

Программное обеспечение

**Система управления данными Polyflow**

РУКОВОДСТВО АДМИНИСТРАТОРА

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

## Аннотация

Настоящий документ является руководством администратора системы управления данными Polyflow.

Документ разработан в соответствии с требованиями ГОСТ Р 59795-2021 «Требования к содержанию документов».

Инв. № подл.	Изм.	Лист	№ докум.	Подп.	Дата	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата
Разраб.						.РЭ			
Пров.									
Н. контр.									
Утв.									
Polyflow Руководство администратора						Лит.	Лист	Листов	
							2	64	
						Наименование исполнителя			

# Содержание

<b>1 Назначение и условия применения.....</b>	<b>8</b>
1.1 Назначение системы.....	8
1.2 Условия применения.....	8
1.2.1 Серверная часть.....	8
1.2.2 Локальная сеть.....	9
<b>2 Подготовка к работе.....</b>	<b>10</b>
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 Развертывание Polyflow в облачных средах.....	12
2.1.4 Гибридные конфигурации Polyflow .....	15
2.1.5 Кластерные конфигурации и высокая доступность.....	17
2.1.6 Описание утилиты управления сервисом Polyflow manage.py.....	24
2.1.7 Описание файлов и структуры проекта .....	28
2.1.8 Поддержка локализации .....	29
2.1.9 Настройки в airflow.cfg.....	29
2.1.10 Требования к СУБД хранилища .....	29
2.2 Порядок проверки работоспособности .....	31
2.3 Порядок обновления системы.....	32
<b>3 Описание операций администрирования .....</b>	<b>33</b>
3.1 Определения и сокращения Polyflow .....	33
3.2 Настройки компонентов платформы.....	34
3.2.1 Общие настройки компонентов платформы .....	34
3.2.2 Ограничение ресурсов для контейнера.....	34
3.2.3 Настройка схем авторизации .....	36
3.3 Настройка уведомлений по SMTP .....	41
3.3.1 Конфигурация SMTP сервера .....	41
3.3.2 EmailHook - системный хук для работы с email.....	42
3.3.3 Настройка шаблонов писем .....	43
3.3.4 Пример пользовательского шаблона HTML .....	43
3.3.5 Практические рекомендации .....	44
3.4 Операции группы безопасность.....	44
3.4.1 Список пользователей .....	44

Инв. №подл.	Подпись и дата	Взам. инв. №	Инв. №дубл.	Подпись и дата	2.1.10 Требования к СУБД хранилища .....29			
					2.2 Порядок проверки работоспособности .....31			
					2.3 Порядок обновления системы.....32			
					<b>3 Описание операций администрирования.....33</b>			
					3.1 Определения и сокращения Polyflow .....33			
Инв. №подл.	Подпись и дата	Взам. инв. №	Инв. №дубл.	Подпись и дата	3.2 Настройки компонентов платформы.....34			
					3.2.1 Общие настройки компонентов платформы .....34			
					3.2.2 Ограничение ресурсов для контейнера.....34			
					3.2.3 Настройка схем авторизации .....36			
					3.3 Настройка уведомлений по SMTP .....41			
Инв. №подл.	Подпись и дата	Взам. инв. №	Инв. №дубл.	Подпись и дата	3.3.1 Конфигурация SMTP сервера .....41			
					3.3.2 EmailHook - системный хук для работы с email.....42			
					3.3.3 Настройка шаблонов писем .....43			
					3.3.4 Пример пользовательского шаблона HTML .....43			
					3.3.5 Практические рекомендации .....44			
Инв. №подл.	Подпись и дата	Взам. инв. №	Инв. №дубл.	Подпись и дата	3.4 Операции группы безопасность.....44			
					3.4.1 Список пользователей .....44			
					<div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div>Изм</div><div>Лист</div><div>№ докум.</div><div>Подп.</div><div>Дата</div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div><div></div><div></div><div></div><div></div></div></div></div></div></div></div></div></div>			

3.4.2	Список ролей .....	46
3.4.3	Статистика пользователя .....	48
3.4.4	Основные разрешения .....	48
3.4.5	Отображения/меню .....	48
3.4.6	Право на просмотр/меню .....	49
3.5	Операции группы Admin .....	50
3.5.1	Variables .....	50
3.5.2	Configuration .....	51
3.5.3	Connections .....	51
3.5.4	Plugins .....	52
3.5.5	Providers .....	53
3.5.6	Pools .....	54
3.5.7	XComs .....	58
3.5.8	DAGs Code Editor .....	58
3.6	Политики безопасности и шифрования .....	59
3.6.1	Шифрование передаваемых данных .....	59
3.6.2	Рекомендации по настройке .....	60
4	Аварийные ситуации .....	61
5	Рекомендации по освоению .....	62

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

## Введение

Модуль Polyflow представляет собой сервис оркестровки, сбора и обработки разнородных данных в хранилищах произвольной архитектуры.

Уровень подготовки персонала, необходимого для работы с ИАС, предполагает наличие следующих групп пользователей:

- служба эксплуатации ИАС;
- разработчики;
- операторы.

### Служба эксплуатации ИАС

В службу эксплуатации ИАС входят специалисты следующих категорий: «Администратор защиты (безопасности) информации», «Администратор операционных систем», «Администратор баз данных».

- Администратор защиты (безопасности) информации обеспечивает:
  - Формирование списка пользователей, допущенных к работе с Системой;
  - Настройку учетных записей пользователей, управление ролями доступа, а также интеграция пользователей с помощью LDAP;
  - Формирование матрицы доступа к ресурсам Системы и данным, а также изменение прав доступа.
- Администратор операционных систем отвечает за:
  - Установку компонентов платформы, активацию и первоначальную настройку.
  - Сопровождение ИАС (тестирование работоспособности, восстановление и т.п.), обновление версий (анализ необходимости перехода на новые версии, разработку перечня мероприятий по переводу на новую версию).
- Администратор баз данных отвечает за:

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата						Лист
Изм.	Лист	№ докум.	Подп.	Дата	.РЭ					5

- Генерацию систем управления базами данных;
- Сопровождение и управление информационными ресурсами;
- Сохранение резервных копий, восстановление искаженной информации, архивирование информации и организацию поступления информации из архива;
- Обработку и анализ статистической информации о характере и интенсивности использования данных, о распределении нагрузки на различные компоненты структуры баз данных, внесение изменений в структуру баз данных в процессе эксплуатации Системы с целью повышения производительности, обеспечивает ввод и поддержание в актуальном состоянии общих разделов баз данных (классификаторов).

Служба эксплуатации обеспечивает функционирование в штатном режиме технических и программных средств АИС, отслеживает процессы наполнения АИС данными.

Поддержка функционирования Системы должна осуществляться силами действующей Службы эксплуатации АИС, состоящей из специалистов, обладающих знаниями в области информационных и сетевых платформ, на которых реализована АИС, и опытом администрирования баз данных.

#### Разработчики

Разработчиками являются специалисты, которые участвуют в процессах разработки процессов загрузки данных в Систему, описания сущностей-получателей и источников, проведения анализа работы процессов, выявления аномалий и их причин.

Разработчики должны иметь опыт разработки в своей отрасли, обладать навыками работы с DWH решениями и базовыми знаниями SQL и Python.

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата	.РЭ	Лист
Изм	Лист	№ докум.	Подп.	Дата		6

Операторы

Операторами являются специалисты, которые участвуют в процессах сбора данных, базовой настройки, запуска и мониторинга таких процессов.

Для работы с системой необходимо ознакомиться со следующим набором эксплуатационной документации:

- Руководство администратора;
- Руководство разработчика;
- Руководство пользователя.

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

Изм	Лист	№ докум.	Подп.	Дата	.РЭ	Лист 7





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

					.РЭ	Лист
Изм	Лист	№ докум.	Подп.	Дата		9







	Образ: Container Optimized Image или стандартный образ Ubuntu/CentOS с предустановленным Docker.
Управляемый Kubernetes	Кластер: 3 ноды (виртуальные машины) с конфигурацией не ниже 2 vCPU, 8 ГБ RAM.  Преимущество: сервис управляет мастер-нодами, обеспечивая отказоустойчивость и упрощая эксплуатацию.
Docker-образы (Container Registry)	Использование: хранение и управление образами df_operator, df_airflow, df_postgres13 и другими кастомными образами.

