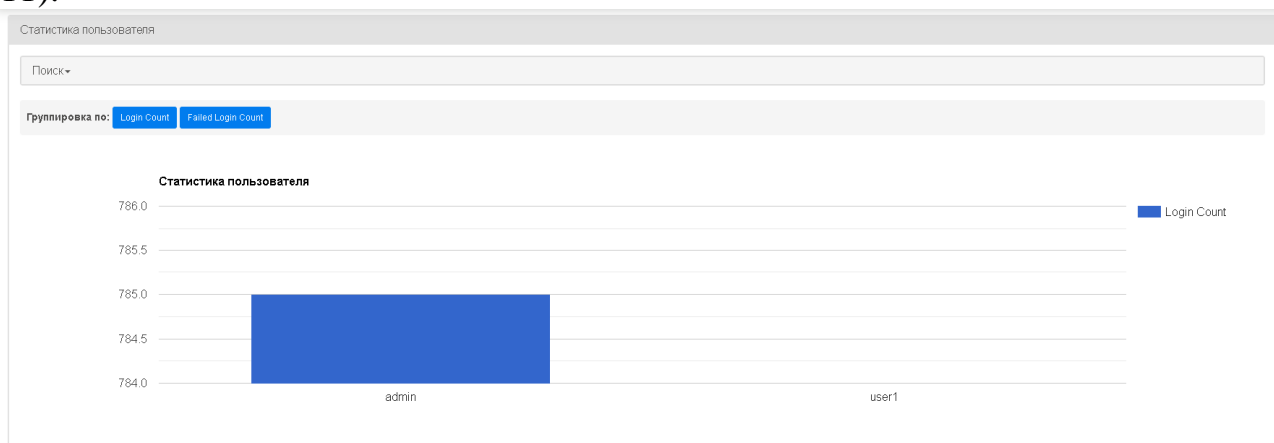


Рисунок 11. Статистика пользователя

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

3.3.4 Основные разрешения

Форма отображает список базовых прав доступа (элементарных действий над ресурсами системы, такими как: DAG, запуски DAG, задачи и подключения) (Рисунок 12).



Рисунок 12. Список базовых прав доступа

3.3.5 Отображения/меню

Форма отображает список отображений/меню, на которые возможна выдача базовых прав доступа (Рисунок 13).

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата	.РЭ					Лист				
										29				
										Изм	Лист	№ докум.	Подп.	Дата

Список прав на просмотр/меню

Поиск

Количество записей: 398

разрешение	Отображение/меню
can_delete	Connections
can_read	Connections
can_edit	Connections
can_create	Connections
can_read	DAGs
can_edit	DAGs
can_delete	DAGs
can_read	DAG Runs
can_read	Task Instances
can_edit	Task Instances
can_delete	DAG Runs
can_edit	DAG Runs
can_read	Audit Logs
can_read	ImportError
can_delete	Pools
can_read	Pools
can_edit	Pools
can_create	Pools
can_read	Providers
can_delete	Variables
can_read	Variables
can_edit	Variables
can_create	Variables
can_read	XComs
can_read	DAG Code

Рисунок 14. Список прав на просмотр/меню

3.4 Операции группы Admin

К операциям администрирование Polyflow относятся: управление переменными, подключениями, провайдерами, плагинами.

Меню Admin содержит пункты:

3.4.1 Variables

Содержит ссылку на страницу с переменными, доступными для использования в DAG (Рисунок 15).

Выберите файл | Файл не выбран | Импорт переменных

List Variable

Поиск

Количество записей: 1

	Key	Val	Description	Is Encrypted
<input type="checkbox"/>	logical_date	31.03.2022		True

Рисунок 15. Переменные

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Форма позволяет просматривать, добавлять, редактировать (Рисунок 16), импортировать и экспортировать переменные (Рисунок 17).

Рисунок 16. Редактирование переменной

Рисунок 17. Действия над переменными

3.4.2 Configuration

Содержит ссылку на страницу с информацией о текущей конфигурации. Доступность страницы и информации может быть задана настройками.

