ETL. Семинар №5

1. Викторина
2. Блок 1

Задание 1

Используя материалы лекции установите Apache Airflow в чистый virtualenv с использованием constraint файла. Убедитесь что по адресу localhost:8080 открывается список Dag сценариев

Решение:
Итак, давайте установим Airflow. Вначале создадим новый virtualenv, чтобы не устанавливать в корень системы все библиотеки. После этого выполним следующие команды

```
AIRFLOW_VERSION=2.5.0
PYTHON_VERSION="$(python --version | cut -d " " -f 2 | cut -d "." -f 1-2)"
# For example: 3.7
CONSTRAINT_URL="https://raw.githubusercontent.com/apache/airflow/constraints-${AIRFLOW_VERSION}/constraints-no-providers-${PYTHON_VERSION}.txt"
# For example: https://raw.githubusercontent.com/apache/airflow/constraints-2.5.0/constraints-no-providers-3.7.txt
pip install "apache-airflow==${AIRFLOW_VERSION}" --constraint "${CONSTRAINT_URL}"
```

Мы указываем версию Airflow которую хотим установить, версию python которую будем использовать, ссылку на файл ограничений и выполняем установку с помощью PyPI используя параметры, определенные нами заранее.

Если после установки команда airflow не распознается системой (чаще всего это происходит на Windows при использовании WSL) необходимо убедиться, что в

переменные окружения добавлен путь к установочной директории Airflow и если нет, то добавить.

```
PATH=$PATH:~/.local/bin
```

Или можно вызывать Airflow с помощью python.

```
python -m airflow
```

Также очень часто встречается ошибка при вызове Airflow Symbol not found: _Py_GetArgcArgv. Эта ошибка может означать что вы используете несовместимую версию python. Суть проблемы в том, что библиотека Airflow зависит от setproctitle, использует непубличный API Python, который недоступен в стандартной установке /usr/local/opt/(которая указывает на путь под /usr/local/Cellar).

Простое решение — просто убедиться, что вы используете версию Python, в которой есть dylib библиотеки Python. Например:

```
# Note: these instructions are for python3.7 but can be loosely modified for other versions
brew install python@3.7
virtualenv -p /usr/local/opt/python@3.7/Frameworks/Python.framework/Versions/3.7/bin/python3 .toy-venv
source .toy-venv/bin/activate
pip install apache-airflow
python
>>> import setproctitle
# Success!
```

Кроме того, вы можете загрузить и установить Python прямо с официального сайта .

Задание 2

Создайте нового пользователя с правами админа и авторизуйтесь с его помощью в интерфейсе airflow

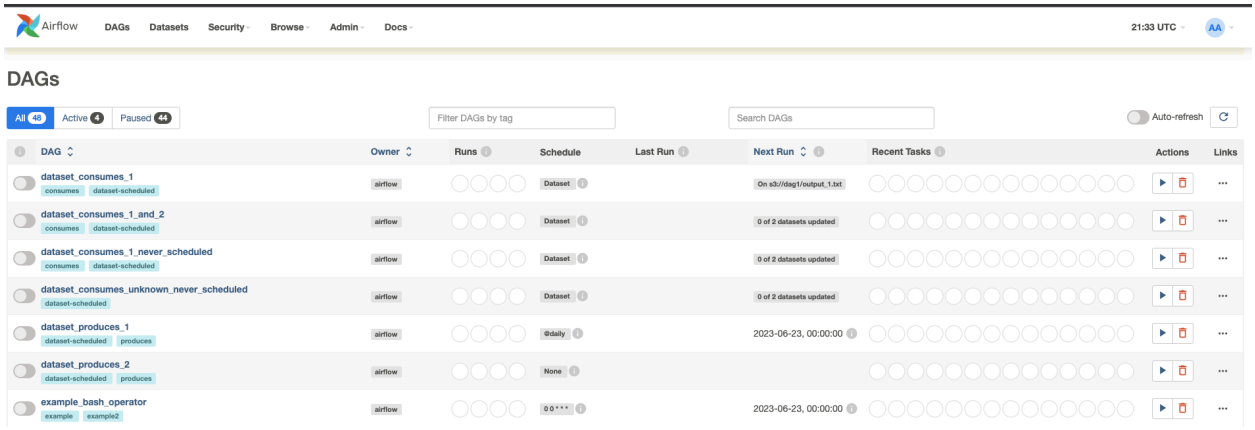
Решение:
Остановить выполнение уже запущенной копии airflow,

```
# create an admin user
airflow users create \
    --username admin \
    --firstname Peter \
    --lastname Parker \
    --role Admin \
    --email spiderman@superhero.org
```

