

Программное обеспечение

Система управления данными Polyflow

РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ

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

Аннотация

Настоящий документ является руководством пользователя системы управления данными Polyflow.

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

[illegible]

Содержание

| | |
|---|-----------|
| Введение | 5 |
| 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 |
| 2.2 Запуск системы..... | 10 |
| 2.2.1 Начало работы | 10 |
| 2.3 Порядок проверки работоспособности | 11 |
| 3 Описание операций | 13 |
| 3.1 Определения и сокращения Polyflow | 13 |
| 3.2 Работа с главной страницей приложения | 15 |
| 3.2.1 Список DAG (DAGs)..... | 15 |
| 3.2.2 Просмотр информации о DAG..... | 17 |
| 3.3 Дерево..... | 19 |
| 3.4 Запуск DAG на выполнение | 24 |
| 3.4.1 Запуск через главную страницу | 24 |
| 3.4.2 Запуск через любое представление | 25 |
| 3.4.3 Запуск через представление Дерево | 25 |
| 3.4.4 Запуск через представление График | 25 |
| 3.5 Просмотр журнала выполнения задачи | 26 |
| 3.5.1 Просмотр через представление Дерево..... | 26 |
| 3.5.2 Просмотр через представление График | 26 |
| 3.6 Безопасность | 26 |
| 3.7 Browse..... | 27 |
| 3.7.1 DAG Runs | 27 |
| 3.7.2 Jobs..... | 27 |
| 3.7.3 Audit Logs..... | 28 |
| 3.7.4 Task Instances | 28 |
| 3.7.5 Task Reschedule..... | 29 |
| 3.7.6 Triggers | 29 |
| 3.7.7 SLA Misses | 29 |

| | | | | | | | | | |
|----------------|------|--------------|-------|--------------|-----|----------------|--|--------------|------|
| Подпись и дата | | Име. № дубл. | | Взам. име. № | | Подпись и дата | | Име. № подл. | |
| | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | | | | Лист |
| | | | | | | | | | 3 |

| | |
|---|-----------|
| 3.7.8 DAG Dependencies..... | 30 |
| 3.8 Admin..... | 31 |
| 3.9 Docs..... | 32 |
| 4 Аварийные ситуации | 34 |
| 5 Рекомендации по освоению | 35 |

| | | | | | | | | | | | | | | | | | | | |
|--------------|----------------|------|----------|-------|--------------|----------------|--|--|--|--------------|--------------|----------------|--------------|---|--|--|--|--|------|
| Инв. № подл. | Подпись и дата | | | | Инв. № дубл. | Подпись и дата | | | | Взам. инв. № | Инв. № дубл. | Подпись и дата | Инв. № подл. | <div style="text-align: right; font-size: 24px; font-weight: bold;">.РЭ</div> | | | | | Лист |
| | | | | | | | | | | | | | | | | | | | 4 |
| | Изм. | Лист | № докум. | Подп. | | Дата | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

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:

| | | | | | | | | | | |
|-------------|----------------|--------------|-------------|----------------|---|--|--|--|--|------|
| Ине. №подл. | Подпись и дата | Взам. инв. № | Ине. №дубл. | Подпись и дата | <div style="text-align: right;">.РЭ</div> | | | | | Лист |
| | | | | | | | | | | 5 |
| | | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | | | | | | |

- 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.

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.

| | | | | | | | | | | |
|--------------|----------------|--------------|--------------|----------------|-----|--|--|--|--|------|
| Инев. №подл. | Подпись и дата | Взам. инв. № | Инев. №дубл. | Подпись и дата | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | | | | | Лист |
| | | | | | | | | | | 7 |

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

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

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. | Подпись и дата | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <tr> <td>Изм.</td> <td>Лист</td> <td>№ докум.</td> <td>Подп.</td> <td>Дата</td> <td rowspan="2">.РЭ</td> <td>Лист</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>9</td> </tr> </table> | | | | | | | | | | | | | Изм. | Лист | № докум. | Подп. | Дата | .РЭ | Лист | | | | | | 9 |
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | Лист | | | | | | | | | | | | | | | | | | | |
| | | | | | | 9 | | | | | | | | | | | | | | | | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

- | Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

- | Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

| Ине. № подл. | Подпись и дата | Взам. ине. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|
| | | | | |

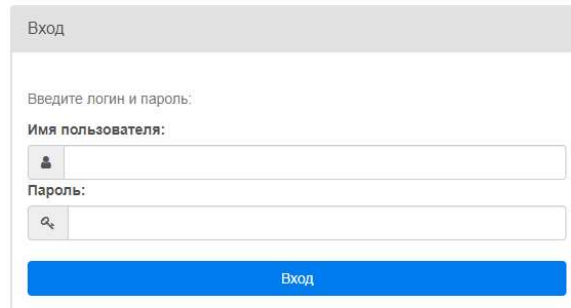


Рисунок 1. Начальная страница

2.3 Порядок проверки работоспособности

Для проверки работы Polyflow необходимо произвести аутентификацию в системе. После успешной аутентификации откроется главная страница приложения, со списком доступных процессов (Рисунок 2). Необходимо выбрать один из них и перейти на форму просмотра информации о выбранном процессе, кликнув на его название (Рисунок 3).

| | | | | | | | |
|--------------|------|----------|-------|------|---|----------------|--|
| | | | | | аутентификацию в системе. После успешной аутентификации откроется главная страница приложения, со списком доступных процессов (Рисунок 2). Необходимо выбрать один из них и перейти на форму просмотра информации о выбранном процессе, кликнув на его название (Рисунок 3). | | |
| | | | | | | Подпись и дата | |
| | | | | | | Инв. № дубл. | |
| | | | | | | Взам. инв. № | |
| | | | | | | Подпись и дата | |
| Инв. № подл. | | | | | | | |
| | | | | | .РЭ | Лист | |
| Изм. | Лист | № докум. | Подп. | Дата | | 11 | |

POLYFLOW

Список DAG

Безопасность

Browse

Admin

Docs

12:50 MSK (+03:00)