3.4.3 Connections

Предназначен для настройки подключений (Рисунок 18).

Име. № дубл.	Подпись и дата
Взам. инв. №	
Име. № подл.	

Изм	Лист	№ докум.	Подп.	Дата	.РЭ	Лист 32

List Connection							
Поиск							
+ Действия							Количество записей: 49
<input type="checkbox"/>	Conn Id	Conn Type	Description	Host	Port	Is Encrypted	Is Extra Encrypted
<input type="checkbox"/>	aws_default	aws				False	False
<input type="checkbox"/>	azure_batch_default	azure_batch				False	True
<input type="checkbox"/>	azure_cosmos_default	azure_cosmos				False	True
<input type="checkbox"/>	azure_data_explorer_default	azure_data_explorer		https://<CLUSTER>.kusto.windows.net		False	True
<input type="checkbox"/>	azure_data_lake_default	azure_data_lake				False	True
<input type="checkbox"/>	azure_default	azure				False	False
<input type="checkbox"/>	cassandra_default	cassandra		cassandra	9042	False	False
<input type="checkbox"/>	databricks_default	databricks		localhost		False	False
<input type="checkbox"/>	dingding_default	http				False	False
<input type="checkbox"/>	drill_default	drill		localhost	8047	False	True
<input type="checkbox"/>	druid_broker_default	druid		druid-broker	8082	False	True
<input type="checkbox"/>	druid_ingest_default	druid		druid-overlord	8081	False	True
<input type="checkbox"/>	elasticsearch_default	elasticsearch		localhost	9200	False	False
<input type="checkbox"/>	emr_default	emr				False	True
<input type="checkbox"/>	facebook_default	facebook_social				False	True
<input type="checkbox"/>	fs_default	fs				False	True
<input type="checkbox"/>	google_cloud_default	google_cloud_platform				False	False
<input type="checkbox"/>	hive_cli_default	hive_cli		localhost	10000	False	True
<input type="checkbox"/>	hiveserver2_default	hiveserver2		localhost	10000	False	False

Рисунок 18. Подключения

Форма позволяет просматривать, добавлять, редактировать подключения (Рисунок 19).

Edit Connection	
Идентификатор соединения *	testpostgresql
Тип подключения *	Postgres <small>Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.</small>
Описание	
Хост	172.24.0.41
Схема	appdb
Логин	dwh
Пароль	
Порт	5432
Дополнительно	{'schema': 'dwh'}
<input type="button" value="Сохранить"/> <input type="button" value="Тест"/>	

Рисунок 19. Редактирование подключения

3.4.4 Plugins

Предназначен для просмотра информации о плагинах (Рисунок 20, Рисунок 21, Рисунок 22).

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Airflow Plugins

1. datflow_api_plugin

Атрибут	Значение
hooks	[]
executors	[]
macros	[]
admin_views	[]
flask_blueprints	[<flask.blueprints.Blueprint object at 0x7fa3efef4876d>]
menu_links	[]
appbuilder_views	[{'category': 'Docs', 'name': 'Dataflow API OpenAPI', 'view': <view.AppView object at 0x7fa3efef52310>}]
appbuilder_menu_items	[{'category': 'Docs', 'name': 'about Dataflow API', 'href': 'https://gitlab-ortv.polymedia.ru/system/wiki'}]
global_operator_extra_links	[]
operator_extra_links	[]
source	<code>SPLUGINS_FOLDER/views/dataflow_api/view.py</code>

Рисунок 20. Информация о плагинах

2. setup

Атрибут	Значение
hooks	[]
executors	[]
macros	[]
admin_views	[]
flask_blueprints	[]
menu_links	[]
appbuilder_views	[]
appbuilder_menu_items	[]
global_operator_extra_links	[]
operator_extra_links	[]
source	<code>SPLUGINS_FOLDER/common/setup.py</code>

Рисунок 21. Информация о плагинах (продолжение)