Порядок развертывания:

1) Создайте виртуальную машину или кластер Managed Service for Kubernetes в Яндекс Облаке.

2) Установите Docker и Docker Compose на ВМ (если используется Compute Cloud).

3) Скопируйте файлы дистрибутива Polyflow (manage.py, docker-compose.yml.tpl и др.) на сервер.

4) Запустите развертывание с помощью утилиты manage.py.

### 2.1.3.2 Конфигурация на Selectel

Selectel предлагает надежную облачную инфраструктуру с возможностью развертывания виртуальных машин и управления контейнерами.

Таблица 2. Рекомендуемая конфигурация сервисов

Компонент Selectel	Рекомендация для Polyflow
Облачные серверы	Конфигурация: 2+ vCPU, 8+ ГБ RAM, 50+ ГБ SSD  Образ: Выберите образ Ubuntu 20.04 LTS или CentOS 7 с последующей установкой Docker.

Име. № подл.	Подпись и дата	Взам. инв. №	Име. № дубл.	Подпись и дата						Лист
Изм.	Лист	№ докум.	Подп.	Дата	.РЭ					13

Управляемый Kubernetes	Кластер: 3 рабочие ноды с аналогичными характеристиками. Selectel полностью управляет жизненным циклом кластера.
------------------------	--

Порядок развертывания:

- 1) Активируйте облачные серверы или управляемый Kubernetes в панели управления Selectel.
- 2) Убедитесь, что на серверах установлены Docker и Docker Compose.
- 3) Разместите файлы дистрибутива Polyflow в домашней директории пользователя.
- 4) Выполните стандартную процедуру установки, как описано в разделе «Подготовка к работе».

### 2.1.3.3 Ключевые преимущества облачного размещения

- Масштабируемость: легко увеличивайте вычислительные ресурсы (vCPU, RAM) или добавляйте ноды в кластер Kubernetes для обработки растущей нагрузки.
- Отказоустойчивость: облачные провайдеры гарантируют высокую доступность дата-центров и сервисов. При использовании Kubernetes достигается отказоустойчивость на уровне приложения.
- Управляемость: использование управляемых сервисов (Kubernetes, реестр контейнеров) снижает операционную нагрузку на администраторов.
- Экономическая эффективность: оплата происходит только за потребленные ресурсы, что позволяет оптимизировать затраты.

### 2.1.3.4 Общие рекомендации

- Безопасность: настройте Security Groups (группы безопасности) для строгого ограничения входящего и исходящего трафика. Откройте для доступа только порт, используемый веб-интерфейсом Polyflow (по умолчанию 8080).

Подпись и дата		Инв. № дубл.		Взам. инв. №		Подпись и дата		Инв. № подл.	
Изм	Лист	№ докум.	Подп.	Дата	.РЭ				Лист
									14



Таблица 3. Пример гибридной конфигурации

Компонент	Рекомендуемая среда размещения	Обоснование
Веб-сервер (Webserver)	Частное облако / Локальный ЦОД	Обеспечение безопасного доступа к UI, соблюдение требований к внутреннему шлюзу.
Планировщик (Scheduler)	Частное облако / Локальный ЦОД	Централизованное управление выполнением DAG, требует минимальной задержки с БД метаданных.
База данных метаданных (PostgreSQL)	Частное облако / Локальный ЦОД	Хранение критических метаданных и информации о выполнении задач в защищенной среде.
Воркеры (Workers)	Публичное облако (Яндекс Облако, Selectel)	Эластичное масштабирование: быстрое развертывание дополнительных воркеров для параллельного выполнения большого числа задач.

Ключевые технические требования:

- Сетевое соединение: необходимо обеспечить защищенное сетевое соединение (например, VPN-туннель или выделенный канал связи) между сегментами частного и публичного облака.
- Доступ к данным: Воркеры в публичном облаке должны иметь сетевой доступ к источникам данных и получателям, а также к очереди сообщений (например, Redis) и БД метаданных в частном облаке.

Преимущества гибридного подхода для Polyflow

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

Изм.	Лист	№ докум.	Подп.	Дата	.РЭ	Лист
						16





2.1.5.1 Архитектура высокодоступного кластера

Polyflow поддерживает следующие режимы кластеризации КОМПОНЕНТОВ:

Таблица 4. Режимы кластеризации компонентов

Компонент	Режим кластеризации	Конфигурация
Веб-сервер	Активный-активный	Несколько экземпляров за балансировщиком нагрузки
Планировщик	Активный-резервный	Один активный планировщик, несколько standby
Воркеры	Активный-активный	Неограниченное количество воркеров в пуле
База метаданных	Кластерная СУБД	PostgreSQL с репликацией или MS SQL Server Always On.

2.1.5.2 Конфигурация кластера с CeleryExecutor

Рекомендуемая конфигурация для продакшн-среды:

```
#yaml
# docker-compose-ha.yml
version: '3'
services:
  postgres:
    image: postgres:13
    environment:
      POSTGRES_DB: airflow
      POSTGRES_USER: airflow
      POSTGRES_PASSWORD: airflow

  redis:
    image: redis:6-alpine
    command: redis-server --appendonly yes
```

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

```

webserver:

  image: df_airflow:latest

  deploy:

    replicas: 2

  command: webserver

  environment:

    AIRFLOW__CORE__EXECUTOR: CeleryExecutor

    AIRFLOW__DATABASE__SQL_ALCHEMY_CONN:
postgresql+psycopg2://airflow:airflow@postgres/airflow

    AIRFLOW__CELERY__BROKER_URL: redis://redis:6379/0

  depends_on:

    - postgres

    - redis

```

```

scheduler:

  image: df_airflow:latest

  deploy:

    replicas: 1

  command: scheduler

  environment:

    AIRFLOW__CORE__EXECUTOR: CeleryExecutor

    AIRFLOW__DATABASE__SQL_ALCHEMY_CONN:
postgresql+psycopg2://airflow:airflow@postgres/airflow

    AIRFLOW__CELERY__BROKER_URL: redis://redis:6379/0

  depends_on:

    - postgres

    - redis

```

```

worker:

  image: df_operator:latest

  deploy:

    replicas: 4

  command: worker

  environment:

    AIRFLOW__CORE__EXECUTOR: CeleryExecutor

    AIRFLOW__DATABASE__SQL_ALCHEMY_CONN:
postgresql+psycopg2://airflow:airflow@postgres/airflow

```

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата						
Изм	Лист	№ докум.	Подп.	Дата	.РЭ					Лист
										19

```
AIRFLOW__CELERY__BROKER_URL: redis://redis:6379/0

depends_on:
  - postgres
  - redis
```

### 2.1.5.3 Кластерная конфигурация в Kubernetes

Пример манифеста для развертывания в Kubernetes:

```
#yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: airflow-webserver
spec:
  replicas: 2
  selector:
    matchLabels:
      app: airflow-webserver
  template:
    metadata:
      labels:
        app: airflow-webserver
    spec:
      containers:
        - name: webserver
          image: df_airflow:latest
          command: ["webserver"]
          env:
            - name: AIRFLOW__CORE__EXECUTOR
              value: "KubernetesExecutor"
            - name: AIRFLOW__DATABASE__SQL_ALCHEMY_CONN
              valueFrom:
                secretKeyRef:
                  name: airflow-secrets
                  key: database_url
      ---
apiVersion: apps/v1
```

Инв. № подл.	Подпись и дата				Лист 20
	Инв. № дубл.				
	Взам. инв. №				
	Подпись и дата				
<div style="text-align: center;">.РЭ</div>					
Изм	Лист	№ докум.	Подп.	Дата	

```

kind: Deployment
metadata:
  name: airflow-scheduler
spec:
  replicas: 2
  selector:
    matchLabels:
      app: airflow-scheduler
  template:
    metadata:
      labels:
        app: airflow-scheduler
    spec:
      containers:
        - name: scheduler
          image: df_airflow:latest
          command: ["scheduler"]
          env:
            - name: AIRFLOW__CORE__EXECUTOR
              value: "KubernetesExecutor"

```

#### 2.1.5.4 Преимущества кластерной конфигурации

Отказоустойчивость:

- Автоматическое переключение при отказе компонентов
- Резервирование критических сервисов
- Сохранение состояния выполнения задач