AS

DAGs

Всего145

Активный00

Приостановлено49

Filter DAGs by tag

Поиск DAG

| DAG | Владелец | Запуски | Расписание | Последний запуск | Следующий запуск | Недавние задачи | Действия |
|---|----------|---|------------|----------------------|----------------------|---|--|
| <div><div></div><div>df_airflow_log_cleanup</div><div>system</div></div> | admin | <div><div></div><div></div><div></div><div></div></div> | @monthly | | 2022-04-01, 03:00:00 | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_backup_environment</div><div>system</div></div> | admin | <div><div></div><div></div><div></div><div></div></div> | None | | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_datadb_populate</div><div>system</div></div> | admin | <div><div></div><div>1</div><div></div><div></div></div> | None | 2022-04-20, 13:08:30 | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div>1</div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_datadb_upgrade</div><div>system</div></div> | admin | <div><div></div><div>1</div><div></div><div></div></div> | None | 2022-04-20, 13:08:04 | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div>4</div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_example_basic_etl_workflow</div><div>example</div></div> | airflow | <div><div></div><div></div><div></div><div></div></div> | None | | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_example_basic_sql</div><div>example</div></div> | airflow | <div><div></div><div></div><div></div><div></div></div> | None | | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_example_batch_sql</div><div>example</div></div> | airflow | <div><div></div><div></div><div></div><div></div></div> | None | | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_example_create_sequence</div><div>example</div></div> | airflow | <div><div></div><div></div><div></div><div></div></div> | None | | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_example_dagmon_observer</div><div>dagmon</div><div>example</div><div>monitoring</div></div> | airflow | <div><div></div><div>5</div><div></div><div></div></div> | None | 2022-04-28, 16:34:56 | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div>5</div></div> | <div><div></div><div>2</div><div></div><div></div></div> |
| <div><div></div><div>df_example_dagmon_subject</div><div>dagmon</div><div>example</div><div>monitoring</div></div> | airflow | <div><div></div><div>4</div><div></div><div></div></div> | @once | 2022-04-28, 16:34:54 | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div>4</div></div> | <div><div></div><div></div></div> |
| <div><div></div><div>df_example_data_move</div><div>example</div></div> | airflow | <div><div></div><div>1</div><div></div><div>4</div></div> | None | 2022-04-20, 22:43:38 | | <div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div>1</div></div> | <div><div></div><div></div></div> |

Рисунок 2. Список доступных процессов

Будет открыта форма информации и управления выбранным DAG. По умолчанию открыта вкладка График (Рисунок 3).

| | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|---|--|--|--|------------------------|--|
| DAG: df_example_dagmon_subject | | | | | | | | | | success | | Расписание: @once | | Следующий запуск: None | |
| <div> <div>Дерево</div> <div>График</div> <div>Календарь</div> <div>Длительность задачи</div> <div>Попытки выполнения задачи</div> <div>Время выгрузки</div> <div>Gantt</div> <div>Подробности</div> <div>Код</div> </div> | | | | | | | | | | | | | | | |
| <div> <div>2022-04-28T16:34:55+00:00</div> <div>Запуски 25</div> <div>Запустить manual_2022-04-28T13:34:54.785604+00:00</div> <div>Макет Left > Right</div> <div>Обновить</div> <div>Найти задачу...</div> </div> | | | | | | | | | | | | | | | |
| <div> <div>BashOperator</div> <div>PythonOperator</div> <div>TriggerDagRunOperator</div> </div> | | | | | | | | | | <div> <div>queued</div> <div>running</div> <div>success</div> <div>failed</div> <div>up_for_retry</div> <div>up_for_reschedule</div> <div>upstream_failed</div> <div>skipped</div> <div>scheduled</div> <div>deferred</div> <div>no_status</div> </div> | | <div> <div>Автообновление</div> <div></div> </div> | | | |
| <div> <div>df_example_dagmon_subject.bash_task1</div> <div>df_example_dagmon_subject.worker_task</div> <div>df_example_dagmon_subject.bash_task2</div> <div>df_example_dagmon_subject.observer_trigger</div> </div> | | | | | | | | | | | | | | | |

Рисунок 3. Форма просмотра информации о процессе

| | | | | | | | | | | | | | | | | | |
|--------------|----------------|------|----------|--|--------------|--------------|----------------|--|-----|--|--------------|--|--|--|----|------|--|
| Ине. № подл. | Подпись и дата | | | | Ине. № дубл. | Взам. ине. № | Подпись и дата | | | | Ине. № подл. | | | | | Лист | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| Изм. | | Лист | № докум. | | Подп. | | Дата | | .РЭ | | | | | | 12 | | |

3 Описание операций

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 дубликатов, сохранение полученных результатов и |

| | | | |
|--------------|--------------|--------------|----------------|
| Ине. № подл. | Взам. инв. № | Ине. № дубл. | Подпись и дата |
| | | | |
| | | | |
| | | | |
| | | | |

| | |
|----------|---|
| | прочие ETL-процессы. |
| 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. |

| | | | | | | | | | | |
|-------------|----------------|--------------|-------------|----------------|-----|--|--|--|--|------|
| Име. №подл. | Подпись и дата | Взам. име. № | Име. №дубл. | Подпись и дата | | | | | | Лист |
| | | | | | | | | | | |
| | | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | | | | | 14 |

3.2 Работа с главной страницей приложения

После успешной аутентификации откроется главная страница приложения (Рисунок 4).

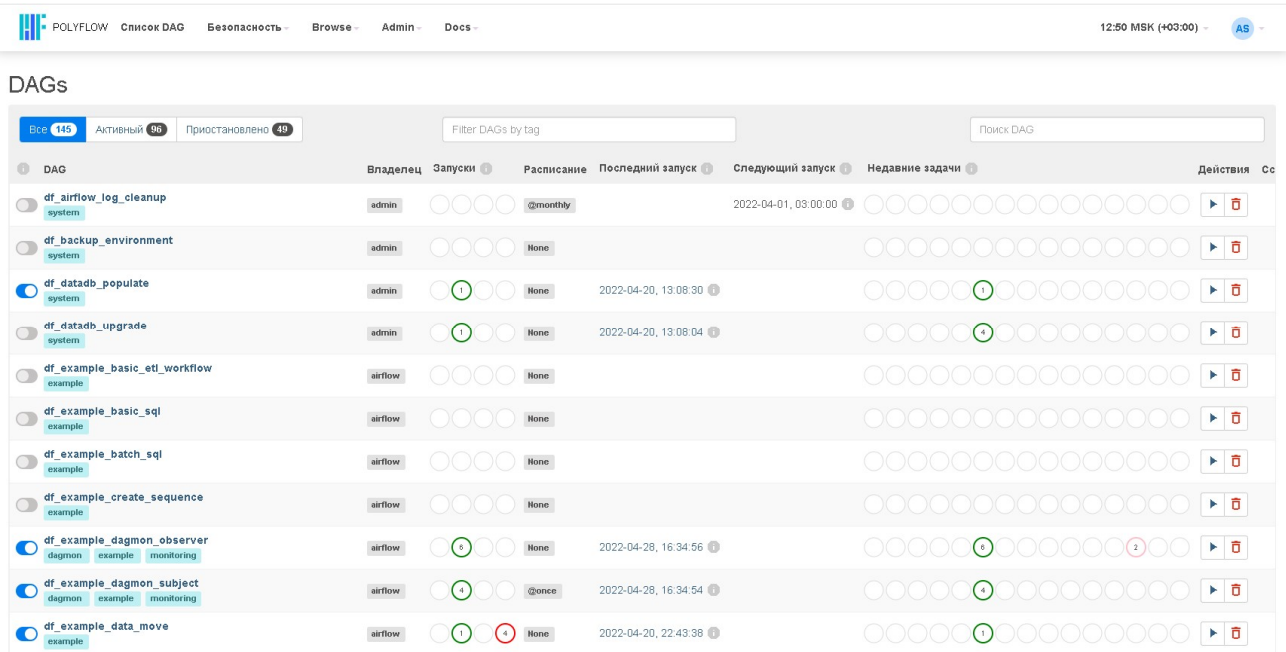


Рисунок 4. Главная страница приложения

В верхней части окна располагается меню, в нижней части - таблица со списком DAG, полем поиска DAG и элементами для управления ими. Через пункты верхнего меню доступны представления для управления и отображения различной информации по DAG (Рисунок 5).