3. editor_plugin

Атрибут	Значение
hooks	[]
executors	[]
macros	[]
admin_views	[]
flask_blueprints	[<flask.blueprints.Blueprint object at 0x7fa3efdfcfa0>]
menu_links	[]
appbuilder_views	[{'category': 'Admin', 'name': 'DAGs Code Editor', 'view': <airflow_code_editor.app_builder_view.AppBuilderCodeEditorView object at 0x7fa3efef1f7f0>}]
appbuilder_menu_items	[]
global_operator_extra_links	[]
operator_extra_links	[]
source	<code>airflow-code-editor==5.1.5: EntryPoint(name='airflow_code_editor', value='airflow_code_editor.airflow_code_editor.CodeEditorPlugin', group='airflow.plugins')</code>

Рисунок 22. Информация о плагинах (продолжение 2)

3.4.5 Providers

Предназначен для просмотр информации о установленных провайдерах (Рисунок 23).

Подпись и дата

Инв. № дубл.

Взам. инв. №

Подпись и дата

Инв. № подл.

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

34

Providers

Имя пакета	Версия	Описание
apache-airflow-providers-cncf-kubernetes	3.0.0	Kubernetes
apache-airflow-providers-docker	2.5.2	Docker
apache-airflow-providers-ftp	2.1.2	File Transfer Protocol (FTP)
apache-airflow-providers-http	2.1.2	Hypertext Transfer Protocol (HTTP)
apache-airflow-providers-imap	2.2.3	Internet Message Access Protocol (IMAP)
apache-airflow-providers-microsoft-mssql	2.1.3	Microsoft SQL Server (MSSQL)
apache-airflow-providers-postgres	4.1.0	PostgreSQL
apache-airflow-providers-sqlite	2.1.3	SQLite

Рисунок 23. Провайдеры

3.4.6 Pools

Содержит ссылку на страницу с информацией о пулах (Рисунок 24).

Pool	Slots	Running Slots	Queued Slots
default_pool	128	0	0

Рисунок 24. Пул процессов

Форма позволяет просматривать, добавлять, редактировать данные по пулам (Рисунок 25).

Pool: default_pool
Slots: 128
Description: Default pool

Сохранить

Рисунок 25. Редактирование данных пула процессов

3.4.7 XComs

Содержит ссылку на страницу с информацией по XComs (Рисунок 26).

Подпись и дата

Инв. № дубл.

Взам. инв. №

Подпись и дата

Инв. № подл.