Масштабируемость:

- Динамическое добавление воркеров под нагрузку
- Распределение задач между узлами кластера
- Балансировка нагрузки на веб-интерфейс

Обслуживание:

- Бесперебойное обновление компонентов
- "Горячее" добавление мощностей

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата						
					.РЭ					Лист
										21
Изм.	Лист	№ докум.	Подп.	Дата						

- Минимизация downtime

### 2.1.5.5 Рекомендации по настройке

- 1) База данных: Используйте кластерную СУБД с репликацией
- 2) Брокер сообщений: Redis Sentinel или RabbitMQ Cluster для высокой доступности
- 3) Сеть: Обеспечьте низкую задержку между узлами кластера
- 4) Мониторинг: Настройте health checks для всех компонентов
- 5) Резервное копирование: Регулярные бэкапы БД метаданных

### 2.1.5.6 Требования к инфраструктуре

- Минимальная конфигурация: 3 узла для отказоустойчивости
- Сетевая связность: < 1ms задержка между узлами
- Дисковое пространство: Раздельные тома для данных и логов
- Безопасность: TLS-шифрование трафика между компонентами

### 2.1.5.7 Принципы отказоустойчивости компонентов

## Обеспечение отказоустойчивости (Failover)

Архитектура Polyflow позволяет настроить отказоустойчивую конфигурацию, которая обеспечивает непрерывность работы сервиса даже в случае сбоя отдельных компонентов. Это достигается за счет дублирования критических служб и использования кластерных решений.

### Ключевые компоненты и принципы их резервирования:

- 1) Веб-сервер (Websaver)
  - Роль: предоставляет пользовательский интерфейс для управления и мониторинга рабочих процессов (DAGs).
  - Обеспечение отказоустойчивости: может быть развернут в нескольких экземплярах в конфигурации «активный-активный» (active-active). Для распределения нагрузки и единой точки входа между экземплярами используется балансировщик нагрузки.

## 2) Планировщик (Scheduler)

- Роль: отслеживает расписания и запускает задачи.
- Обеспечение отказоустойчивости: хотя планировщик является ключевым компонентом, для него может быть настроена схема «активный-резервный» (active-passive). В этом случае работает один основной планировщик, в то время как резервный находится в режиме ожидания и готов взять на себя функции в случае сбоя основного. Это позволяет избежать единой точки отказа.

## 3) Исполнители (Workers/Executors)

- Роль: непосредственно выполняют задачи рабочих процессов.
- Обеспечение отказоустойчивости: развертываются в конфигурации «активный-активный». Количество воркеров может масштабироваться горизонтально. При использовании CeleryExecutor задачи автоматически перераспределяются между живыми узлами кластера в случае падения одного из них.

## 4) База данных метаданных (Metadata Database)

- Роль: хранит состояние всех DAG, задач, переменных и подключений.
- Обеспечение отказоустойчивости: для базы данных (например, PostgreSQL) необходимо настроить кластерную конфигурацию с репликацией. Это гарантирует, что в случае отказа основной базы данных одна из реплик сможет принять на себя нагрузку, обеспечивая сохранность и доступность метаданных.

## 5) Брокер сообщений (Message Broker)

- Роль: обеспечивает очередь задач для CeleryExecutor.

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата						
Изм.	Лист	№ докум.	Подп.	Дата	.РЭ					Лист
										23





- `--run-dags` запуск системных dag'ов `df\_datadb\_upgrade` и `df\_datadb\_populate` (для запуска необходимо наличие системного подключения)
- `--sync-perm` синхронизации разрешений и добавление ролей
- `--df-conn-id` идентификатор системного подключения, по умолчанию: `df`
- `--backup` бэкап бд и конфигурации
- `--restore` восстановление бд и конфигурацию
- `--backup-ws` бэкап рабочего пространства, содержащего системный исходный код и артефакты проектов
- `--restore-ws` восстановление рабочего пространства, содержащего системный исходный код и артефакты проектов
- `--init-project` создание структуры проекта
- `--no-rebuild` не пересобирать локальные образы
- `--no-restart` не запускать (перезапускать) сервисы
- `--no-upgrade` не обновлять базу данных
- `--docker-repo` задать репозиторий docker-образов, например: `test.polymedia.ru/r5/`
- `--timeout` задать таймаут выполнения шагов утилиты в секундах (по умолчанию 10 секунд)
- `--localize` пересобрать библиотеки интерфейса (в том числе появляется возможность скомпилировать пользовательские файлы переводов)
- `--rm-default-conns` удалить примеры подключений Airflow
- `--rm-temp-tables` удалить временные таблицы Airflow.

Примеры запуска:

- `python3 manage.py --backup --no-restart --no-rebuild --no-upgrade`  
создать бэкап базы и конфигурации сервиса (при обновлении)

Инв. № подл.	Подпись и дата				Лист 25
	Инв. № дубл.				
	Взам. инв. №				
	Подпись и дата				
<div style="text-align: center;">.РЭ</div>					
Изм	Лист	№ докум.	Подп.	Дата	

- `python3 manage.py --update --backup --no-upgrade --compile` запустить обновление
- `python3 manage.py --deploy` развернуть сервис
- `python3 manage.py --run-dags --no-restart --no-rebuild --no-upgrade` запустить системные dag'и
- `python3 manage.py --no-restart --no-upgrade` пересобрать образы
- `python3 manage.py --no-restart --no-upgrade --localize` пересобрать образы с компиляцией библиотек интерфейса
- `python3 manage.py --add-df-conn --no-restart --no-rebuild --no-upgrade` создать подключение `df` к базе данных со схемой с системными объектами.

### 2.1.6.1 Резервное копирование

Резервное копирование предусматривает использование утилиты manage.py. Сервис при этом останавливается.

Доступно создание резервных копий базы и конфигурации сервиса, а также рабочего пространства.

При создании и восстановлении из резервных копий по умолчанию файлы сохраняются в папке backups директории сервиса.

#### 2.1.6.1.1 Создание бэкапа базы и конфигурации сервиса

Пример создания бэкапа базы и конфигурации сервиса в указанную директорию с последующим запуском сервиса:

- `python3 manage.py --backup ~/backups/polyflow/ --no-rebuild --no-upgrade`

При этом будут созданы:

- копии файлов:
- Dockerfile\_operator
- airflow.env

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата						
Изм.	Лист	№ докум.	Подп.	Дата	.РЭ					Лист
										26



В некоторых случаях (например, при ошибке `Operation not permitted`) команды должны быть запущены с использованием привилегий суперпользователя: `sudo python3 manage.py --backup-ws` или `sudo python3 manage.py --restore-ws`. Это позволит корректно выполнить процесс резервного копирования и восстановления.

### 2.1.7 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
- 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.

Томы:

- /home/osuser/docker/dataflow/volumes/projects - том df\_projects;

Имя. № подл.	Подпись и дата	Имя. № дубл.	Взам. инв. №	Подпись и дата						Лист
					.РЭ					28
Изм	Лист	№ докум.	Подп.	Дата						





```
create extension if not exists "uuid-oss";  
...
```

Хранение данных допустимо также на СУБД MS SQL Server 2008 R2 SP3 10.50.6560.0+.

2.2 Порядок проверки работоспособности

Для проверки работы Polyflow необходимо произвести аутентификацию в системе. После успешной аутентификации откроется главная страница приложения, со списком доступных процессов (Рисунок 1). Необходимо выбрать один из них и перейти на форму просмотра информации о выбранном процессе, кликнув на его название (Рисунок 2).