Рисунок 5. Пункты верхнего меню

3.2.1 Список DAG (DAGs)

Представление Список DAG (DAGs) – страница, загружаемая при входе в приложение (Рисунок 6).

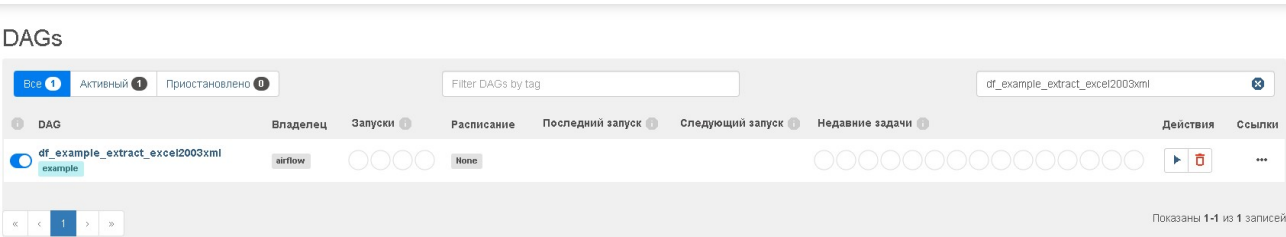
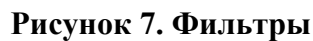


Рисунок 6. Список DAG

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

Имеется возможность задать фильтры по состоянию, тэгам или названию DAG (Рисунок 7).



- в первом столбце расположены переключатели, позволяющие приостановить или снять с паузы DAG (исключает или включает запуск DAG планировщиком) (Рисунок 8)



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

- Действия – действия над DAG (запуск, запуск с конфигурацией или удаление DAG);
- Ссылки - быстрые переходы на другие формы и представления DAG (Рисунок 9).



Рисунок 9. Ссылки быстрых переходов

3.2.2 Просмотр информации о DAG

При клике в таблице на название DAG будет открыта форма информации и управления выбранным DAG. По умолчанию открыта вкладка График (Рисунок 10).

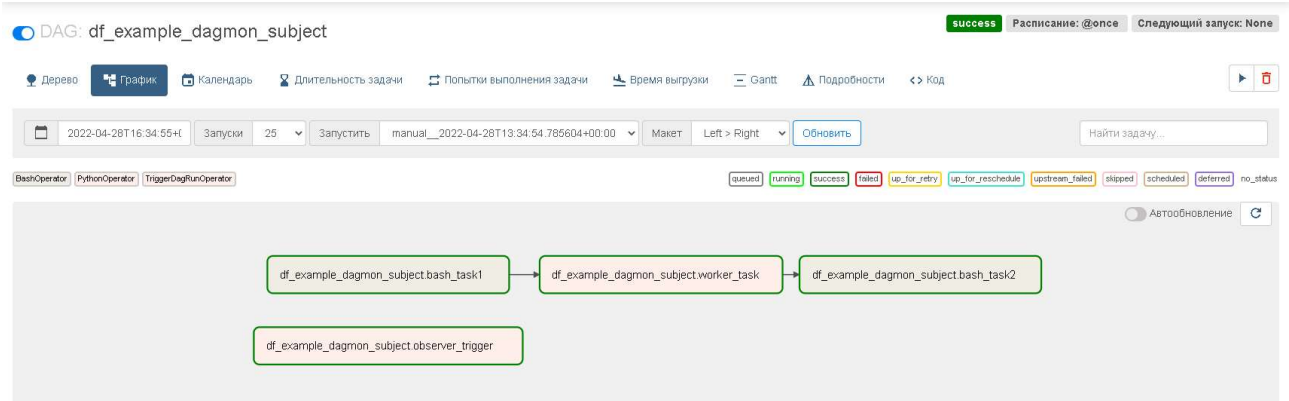


Рисунок 10. Представление «График»

3.2.2.1 График

Визуализация списка задач в виде графа с учетом их зависимостей и текущее их состояние

По клику на задачу в графе доступна модальная форма с дополнительными представлениями и действиями над экземпляром задачи (Рисунок 11):

| | | | | | | |
|--------------|----------------|----------|-------|------|------|----|
| Инв. № подл. | Подпись и дата | | | | Лист | |
| | Инв. № дубл. | | | | | |
| | Взам. инв. № | | | | | |
| | Подпись и дата | | | | | |
| Инв. № подл. | | | | | .РЭ | 17 |
| Изм. | Лист | № докум. | Подп. | Дата | | |

```
graph LR; A[df_example_dagmon_subject.bash_task1] --> B[df_example_dagmon_subject.worker_task]; B --> C[df_example_dagmon_subject.bash_task2]; D[df_example_dagmon_subject.observer_trigger];
```

Рисунок 10. Представление «График»

3.2.2.1 График

Визуализация списка задач в виде графа с учетом их зависимостей и текущее их состояние

По клику на задачу в графе доступна модальная форма с дополнительными представлениями и действиями над экземпляром задачи (Рисунок 11):

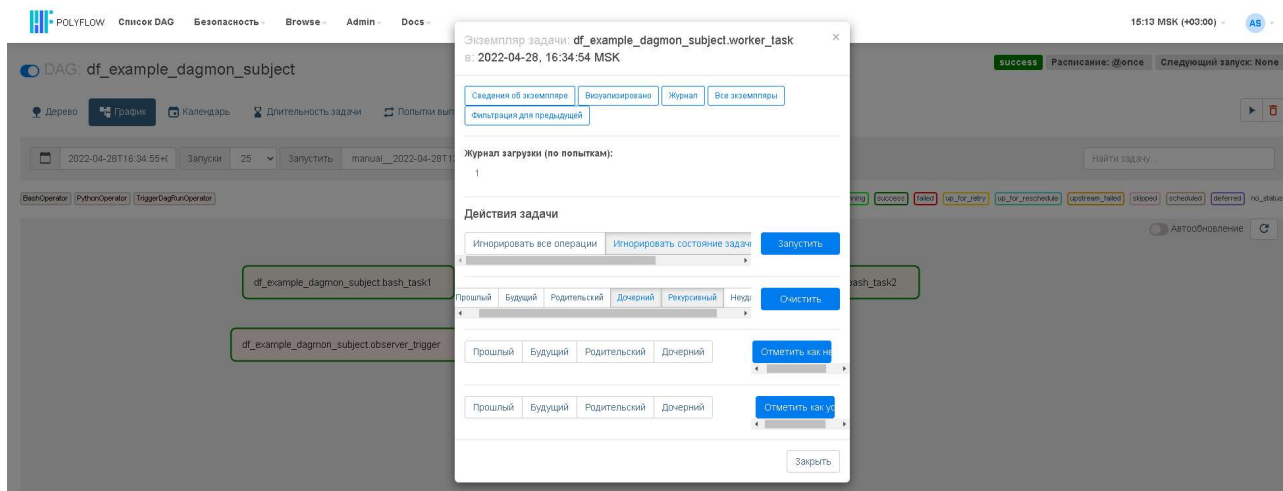


Рисунок 11. Дополнительные действия над задачей

- Сведения об экземпляре — подробная информация о задаче (атрибуты, значения, шаблоны);
- Визуализировано — метаданные задачи, после ее создания;
- Журнал — журнал выполнения экземпляра задачи;
- Все экземпляры — информация по выполнению экземпляров задачи;
- Фильтрация для предыдущей – отображает график до текущей задачи;
- Журналы загрузки (по попыткам) - журналы доступные для скачивания.