Изм	Лист	№ докум.	Подп.	Дата
-----	------	----------	-------	------

.РЭ

Лист

35

ListXComs

Поиск-

Количество записей: 2015

	Key	Value	Timestamp	Logical Date	Task Id	Dag Id
<input type="checkbox"/>	return_value	Всего загружено записей 6 в tuser.test_const	2022-04-20, 13:52:06	2022-04-20, 13:51:56	df_example_extract_excel_const.extract	df_example_extract_excel_const
<input type="checkbox"/>	return_value	Выполняется слияние данных в таблицу <rule>	2022-04-20, 13:08:41	2022-04-20, 13:08:30	merge	df_dataadb_populate
<input type="checkbox"/>	return_value	Выполнение запроса SQL завершено.	2022-04-20, 13:08:16	2022-04-20, 13:08:04	callable_common	df_dataadb_upgrade
<input type="checkbox"/>	return_value	INFO [alembic.runtime.migration] Will assume transactional DDL.	2022-04-20, 13:08:09	2022-04-20, 13:08:04	ddl	df_dataadb_upgrade
<input type="checkbox"/>	return_value	Выполнение запроса SQL завершено.	2022-04-20, 13:08:23	2022-04-20, 13:08:04	callable_json	df_dataadb_upgrade
<input type="checkbox"/>	return_value	Выполнение запроса SQL завершено.	2022-04-20, 13:08:21	2022-04-20, 13:08:04	callable_rules	df_dataadb_upgrade
<input type="checkbox"/>	return_value	'df_example_dagmon_subject' начал свою работу	2022-04-28, 01:24:54	2021-01-01, 03:00:00	df_example_dagmon_subject.bash_task1	df_example_dagmon_subject
<input type="checkbox"/>	trigger_run_id	manual__2022-04-27T08:59:43.140306+00:00	2022-04-27, 11:59:43	2021-01-01, 03:00:00	df_example_monitoring_observer.subject_trigger	df_example_monitoring_observer
<input type="checkbox"/>	trigger_execution_date_iso	2022-04-27T22:24:54.912091+00:00	2022-04-28, 01:24:55	2021-01-01, 03:00:00	df_example_dagmon_subject.observer_trigger	df_example_dagmon_subject
<input type="checkbox"/>	trigger_execution_date_iso	2022-04-27T08:59:43.203879+00:00	2022-04-27, 11:59:43	2021-01-01, 03:00:00	df_example_monitoring_observer.observer_trigger	df_example_monitoring_observer
<input type="checkbox"/>	trigger_run_id	manual__2022-04-27T22:24:54.912091+00:00	2022-04-28, 01:24:55	2021-01-01, 03:00:00	df_example_dagmon_subject.observer_trigger	df_example_dagmon_subject
<input type="checkbox"/>	trigger_run_id	manual__2022-04-27T08:59:43.203879+00:00	2022-04-27, 11:59:43	2021-01-01, 03:00:00	df_example_monitoring_observer.observer_trigger	df_example_monitoring_observer
<input type="checkbox"/>	return_value	'df_example_dagmon_subject' завершает свою работу	2022-04-28, 01:25:08	2021-01-01, 03:00:00	df_example_dagmon_subject.bash_task2	df_example_dagmon_subject
<input type="checkbox"/>	return_value	df_example_monitoring_subject начал свою работу	2022-04-27, 11:50:11	2021-01-01, 03:00:00	df_example_monitoring_subject.bash_task1	df_example_monitoring_subject
<input type="checkbox"/>	trigger_execution_date_iso	2022-04-27T08:59:43.140306+00:00	2022-04-27, 11:59:43	2021-01-01, 03:00:00	df_example_monitoring_observer.subject_trigger	df_example_monitoring_observer

Рисунок 26. XComs

3.4.8 DAGs Code Editor

Содержит ссылку на страницу редактора DAG (Рисунок 27, Рисунок 28).