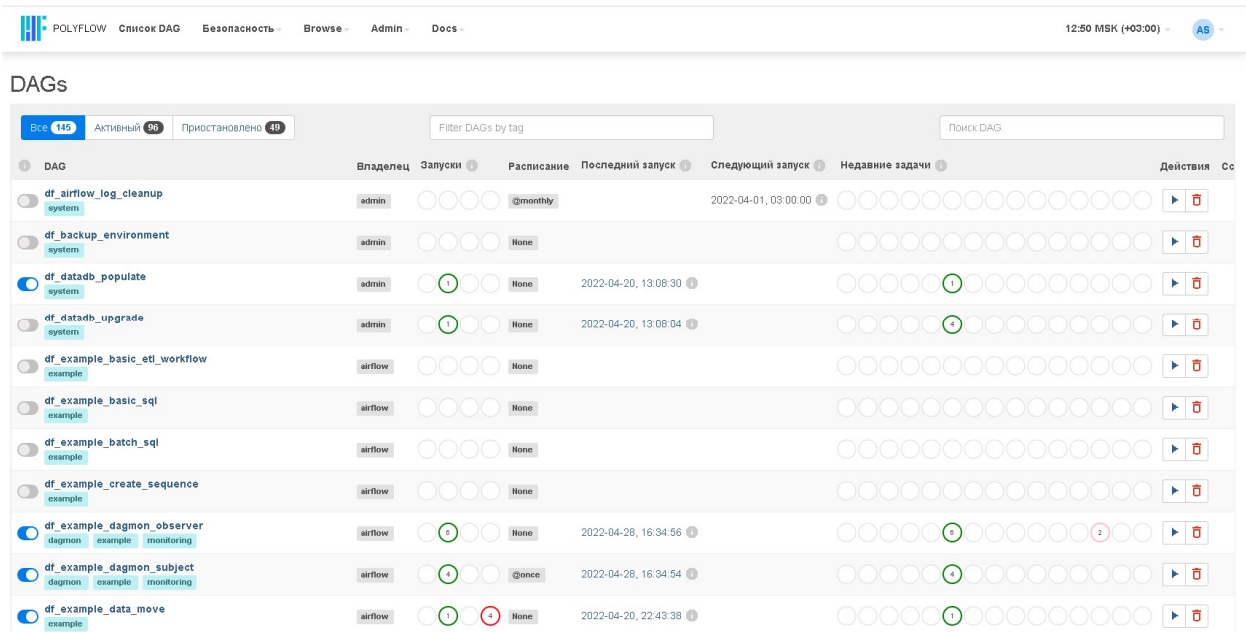


Рисунок 1. Список доступных процессов

Будет открыта форма информации и управления выбранным DAG. По умолчанию открыта вкладка График (Рисунок 2).

Подпись и дата

Инв. № дубл.

Взам. инв. №

Подпись и дата

Инв. № подл.

.РЭ

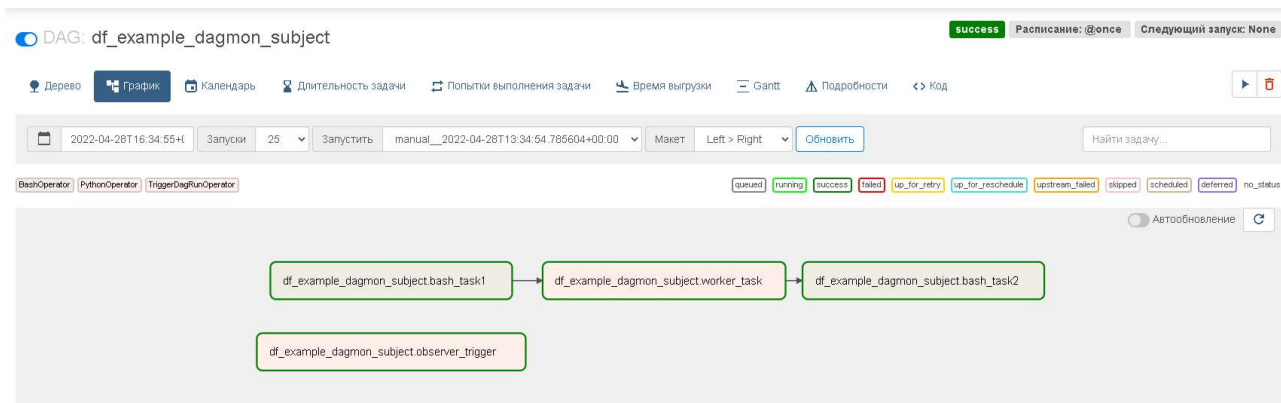


Рисунок 2. Форма просмотра информации о процессе

## 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
```

Име. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата	.РЭ				Лист
									32
Изм.	Лист	№ докум.	Подп.	Дата					



3 Описание операций администрирования

3.1 Определения и сокращения Polyflow

Определения и сокращения Polyflow представлены в таблице 1.

Таблица 5. 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

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



- 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.

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата	.РЭ	Лист				
						35				
						Изм	Лист	№ докум.	Подп.	Дата

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

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

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

- 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
```

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист	№ докум.	Подп.	Дата	<div>РЭ</div>	Лист 37

Изм	Лист</
-----	--------

```

- https://flask-
appbuilder.readthedocs.io/en/latest/security.html#authentication-oauth
"""

import os
import logging
import jwt

from flask import redirect, session, request
from flask_appbuilder import expose
from flask_appbuilder.security.views import AuthOAuthView
from airflow.www.fab_security.manager import AUTH_OAUTH, AUTH_DB
from airflow.www.security import AirflowSecurityManager

basedir = os.path.abspath(os.path.dirname(__file__)) # default
log = logging.getLogger(__name__)

DATAFLOW_HOST = 'https://dataflow.local:8080'
PROVIDER = 'keycloak'

PROVIDER_HOST = 'https://keycloak.local:7070' # visible from dataflow docker
network

REALM = 'testrealm'
CLIENT_ID = 'dataflow'
CLIENT_SECRET = 'Bc4tBp05HoUXZxOTigXhY9os4sXSNxCZ' # set in provider
KEYCLOAK_BASE_URL = f'{PROVIDER_HOST}/realms/{REALM}/protocol/openid-connect'
KEYCLOAK_TOKEN_URL = f'{KEYCLOAK_BASE_URL}/token'
KEYCLOAK_AUTH_URL = f'{KEYCLOAK_BASE_URL}/auth'
KEYCLOAK_LOGOUT_URL = f'{KEYCLOAK_BASE_URL}/logout'

WTF_CSRF_ENABLED = True # default = True
AUTH_TYPE = AUTH_OAUTH # default = AUTH_DB
AUTH_USER_REGISTRATION = False
AUTH_USER_REGISTRATION_ROLE = "Public"
AUTH_ROLES_SYNC_AT_LOGIN = True # sync roles on each login or only on
registration
PERMANENT_SESSION_LIFETIME = 43200

AUTH_ROLES_MAPPING = {

```

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата						
					.РЭ					Лист
										38
Изм	Лист	№ докум.	Подп.	Дата						

```

"dataflow_admin": ["Admin"],
"dataflow_op": ["Op"],
"dataflow_user": ["User"],
"dataflow_viewer": ["Viewer"],
"dataflow_public": ["Public"],
}

OAUTH_PROVIDERS = [
    {
        'name': 'keycloak',
        'icon': 'fa-key',
        'token_key': 'access_token',
        'remote_app': {
            'client_id': CLIENT_ID,
            'client_secret': CLIENT_SECRET,
            'client_kwargs': {
                'scope': 'email profile'
            },
            'api_base_url': KEYCLOAK_BASE_URL,
            'request_token_url': None,
            'access_token_url': KEYCLOAK_TOKEN_URL,
            'authorize_url': KEYCLOAK_AUTH_URL,
        },
    },
]

class CustomAuthRemoteUserView(AuthOAuthView):
    @expose("/logout/")
    def logout(self):
        """Delete access token before logging out."""
        log.info('Logout')
        super().logout()
        return
        redirect(f'{KEYCLOAK_LOGOUT_URL}?redirect_uri={DATAFLOW_HOST}')

```

Инв. № подл.	Подпись и дата	Взам. инв. №	Инв. № дубл.	Подпись и дата	.РЭ	Лист
						39
Изм	Лист	№ докум.	Подп.	Дата		

```

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				Лист
									40







```
with DAG('email_notification_dag',
        start_date=datetime(2024, 1, 1),
        schedule_interval='@daily') as dag:

    send_email_task = PythonOperator(
        task_id='send_email_notification',
        python_callable=send_custom_notification
    )
```

3.3.3 Настройка шаблонов писем

В секции [email] файла airflow.cfg можно настроить пользовательские шаблоны:

```
#ini

[email]

email_backend = airflow.utils.email.send_email_smtp
subject_template = /path/to/custom_subject_template.j2
html_content_template = /path/to/custom_html_template.j2
```

3.3.4 Пример пользовательского шаблона HTML