Запустить — ручной запуск задачи, с возможностью игнорирования зависимостей и текущего состояния задачи. Список доступных опций:

- Игнорировать все операции;
- Игнорировать состояние задачи;
- Игнорировать задачи.

Очистить - удаляет экземпляр задачи из базы данных метаданных. Это один из способов вручную повторно запустить задачу (и любые последующие задачи, если выбрано). Список доступных опций:

- Прошлый;

| | | | | | | | | | | |
|----------------|--------------|--------------|----------------|--------------|-----|--|--|--|--|------|
| Подпись и дата | Име. № дубл. | Взам. инв. № | Подпись и дата | Име. № подл. | .РЭ | | | | | Лист |
| | | | | | | | | | | 18 |
| | | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | | | | | | |

- Будущий;
- Родительский;
- Дочерний;
- Рекурсивный;
- Неудачный.

Отметить как неуспешный — изменяет статус задачи на неуспешный. Это обновит базу данных метаданных и остановит выполнение последующих задач, если таким образом определены зависимости в DAG. Список доступных опций:

- Прошлый;
- Будущий;
- Родительский;
- Дочерний.

Отметить как успешный — изменяет статус задачи на успешный. Это обновит базу данных метаданных и позволит запустить выполнение последующих задач, если таким образом определены зависимости в DAG. Список доступных опций:

- Прошлый;
- Будущий;
- Родительский;
- Дочерний.

3.3 Дерево

Древовидное представление DAG (Рисунок 12). Позволяет быстро увидеть, время выполнения различных этапов DAG, и определить статус отдельных задач. Каждый столбец представляет запуск DAG, а каждый квадрат - экземпляр задачи в этом запуске DAG. Экземпляры задач имеют цветовую маркировку в соответствии с их статусом.

| | | | | | | | | | | |
|--------------|----------------|--------------|--------------|----------------|--|--|--|--|--|------|
| Име. № подл. | Подпись и дата | Взам. инв. № | Име. № дубл. | Подпись и дата | | | | | | Лист |
| | | | | | | | | | | 19 |
| Изм. | Лист | № докум. | Подп. | Дата | | | | | | |

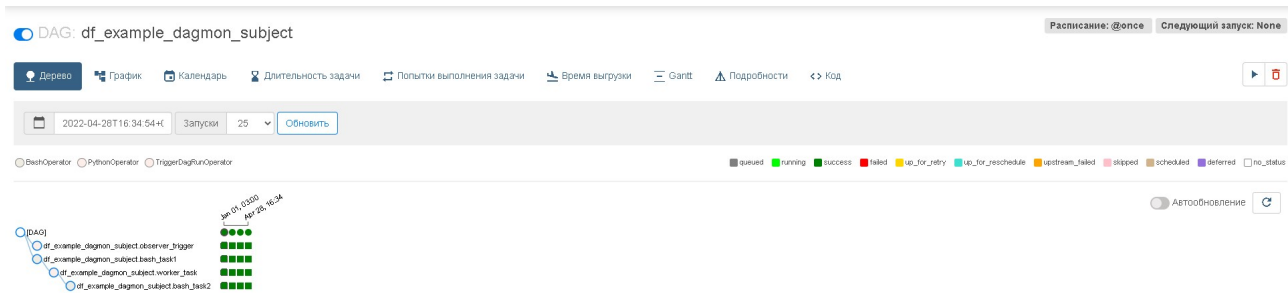


Рисунок 12. Представление «Дерево»

Клик по конкретному запуску DAG позволяет перезапустить DAG или перейти представлению График; по экземпляру задачи в дереве - открывает форму с теми же дополнительными представлениями и действиями, которые описаны в разделе «График» выше, и кнопкой перезапуска задачи (Рисунок 13).

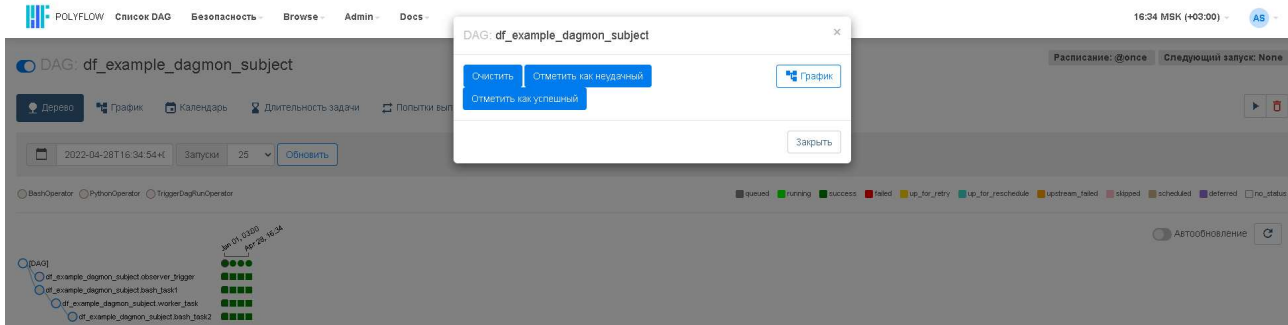


Рисунок 13. Дополнительные действия над экземпляром запуска DAG

3.3.1.1 Календарь

Представление календаря (Рисунок 14). Показывает статус выполнения DAG, наложенный на календарь. Статусы представлены цветом. Если в один и тот же день было несколько запусков DAG с разными статусами (например, один неудачный, другой успешный), цвет будет градиентом между зеленым (успешно) и красным (неудачно). При клике на день будет открыто представление «Дерево».

| | | | | |
|--------------|----------------|--------------|--------------|----------------|
| Ине. № подл. | Подпись и дата | Взам. инв. № | Ине. № дубл. | Подпись и дата |
|--------------|----------------|--------------|--------------|----------------|