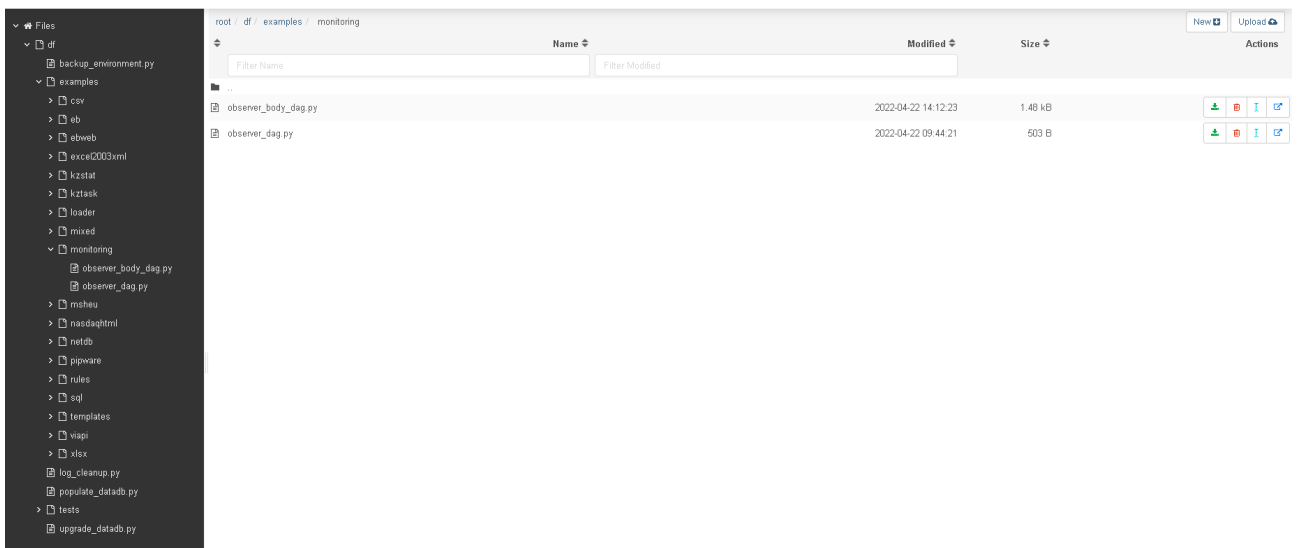


Рисунок 27. Редактор DAG

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата	.РЭ	Лист 36

```

root / df / examples / monitoring / observer_body_dag.py
1  """
2  """
3  import pendulum
4
5  from airflow import DAG
6  from airflow.decorators import task
7  from airflow.operators.bash import BashOperator
8
9
10 @task(task_id="run_this")
11 def run_this_func(dag_run=None):
12     """
13     Print the payload "message" passed to the DagRun conf attribute.
14     :param dag_run: The DagRun conf object
15     :type dag_run: DagRun
16     """
17     print(f"Remotely received value of {dag_run.conf.get('message')} for key=message")
18
19
20 def find_str(file_path, pattern):
21     """
22     Поиск строк, реализация в зависимости от решаемых задач
23     """
24     import re
25
26     my_regex = re.compile(pattern)
27     found_strs = set()
28     for line in open(file_path):
29         match = my_regex.match(line)
30         if match:
31             found_strs.add(match.group(0))
32
33     s = None
34     if found_strs:
35         s = '\n'.join(found_strs)
36     return s
37
38
39 def get_task_log_path(dag_id, task_id):
40     """
41     TODO
42     """
43     return f'home/airflow/logs/{dag_id}/{task_id}/2022-04-22T10:14:34.378446+00:00/1.log'
44
45
46 DAG_ID = 'df_example_observer_body'
47 state = 'wal'
48
49 with DAG(
50     dag_id=DAG_ID,
51     start_date=pendulum.datetime(2021, 1, 1, tz="UTC"),
52     catchup=False,
53     schedule_interval=None,
54     tags=['example'],
55 ) as dag:
56     run_this = run_this_func()
57
58     bash_task = BashOperator(
59         task_id="print_the_bash_task"

```

Рисунок 28. Редактор DAG с открытым файлом

Инв. № подл.	Подпись и дата			
	Инв. № дубл.			
Инв. № докл.	Взам. инв. №			
	Подпись и дата			
Изм	Лист	№ докум.	Подп.	Дата

.РЭ

4 Аварийные ситуации

Для Системы определены следующие режимы функционирования:

- штатный;
- аварийный.

Аварийный режим функционирования Системы используется при отказе одного или нескольких компонент программного и (или) технического обеспечения.

При переходе в аварийный режим в Системе предусмотрено формирование соответствующего информационного сообщения.

После выдачи сообщения, администратору необходимо выполнить комплекс мероприятий по устранению причины перехода Системы в аварийный режим.

При работе с АИС могут возникнуть следующие неисправности, приводящие к аварийным ситуациям:

- Превышение нагрузки на АИС. В этом случае необходимо ограничить количество тяжело-нагруженных процессов или общее их количество;
- Недостаток свободной оперативной памяти на сервере. В этом случае необходимо ограничить ресурсы для контейнера.
- Другие неисправности. В случае нарушения технологического процесса или при длительных отказах технических средств администратор системы обязан сообщить о возникшей проблеме в службу технической поддержки, провести диагностику работы Системы, определить вероятную причину неисправности и передать лог-файлы из соответствующего docker-контейнера. Чтобы связаться с службой поддержки необходимо сообщить о возникшей неисправности по электронному адресу: support@polyanalitika.ru.

Подпись и дата	
Инв. № дубл.	
Взам. инв. №	
Подпись и дата	
Инв. № подл.	

Изм	Лист	№ докум.	Подп.	Дата

.РЭ

Лист

38

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата

Изм	Лист	№ докум.	Подп.	Дата

.РЭ