```
jinja2

<h3>Уведомление о выполнении DAG: {{ ti.dag_id }}</h3>

<p><strong>Задача:</strong> {{ ti.task_id }}</p>

<p><strong>Статус:</strong> {{ ti.state }}</p>

<p><strong>Время выполнения:</strong> {{ df_now }}</p>

<p><strong>Ссылка на логи:</strong> <a href="{{ log_url }}">Перейти к логам</a></p>

<p><strong>Базовая ссылка:</strong> {{ base_url }}</p>

{% if ti.state == 'failed' %}

<div style="color: red;">

    <h4>Ошибка выполнения:</h4>

    <pre>{{ exception_html }}</pre>

</div>

{% endif %}
```

3.3.4 Пример пользовательского шаблона HTML				
Подпись и дата				
Име. № дубл.				
Взам. ине. №				
Подпись и дата				
Име. №подл.				
<pre>jinja2  &lt;h3&gt;Уведомление о выполнении DAG: {{ ti.dag_id }}&lt;/h3&gt;  &lt;p&gt;&lt;strong&gt;Задача:&lt;/strong&gt; {{ ti.task_id }}&lt;/p&gt;  &lt;p&gt;&lt;strong&gt;Статус:&lt;/strong&gt; {{ ti.state }}&lt;/p&gt;  &lt;p&gt;&lt;strong&gt;Время выполнения:&lt;/strong&gt; {{ df_now }}&lt;/p&gt;  &lt;p&gt;&lt;strong&gt;Ссылка на логи:&lt;/strong&gt; &lt;a href="{{ log_url }}"&gt;Перейти к логам&lt;/a&gt;&lt;/p&gt;  &lt;p&gt;&lt;strong&gt;Базовая ссылка:&lt;/strong&gt; {{ base_url }}&lt;/p&gt;  {% if ti.state == 'failed' %}  &lt;div style="color: red;"&gt;      &lt;h4&gt;Ошибка выполнения:&lt;/h4&gt;      &lt;pre&gt;{{ exception_html }}&lt;/pre&gt;  &lt;/div&gt;  {% endif %}</pre>				

3.3.5 Практические рекомендации

- 1) Безопасность учетных данных: Для хранения SMTP пароля рекомендуется использовать секреты Airflow или переменные окружения
- 2) Тестирование подключения: Используйте утилиту airflow config get-value smtp smtp\_host для проверки настроек
- 3) Шаблоны по умолчанию: Polyflow включает базовые шаблоны в директории source/df/common/email

3.4 Операции группы безопасность

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

3.4.1 Список пользователей

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

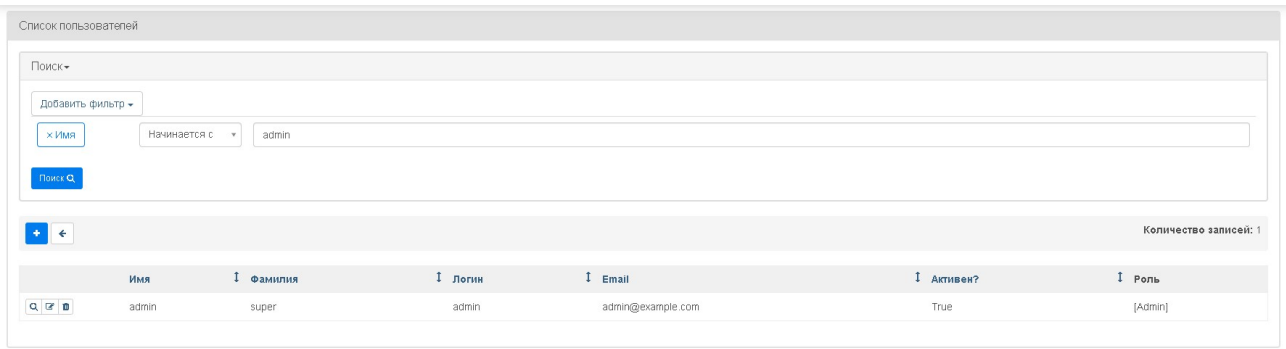


Рисунок 3. Список пользователей

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

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

Показать пользователя

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 посчитать 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

Сбросить пароль

Сброс пароля

Рисунок 4. Информация по пользователю

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

Редактировать 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

Admin

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):

Подпись и дата

Инв. № дубл.

Взам. инв. №

Подпись и дата

Инв. № подл.

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-мэйл пользователя. 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

### 3.4.2 Список ролей

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

Список ролей

Поиск

Действия

Количество записей: 5

<input type="checkbox"/>	Название	Права
<input type="checkbox"/>	Admin	[can delete on Connections, can read on Connections, can edit on Connections, can create on Connections, can read on DAGs, can edit on DAGs, can delete on DAGs, can read on DAG Runs, can read on Task Instances, can edit on Task Instances, can delete on DAG Runs, can edit on DAG Runs, can read on Audit Logs, can read on ImportError, can delete on Pools, can read on Pools, can edit on Pools, can create on Pools, can read on Providers, can delete on Variables, can read on Variables, can edit on Variables, can create on Variables, can read on XComs, can read on DAG Code, can read on Configurations, can read on Plugins, can read on Roles, can read on Permissions, can delete on Roles, can edit on Roles, can create on Roles, can delete on Users, can create on Users, can edit on Users, can delete on Users, can edit on Passwords, can read on Passwords, can edit on My Password, can read on My Password, can edit on My Profile, can read on My Profile, can read on User Stats Chart, can read on View Menus, can read on Permission Views, can get on MenuApi, menu access on DAG Runs, can read on Jobs, menu access on Jobs, menu access on Audit Logs, menu access on Variables, can create on Task Instances, can delete on Task Instances, menu access on Task Instances, can read on Task Reschedules, menu access on Task Reschedules, can read on Triggers, menu access on Triggers, menu access on Configurations, menu access on Connections, can read on SLAMisses, menu access on SLAMisses, menu access on Plugins, menu access on Providers, menu access on Pools, can create on XComs, can delete on XComs, menu access on XComs, can list on AppBuilderCodeEditorView, can create on AppBuilderCodeEditorView, menu access on AppBuilderCodeEditorView, menu access on Security, menu access on List Users, menu access on User's Statistics, menu access on Base Permissions, menu access on ViewsMenus, menu access on Permission on ViewsMenus, menu access on DAG Browse, menu access on DAG Dependencies, menu access on DAG Dependencies, menu access on Admin, menu access on DAGs Code Editor, menu access on Docs, menu access on Documentation, can read on DAG Dependencies, can read on Task Logs, can read on Website, can create on DAG Runs]
<input type="checkbox"/>	Public	[]