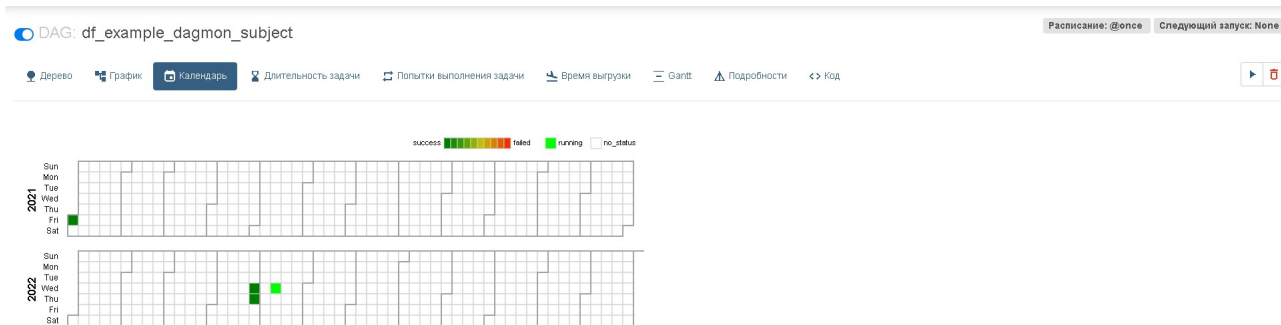


Рисунок 14. Представление «Календарь»

3.3.1.2 Длительность задачи

Показывает линейный график продолжительности каждой задачи во времени (Рисунок 15).



Рисунок 15. Представление «Длительность задачи»

3.3.1.3 Попытки выполнения задачи

Показывает линейный график количества попыток для каждой задачи в DAG с течением времени (Рисунок 16).

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

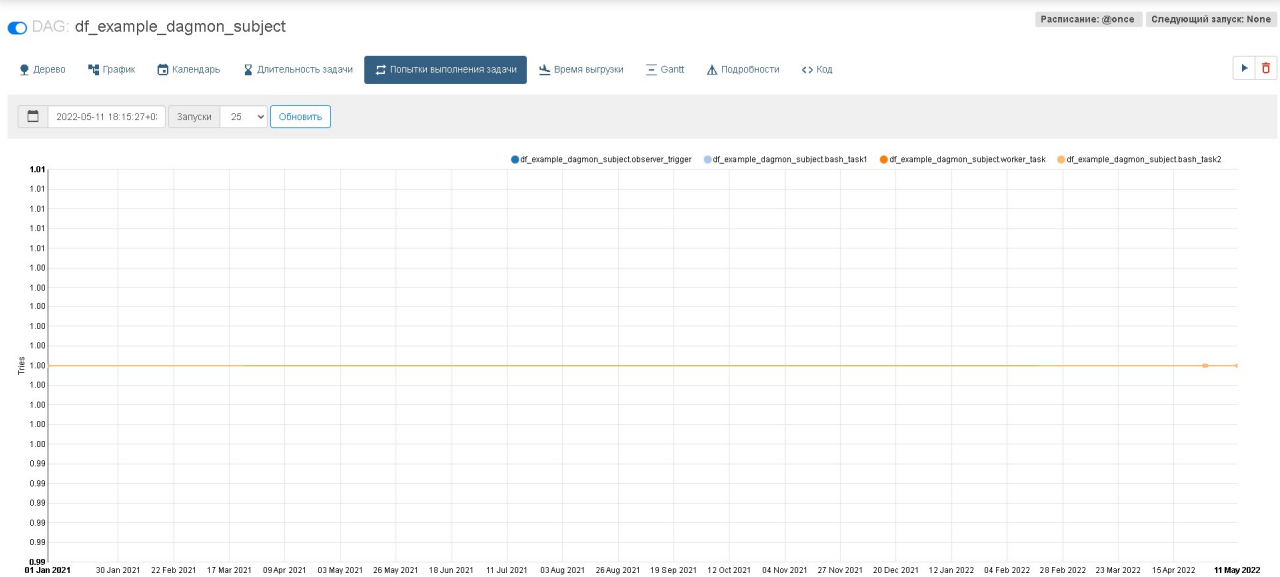


Рисунок 16. Представление «Попытки выполнения задачи»

3.3.1.4 Время выгрузки

Показывает линейный график времени суток, когда каждая задача была запущена с течением времени (Рисунок 17).

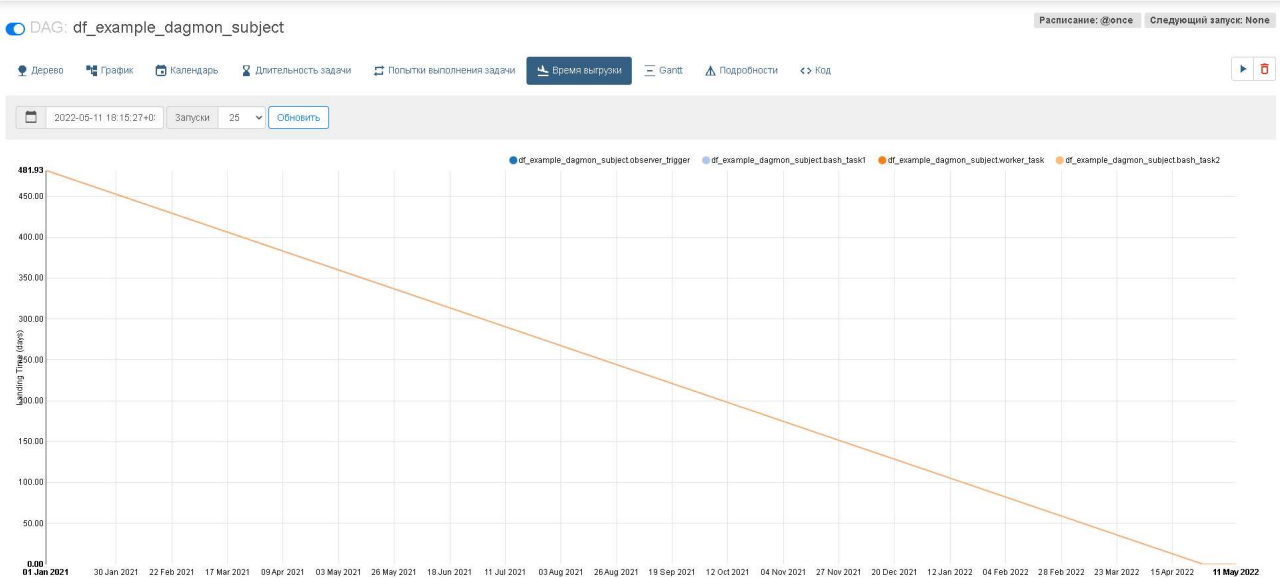


Рисунок 17. Представление «Время выгрузки»

3.3.1.5 Gantt

Показывает диаграмму Ганта с продолжительностью каждой задачи для выбранного прогона DAG (Рисунок 18).