<input type="checkbox"/>	Viewer	[can read on DAGs, can read on DAG Runs, can read on Task Instances, can read on Audit Logs, can read on ImportError, can read on XComs, can read on DAG Code, can read on Plugins, can edit on My Password, can read on My Password, can edit on My Profile, can read on My Profile, menu access on DAG Runs, can read on Jobs, menu access on Jobs, menu access on Audit Logs, menu access on Task Instances, can read on SLAMisses, menu access on SLAMisses, menu access on Plugins, menu access on Browse, menu access on DAG Dependencies, menu access on Docs, menu access on Documentation, can read on DAG Dependencies, can read on Task Logs, can read on Website]
<input type="checkbox"/>	User	[can read on DAGs, can edit on DAGs, can delete on DAGs, can read on DAG Runs, can read on Task Instances, can edit on Task Instances, can delete on DAG Runs, can edit on DAG Runs, can read on Audit Logs, can read on ImportError, can read on XComs, can read on DAG Code, can read on Plugins, can edit on My Password, can read on My Password, can edit on My Profile, can read on My Profile, menu access on DAG Runs, can read on Jobs, menu access on Jobs, menu access on Audit Logs, menu access on Task Instances, can read on SLAMisses, menu access on SLAMisses, menu access on Plugins, menu access on Browse, menu access on DAG Dependencies, menu access on Docs, menu access on Documentation, can read on DAG Dependencies, can read on Task Logs, can read on Website, can create on DAG Runs]
<input type="checkbox"/>	Op	[can delete on Connections, can read on Connections, can edit on Connections, can create on Connections, can read on DAGs, can edit on DAGs, can delete on DAGs, can read on DAG Runs, can read on Task Instances, can edit on Task Instances, can delete on DAG Runs, can edit on DAG Runs, can read on Audit Logs, can read on ImportError, can read on XComs, can read on DAG Code, can read on Plugins, can edit on My Password, can read on My Password, can edit on My Profile, can read on My Profile, menu access on DAG Runs, can read on Jobs, menu access on Jobs, menu access on Audit Logs, menu access on Task Instances, can read on SLAMisses, menu access on SLAMisses, menu access on Plugins, menu access on Pools, can delete on Variables, can read on Variables, can edit on Variables, can create on Variables, can read on XComs, can read on DAG Code, can read on Configurations, can read on Plugins, can edit on My Password, can read on My Password, can edit on My Profile, can read on My Profile, menu access on DAG Runs, can read on Jobs, menu access on Jobs, menu access on Audit Logs, menu access on Task Instances, can read on SLAMisses, menu access on SLAMisses, menu access on Plugins, menu access on Pools, can delete on XComs, menu access on XComs, menu access on Browse, menu access on DAG Dependencies, menu access on Admin, menu access on Docs, menu access on Documentation, can read on DAG Dependencies, can read on Task Logs, can read on Website, can create on DAG Runs]

Рисунок 7. Список ролей

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

Инв. №подл.	Подпись и дата	Взам. инв. №	Инв. №дубл.	Подпись и дата	.РЭ					Лист
Изм	Лист	№ докум.	Подп.	Дата						46

Показать роль	
<div> <div>Detail</div> <div>Список пользователей</div> </div>	
Название	Viewer
Права	[can read on DAGs, can read on DAG Runs, can read on Task Instances, can read on Audit Logs, can read on ImportError, can read on xComs, can read on DAG Code, can read on Plugins, can edit on My Password, can read on My Password, can edit on My Profile, can read on My Profile, menu access on DAG Runs, can read on Jobs, menu access on Jobs, menu access on Audit Logs, menu access on Task Instances, can read on SLA Misses, menu access on SLA Misses, menu access on Plugins, menu access on Browse, menu access on DAG Dependencies, menu access on Docs, menu access on Documentation, can read on DAG Dependencies, can read on Task Logs, can read on Website]
<div> <div>←</div> </div>	

### Рисунок 8. Информация по правам роли

Редактировать роль

Detail

Список пользователей

+

←

Количество записей: 1

Имя	Фамилия	Логин	Email	Активен?	Роль
<div><div>🔍</div><div>🗨</div><div>🗑</div></div> user	one	user1	user1@example.com	True	[User]

### Рисунок 9. Информация по пользователям роли

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

Редактировать роль

Detail

Список пользователей

Название \*

Viewer

\* can read on DAGs

\* can read on DAG Runs

\* can read on Task Instances

\* can read on Audit Logs

\* can read on ImportError

\* can read on XComs

\* can read on DAG Code

\* can read on Plugins

\* can edit on My Password

\* can read on My Password

\* can edit on My Profile

\* can read on My Profile

\* menu access on DAG Runs

\* can read on Jobs

\* menu access on Jobs

\* menu access on Audit Logs

\* menu access on Task Instances

\* can read on SLA Misses

\* menu access on SLA Misses

\* menu access on Plugins

\* menu access on Browse

\* menu access on DAG Dependencies

\* menu access on Docs

\* menu access on Documentation

\* can read on DAG Dependencies

\* can read on Task Logs

\* can read on Website

Права

Сохранить & ←

### Рисунок 10. Редактирование роли

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

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

**Рисунок 10. Редактирование роли**

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

.РЭ

Лист

47