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

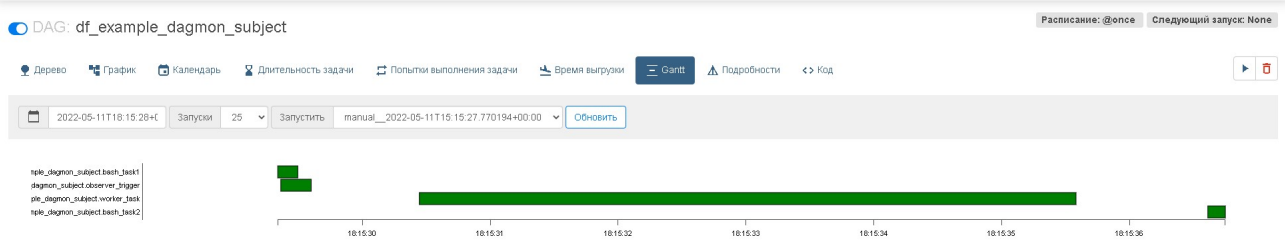


Рисунок 18. Представление «Gantt»

3.3.1.6 Подробности

Подробная информация о DAG (Рисунок 19).

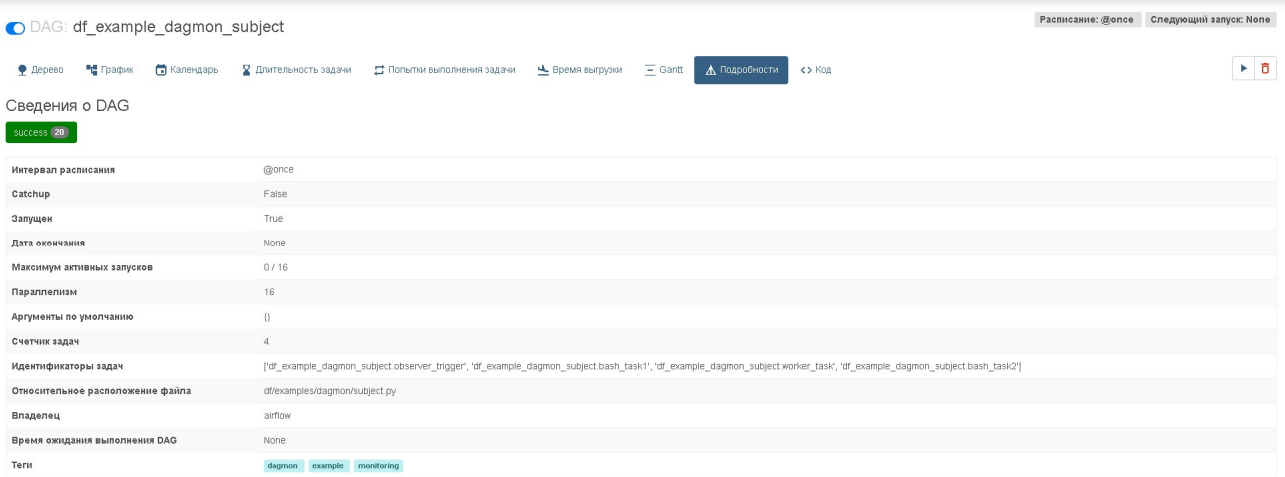


Рисунок 19. Представление «Подробности»

3.3.1.7 Код

В представлении кода показан код, который используется для генерации DAG (Рисунок 20). Данное представление — это быстрый способ получения информации о том, что происходит в DAG.

| | | | | | | | |
|-------------|----------------|-------------|----------------|--------------|-------------|----------------|-------------|
| Ине. №подл. | Подпись и дата | Ине. №дубл. | Подпись и дата | Взам. инв. № | Ине. №дубл. | Подпись и дата | Ине. №подл. |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

3.5 Просмотр журнала выполнения задачи

Журнал выполнения задачи предоставляет дополнительную информацию, которая в большинстве случаев необходима для диагностики некорректной работы DAG (Рисунок 24).

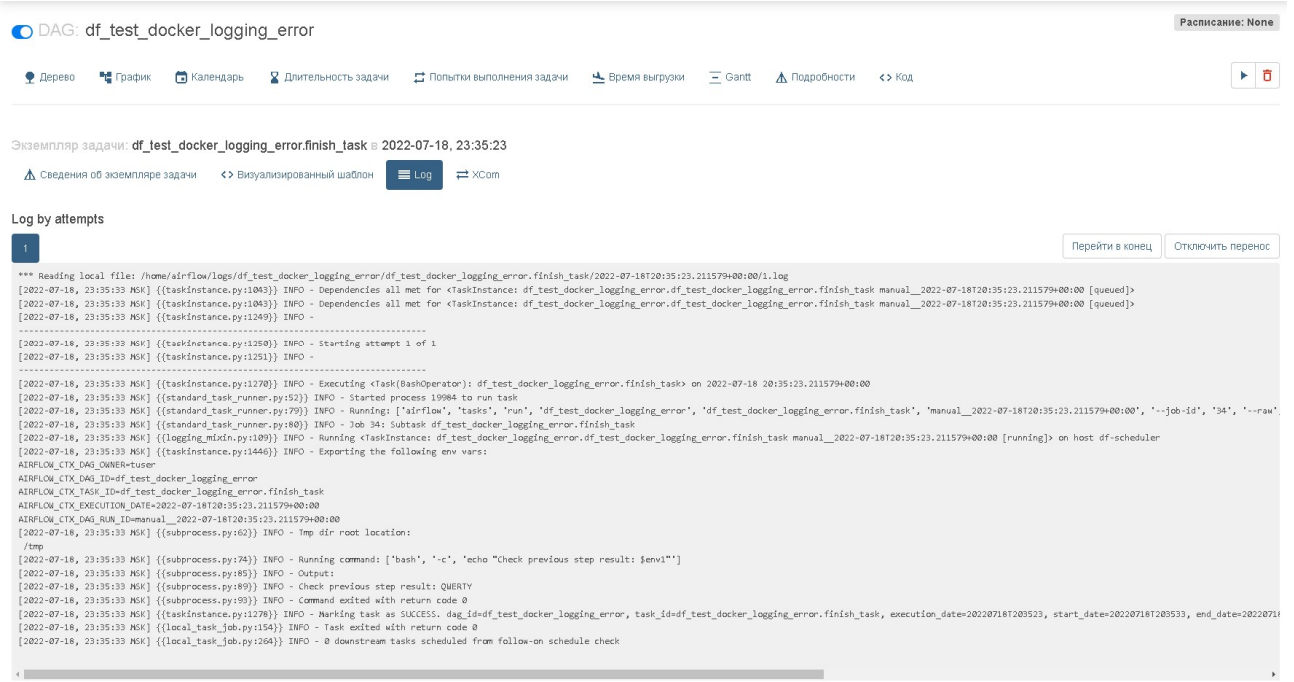


Рисунок 24. Журнал выполнения задачи

3.5.1 Просмотр через представление Дерево

Для просмотра журнала из представления «Дерево» необходимо кликнуть на экземпляре задачи и выбрать кнопку «Журнал».

3.5.2 Просмотр через представление График

Для просмотра журнала из представления «График» необходимо кликнуть на экземпляре задачи и кнопку «Журнал».

3.6 Безопасность

Описание приведено в документе «Руководство администратора» в разделе «Операции группы безопасность»

| | | | | | | | | | | | |
|--------------|----------------|--------------|--------------|----------------|---------------|--|--|--|--|------------|--|
| Име. № подл. | Подпись и дата | Взам. име. № | Име. № дубл. | Подпись и дата | <div>РЭ</div> | | | | | Лист 26 | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | | | | | | | |

3.7 Browse

Содержит ссылки на страницы, которые предоставляют дополнительную информацию и контроль выполнения запусков DAG и экземпляров задач для всех DAG в одном месте.

3.7.1 DAG Runs

Форма предназначена для просмотра информации о запусках DAG (Рисунок 25)

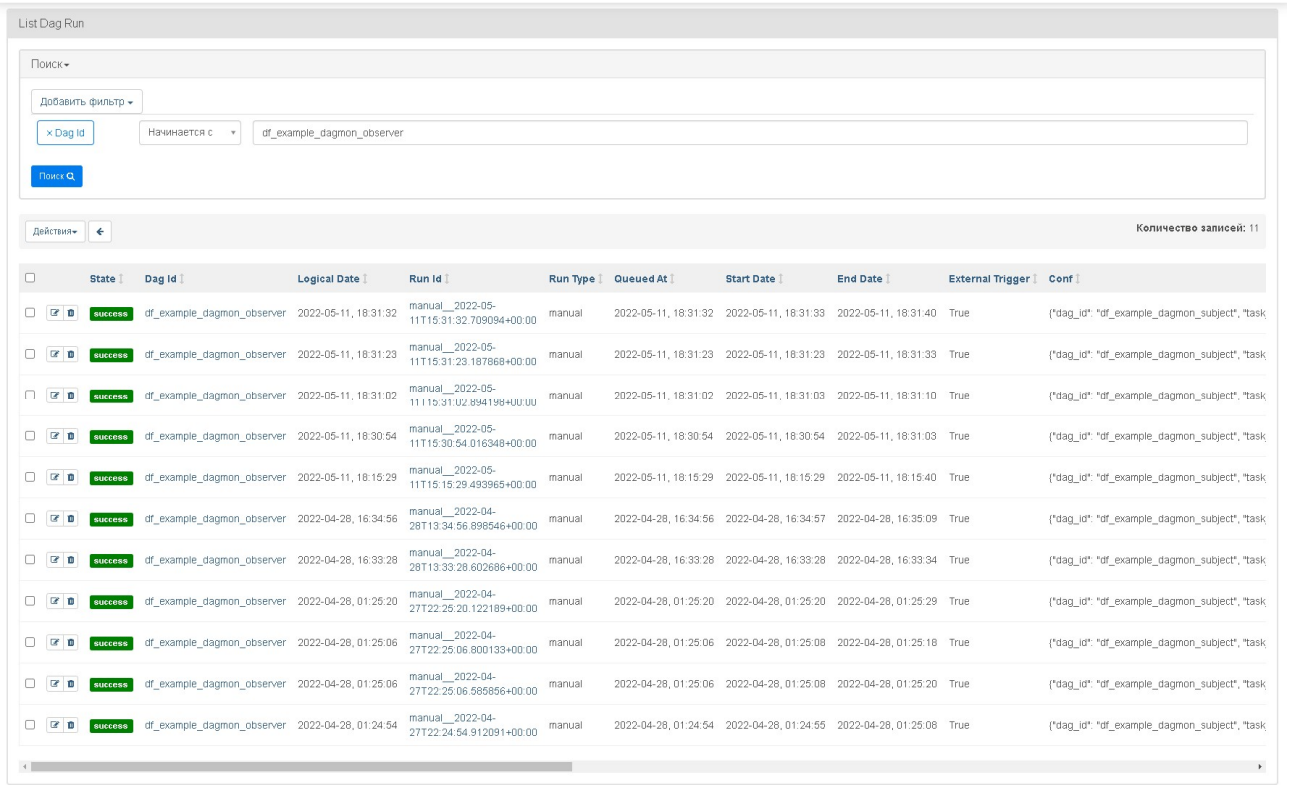


Рисунок 25. Список запусков DAG

3.7.2 Jobs

Форма предназначена для просмотра информации о выполненных заданиях (Рисунок 26). Сюда входят выполненные задачи, а также задания планировщика;

List Task Instance

Поиск*

1234567>

размер страницы

Действия

Количество записей: 4812

| <input type="checkbox"/> | State | Dag Id | Task Id | Run Id | Logical Date | Operator | Start Date |
|--------------------------|---|-------------------------------|--------------------------------------|--|----------------------|----------------|----------------------|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> success | of_datadb_upgrade | ddl | manual__2022-04-20T10:08:04.963483+00:00 | 2022-04-20, 13:08:04 | DockerOperator | 2022-04-20, 13:08:04 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> success | of_datadb_upgrade | callable_common | manual__2022-04-20T10:08:04.963483+00:00 | 2022-04-20, 13:08:04 | DockerOperator | 2022-04-20, 13:08:04 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> success | of_datadb_upgrade | callable_rules | manual__2022-04-20T10:08:04.963483+00:00 | 2022-04-20, 13:08:04 | DockerOperator | 2022-04-20, 13:08:04 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> success | of_datadb_upgrade | callable_json | manual__2022-04-20T10:08:04.963483+00:00 | 2022-04-20, 13:08:04 | DockerOperator | 2022-04-20, 13:08:04 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> success | of_datadb_populate | merge | manual__2022-04-20T10:08:30.217870+00:00 | 2022-04-20, 13:08:30 | DockerOperator | 2022-04-20, 13:08:30 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> success | of_example_loader_ftp_generic | of_example_loader_ftp_generic loader | manual__2022-04-20T12:49:22.078606+00:00 | 2022-04-20, 15:49:22 | DockerOperator | 2022-04-20, 15:49:22 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> failed | of_example_data_move | of_example_docker_data_move | manual__2022-04-20T14:38:55.769358+00:00 | 2022-04-20, 17:38:55 | DockerOperator | 2022-04-20, 17:38:55 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> failed | of_example_data_move | of_example_docker_data_move | manual__2022-04-20T15:39:48.567008+00:00 | 2022-04-20, 18:39:48 | DockerOperator | 2022-04-20, 18:39:48 |

Рисунок 28. Список экземпляров задач

3.7.5 Task Reschedule

Форма перепланированных задач (Рисунок 29): показывает список всех задач, которые были перепланированы.

List Task Reschedule

Поиск*

←

Количество записей: 0

Не найдено ни одной записи

Рисунок 29. Список перепланированных задач

3.7.6 Triggers

Форма предназначена для просмотра списка триггеров (Рисунок 30).