### 3.4.3 Статистика пользователя

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

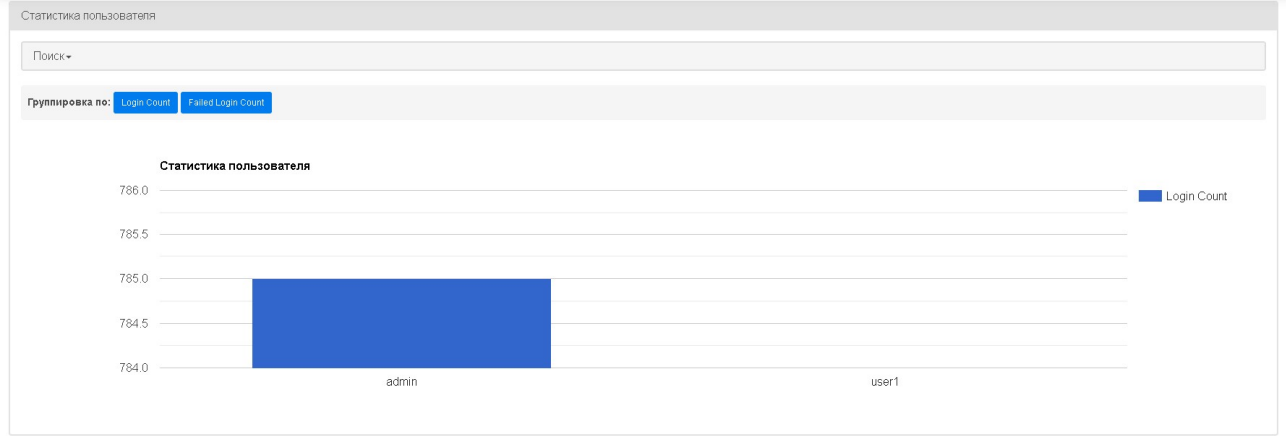


Рисунок 11. Статистика пользователя

### 3.4.4 Основные разрешения

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

Список базовых прав доступа	
Поиск -	
←	
Количество записей: 8	
Название	1
can_delete	
can_read	
can_edit	
can_create	
can_get	
menu_access	
can_list	
menu_access	

Рисунок 12. Список базовых прав доступа

### 3.4.5 Отображения/меню

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

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



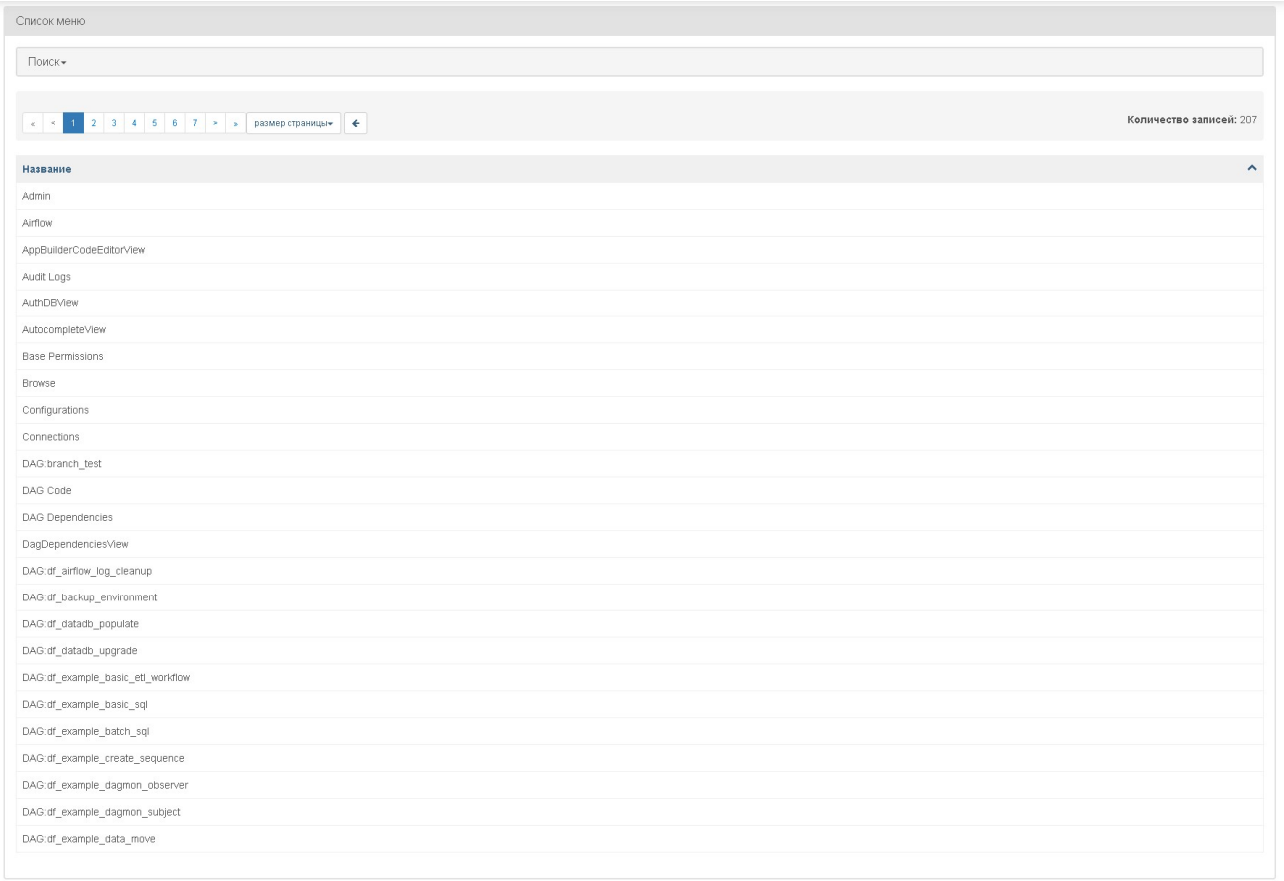


Рисунок 13. Список меню

3.4.6 Право на просмотр/меню

Форма отображает список прав на просмотр/меню (Рисунок 14).

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

					.РЭ	Лист
						49
Изм	Лист	№ докум.	Подп.	Дата		

Список прав на просмотр/меню

Поиск

<

1

2

3

4

5

6

7

>

размер страницы

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

3.5 Операции группы Admin

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

Меню Admin содержит пункты:

3.5.1 Variables

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

Выберите файл | Файл не выбран | Импорт переменных

List Variable

Поиск

+

Действия

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.	

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

Рисунок 16. Редактирование переменной

Рисунок 17. Действия над переменными

### 3.5.2 Configuration

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

### 3.5.3 Connections

Предназначен для настройки подключений (Рисунок 18).

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

List Connection

Поиск

Действия

Количество записей: 49

	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

Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.

Описание

Хост

172.24.0.41

Схема

appdb

Логин

dwh

Пароль

Порт

5432

Дополнительно

(\*schema\*: "dwh")

Сохранить

Test

Рисунок 19. Редактирование подключения

3.5.4 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.	

Airflow Plugins

1. datflow\_api\_plugin

Атрибут	Значение
hooks	[]
executors	[]
macros	[]
admin_views	[]
flask_blueprints	[<flask.blueprints.Blueprint object at 0x7fa3efe48760>]
menu_links	[]
appbuilder_views	[{'category': 'Docs', 'name': 'Dataflow API OpenAPI', 'view': <view.AppView object at 0x7fa3efe52310>}]
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	\$PLUGINS_FOLDER/views/dataflow_api/view.py

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	\$PLUGINS_FOLDER/common/setup.py

Рисунок 21. Информация о плагинах (продолжение)

3. editor\_plugin

Атрибут	Значение
hooks	[]
executors	[]
macros	[]
admin_views	[]
flask_blueprints	[<flask.blueprints.Blueprint object at 0x7fa3efdfcf40>]
menu_links	[]
appbuilder_views	[{'category': 'Admin', 'name': 'DAGs Code Editor', 'view': <airflow_code_editor.app_builder_view.AppBuilderCodeEditorView object at 0x7fa3efe1f7f0>}]
appbuilder_menu_items	[]
global_operator_extra_links	[]
operator_extra_links	[]
source	airflow-code-editor==5.1.5: EntryPoint(name='airflow_code_editor', value='airflow_code_editor.airflow_code_editor:CodeEditorPlugin', group='airflow.plugins')

Рисунок 22. Информация о плагинах (продолжение 2)

3.5.5 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

53

## 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

### 3.5.6 Pools

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

List Pool

Поиск

Действия

←

Количество записей: 1

<input type="checkbox"/>	Pool	Slots	Running Slots	Queued Slots
<input type="checkbox"/>	<input checked="" type="checkbox"/> default_pool	128	0	0

Рисунок 24. Пул процессов

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

Edit Pool

Pool \*

default\_pool

Slots

128

Description

Default pool

Сохранить

←

Рисунок 25. Редактирование данных пула процессов

### 3.5.6.1 Масштабирование и управление задачами через пулы воркеров

Архитектура Polyflow предоставляет мощные механизмы для распределения задач между несколькими воркерами (исполнителями), что позволяет эффективно масштабировать вычислительную мощность и управлять нагрузкой системы.

3.5.6.2 Механизм выполнения задач

Polyflow поддерживает различные типы Executor - компонентов, отвечающих за распределение и выполнение задач:

Таблица 6. Типы Executor

Executor	Описание	Преимущества
LocalExecutor	Выполняет задачи в параллельных процессах на одном сервере	Простая настройка, подходит для средних нагрузок
CeleryExecutor	Распределяет задачи между несколькими воркерами через брокер сообщений (Redis/RabbitMQ)	Горизонтальное масштабирование, отказоустойчивость
KubernetesExecutor	Запускает каждую задачу в отдельном Pod в Kubernetes кластере	Максимальная изоляция, эластичное масштабирование

3.5.6.3 Настройка пулов воркеров (Pools)

Пул воркеров - это механизм ограничения параллельного выполнения определенных типов задач для предотвращения перегрузки системы и управления ресурсами.

Создание и настройка пулов через веб-интерфейс:

- 1) Перейдите в меню Admin → Pools
- 2) Нажмите кнопку "+" для создания нового пула
- 3) Задайте параметры:
  - Pool Name: Уникальное имя пула (например, heavy\_queries\_pool)
  - Slots: Количество параллельных задач в пуле (например, 3 для ресурсоемких операций)
  - Description: Описание назначения пула

Пример использования пула в DAG:

```
#python

from airflow import DAG

from airflow.operators.docker_operator import DockerOperator

from datetime import datetime


default_args = {
    'owner': 'data_team',
    'start_date': datetime(2024, 1, 1),
}

with DAG('example_pool_dag',
        default_args=default_args,
        schedule_interval='@daily') as dag:

    # Задача с ограничением через пул
    heavy_etl_task = DockerOperator(
        task_id='heavy_etl_processing',
        image='df_operator:latest',
        command='python /app/run_heavy_etl.py',
        pool='heavy_queries_pool', # Использование специального пула
        pool_slots=1, # Количество слотов, занимаемых задачей
        environment_variables={
            'AF_DWH_DB_SERVER': '{{ var.value.get("db_host") }}',
            'AF_DWH_DB_NAME': 'dwh_db'
        }
    )

    # Обычная задача без ограничений
    light_transform_task = DockerOperator(
        task_id='light_data_transform',
        image='df_operator:latest',
        command='python /app/run_light_transform.py',
        # Не указан pool - выполняется в default_pool
    )
```

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

Изм	Лист	№ докум.	Подп.	Дата

.РЭ



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

Сценарий	Настройка пула	Преимущество
Ограничение тяжелых ETL-задач	heavy_etl_pool с 2-3 слотами	Предотвращение перегрузки БД при параллельном выполнении ресурсоемких запросов
Контроль API-вызовов	external_api_pool с 5-10 слотами	Соблюдение лимитов RPS (requests per second) внешних API
Приоритизация задач	critical_pool (10 слотов), normal_pool (20 слотов)	Гарантированная производительность для критически важных процессов

- Просмотр загрузки пулов: В интерфейсе Polyflow в разделе Pools отображается количество занятых/свободных слотов
- Динамическое изменение: Размер пула можно изменять в runtime без перезапуска системы
- Интеграция с мониторингом: Метрики использования пулов доступны для интеграции с Prometheus/Grafana

- 1) Размер `default_pool`: Увеличьте количество слотов в пуле по умолчанию в соответствии с ресурсами сервера
- 2) Балансировка нагрузки: Создавайте специализированные пулы для разных типов задач (ETL, API-вызовы, файловые операции)
- 3) Резервирование ресурсов: Оставляйте 20-30% слотов свободными для внеплановых задач и обработки пиковой нагрузки.

3.5.6.7 Масштабирование в облачных средах

Механизмы пулов воркеров Polyflow являются фундаментом для эластичного горизонтального масштабирования в облачных средах. Для автоматизации этого процесса в продакшн-развертываниях рекомендуется использовать оркестраторы, такие как Kubernetes, в связке с механизмом Horizontal Pod Autoscaler (HPA). Данная конфигурация позволяет автоматически увеличивать количество подов-воркеров при росте нагрузки и уменьшать его при его снижении, обеспечивая отказоустойчивость и экономическую эффективность.

3.5.7 XComs

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

List XComs						
Поиск						
<div>23242526272829&gt;размер страницыДействия</div>						
Количество записей: 2815						
<input type="checkbox"/>	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 слание данных в таблицу <rule>	2022-04-20, 13:08:41	2022-04-20, 13:08:30	merge	df_datadb_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_datadb_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_datadb_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_datadb_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_datadb_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

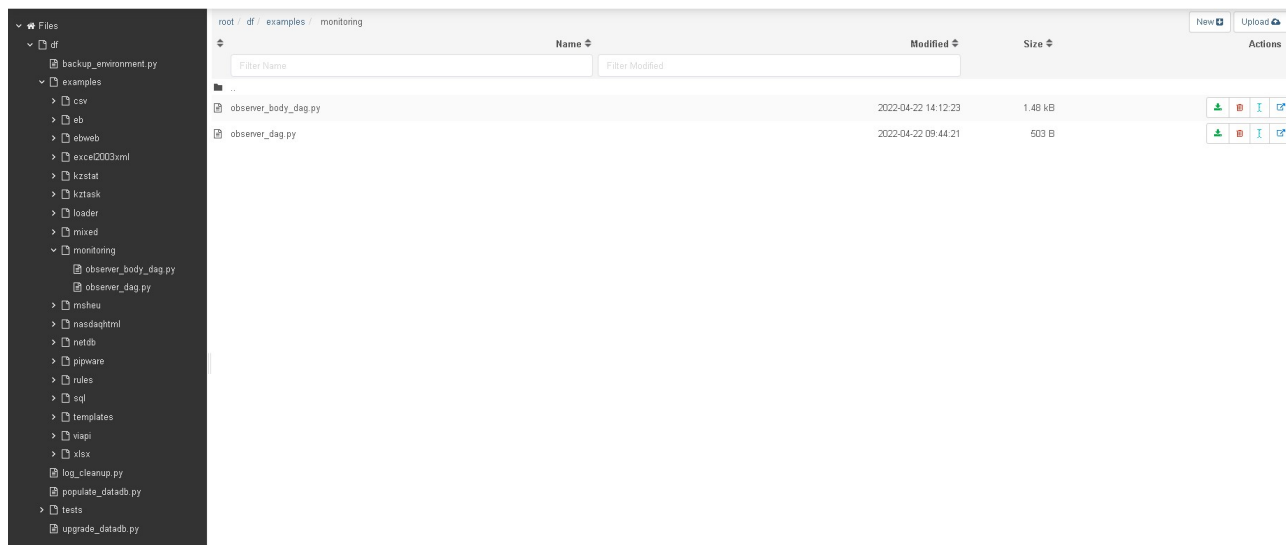
Рисунок 26. XComs

3.5.8 DAGs Code Editor

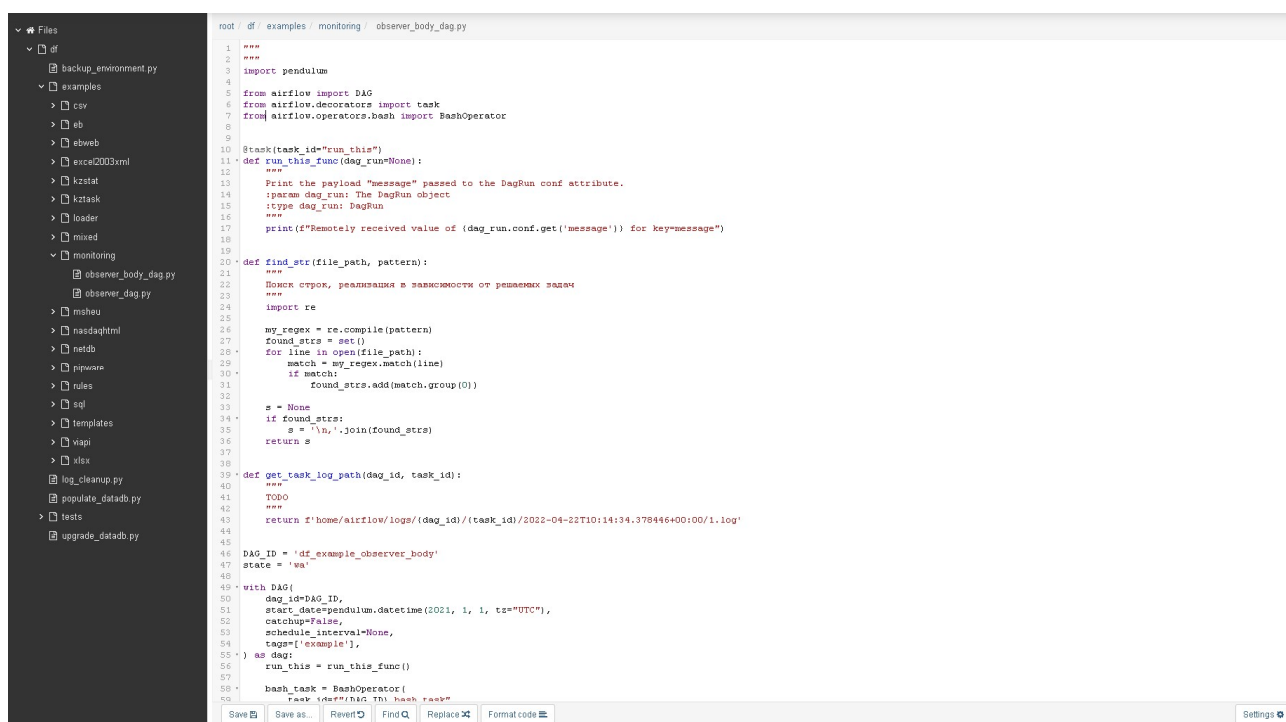
Содержит ссылку на страницу редактора DAG (Рисунок 27, Рисунок 28).

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

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



### Рисунок 27. Редактор DAG



### Рисунок 28. Редактор DAG с открытым файлом

### 3.6 Политики безопасности и шифрования

### 3.6.1 Шифрование передаваемых данных

Polyflow обеспечивает многоуровневое шифрование данных:

## Уровень 1: Веб-интерфейс

- HTTPS с SSL/TLS сертификатами

- Шифрование аутентификационных данных
- Защита сессий пользователей

#### Уровень 2: Внутренние коммуникации

- Шифрование трафика между компонентами

#### Уровень 3: Внешние подключения

- SSL/TLS для подключений к СУБД
- Шифрование API-вызовов к внешним системам
- Защита данных в транзите при работе с облачными провайдерами

### 3.6.2 Рекомендации по настройке

- 1) Сертификаты: Используйте доверенные SSL-сертификаты
- 2) Алгоритмы: Применяйте современные криптографические алгоритмы
- 3) Обновление: Регулярно обновляйте сертификаты и ключи
- 4) Мониторинг: Контролируйте состояние шифрованных соединений

Инв. № подл.	Подпись и дата				Инв. № дубл.	Подпись и дата				Взам. инв. №	Подпись и дата				Инв. № подл.	Подпись и дата				Изм	Лист	№ докум.	Подп.	Дата	.РЭ	Лист	60

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

Изм	Лист	№ докум.	Подп.	Дата

- .P3*

	Лист
	61

Лист
61

7учт

---

61

7учт

---

61

- 7учт
- 61

5 Рекомендации по освоению

Основным источником информации, используемым при освоении Системы, является данное руководство.

Начинать работу с Системой следует со знакомства с разделами руководства «Подготовка к работе», «Описание операций».

Для обеспечения успешной работы пользователям необходимо обладать основными навыками работы с веб-приложениями, опубликованными в сети Интернет.

Инв. № подл.	Подпись и дата				Инв. № дубл.	Подпись и дата																									
Взам. инв. №					Инв. № дубл.																										
Лист					Инв. № дубл.																										
Изм					Лист					№ докум.					Подп.					Дата					.РЭ					Лист	
																				62											

### Лист регистрации изменений

[illegible]

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


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

Изм	Лист	№ докум.	Подп.	Дата


.РЭ

Лист
64