List Trigger

Поиск*

←

Количество записей: 0

Не найдено ни одной записи

Рисунок 30. Список триггеров

3.7.7 SLA Misses

Форма содержит информацию о всех экземплярах задач, для которых не соблюдаются SLA (Рисунок 31)



Рисунок 31. Список задач, для которых не соблюдаются SLA

3.7.8 DAG Dependencies

Форма показывает графическое представление любых зависимостей между DAG в вашей среде Airflow (Рисунок 32).



Рисунок 32. Зависимости DAG

3.7.8.1 Отслеживание происхождения данных (Data Lineage)

Polyflow предоставляет функционал отслеживания происхождения данных (Data Lineage), позволяя визуализировать зависимости между процессами и наборами данных.

Ключевые возможности:

1) Зависимости между DAG (DAG Dependencies)

- В интерфейсе Polyflow доступен граф зависимостей между всеми DAG.

- Система автоматически визуализирует связи, установленные через Dataset, TriggerDagRunOperator и ExternalTaskSensor.

- Раздел интерфейса: Browse → DAG Dependencies.

2) Управление зависимостями на основе наборов данных (Dataset-Driven Scheduling)

| | | | | | | |
|--------------|----------------|--------------|--------------|----------------|---|------------|
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | Лист 30 |
| | | | | | | |
| | | | | | | |
| Име. № подл. | Подпись и дата | Взам. инв. № | Име. № дубл. | Подпись и дата | <h3>3.7.8.1 Отслеживание происхождения данных (Data Lineage)</h3> <p>Polyflow предоставляет функционал отслеживания происхождения данных (Data Lineage), позволяя визуализировать зависимости между процессами и наборами данных.</p> <p>Ключевые возможности:</p> <p>1) Зависимости между DAG (DAG Dependencies)</p> <ul style="list-style-type: none">- В интерфейсе Polyflow доступен граф зависимостей между всеми DAG.- Система автоматически визуализирует связи, установленные через Dataset, TriggerDagRunOperator и ExternalTaskSensor.- Раздел интерфейса: Browse → DAG Dependencies. <p>2) Управление зависимостями на основе наборов данных (Dataset-Driven Scheduling)</p> | |

- Задачи могут быть объявлены как производители данных, указывая выходной Dataset в параметре outlets.
- DAG может быть запланирован на обновление определенного Dataset, что создает прозрачную связь "производитель-потребитель".
- Airflow/Polyflow автоматически отслеживает и отображает эти связи в графе зависимостей.

Пример настройки зависимости через Dataset:

```
#python
from airflow import Dataset
from airflow.decorators import task, dag

# Объявляем Dataset
report_dataset = Dataset("s3://bucket/data/report.csv")

@task(outlets=[report_dataset])
def generate_report():
    # Задача, которая создает данные
    ...

@dag(schedule=[report_dataset], start_date=datetime(2023, 1, 1))
def consuming_dag():
    # Этот DAG будет запущен после обновления report_dataset
    ...
```

Этот функционал обеспечивает базовое, но эффективное отслеживание происхождения данных на уровне графа процессов и их зависимостей.

3.8 Admin

Описание приведено в документе «Руководство администратора» в разделе «Операции пункта меню Admin»

| | |
|----------------|----------------|
| Ине. № подл. | Подпись и дата |
| Взам. инв. № | Ине. № дубл. |
| Подпись и дата | |
| Ине. № подл. | |

3.9 Docs

- Документация – переход на веб-страницу с документацией
- Веб-сайт Polyflow – переход на веб-страницу вендора.

В правой части формы отображаются: текущее время сервера, кнопка с пунктами меню «Ваш профиль» и «Выйти» (Рисунок 33).

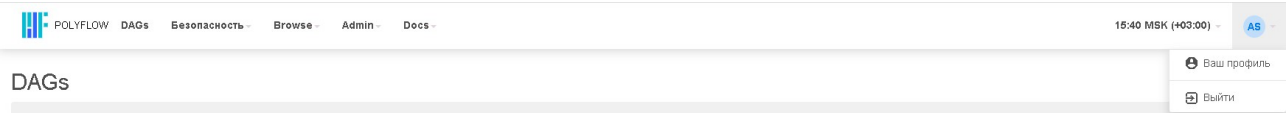


Рисунок 33. Пункты меню профиля



Рисунок 34. Данные пользователя

Через форму пользовательских данных (Рисунок 34) доступны переходы на форму сброса пароля (Рисунок 35) и форму редактирования отображаемых данных пользователя (Рисунок 36).

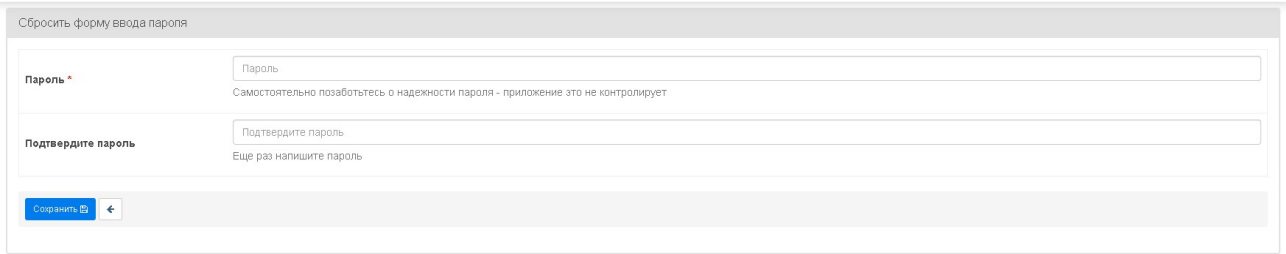


Рисунок 35. Сброс пароля пользователя

| | | |
|-------------|----------------|--|
| Ине. №подл. | Подпись и дата | |
| | Ине. №дубл. | |
| | Взам. инв. № | |
| | Подпись и дата | |

Edit User Information

Имя *

admin

Выводить имя пользователя

Фамилия *

super

Выводить фамилию

Сохранить

←

Рисунок 36. Редактирование данных пользователя

| | | | | | | | | | | | | |
|--------------|----------------|----------|-------|------|--------------|--------------|----------------|------|--|--|--|--|
| Инв. № подл. | Подпись и дата | | | | Взам. инв. № | Инв. № дубл. | Подпись и дата | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | | | Лист | | | | |
| | | | | | | | | 33 | | | | |

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.

| | | | | | | | | | |
|--------------|----------------|--------------|--------------|----------------|------|--|--|--|--|
| Ине. № подл. | Подпись и дата | Взам. инв. № | Ине. № дубл. | Подпись и дата | | | | | |
| Изм. | Лист | № докум. | Подп. | Дата | .РЭ | | | | |
| | | | | | Лист | | | | |
| | | | | | 34 | | | | |

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

Лист регистрации изменений

[illegible]

| | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| | | | | |
|-------------|----------------|--------------|-------------|----------------|
| Ине. №подл. | Подпись и дата | Взам. инв. № | Ине. №дубл. | Подпись и дата |
| | | | | |

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

.РЭ