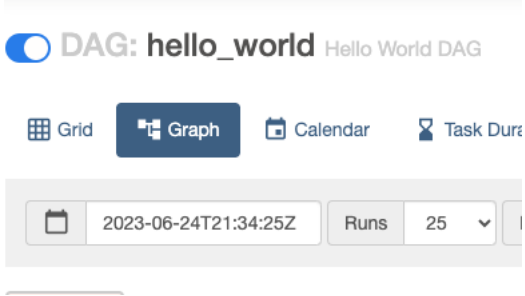
Задание 3

Посмотрите список существующих пайплайнов. Запустите некоторые из них, посмотрите на результат выполнения. Откройте логи выполнения пайплайнов. Отправьте в чат скриншот из логов, говорящий об успешном завершении пайплайна

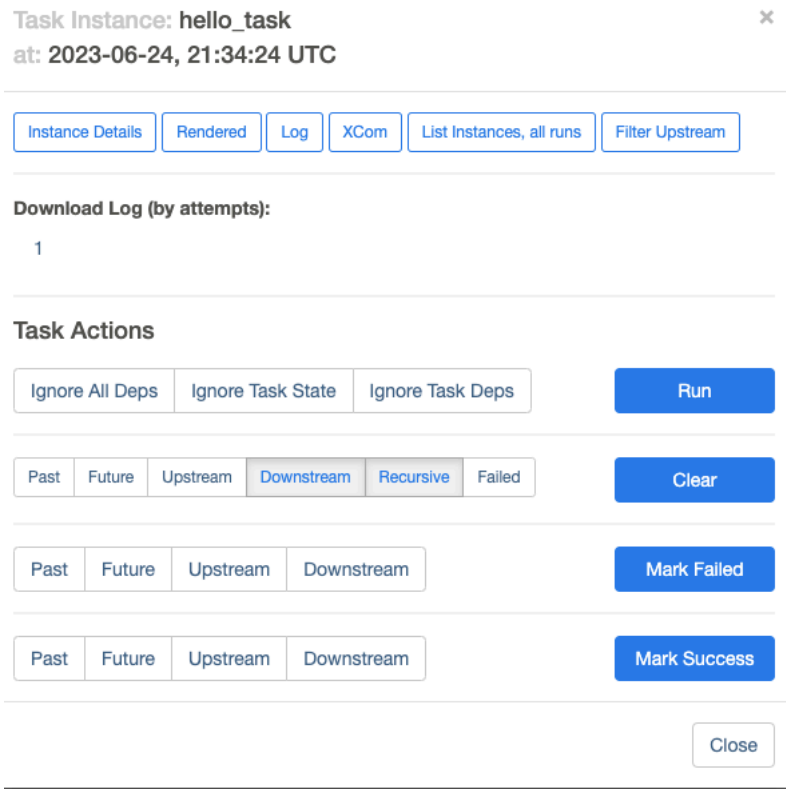
Решение:



Запустить dag -> зайти внутрь dag в секцию graph



Нажать на оператор логи которого мы хотим посмотреть и в открывшемся окне нажать кнопку Log


Task Instance: hello_task
at: 2023-06-24, 21:34:24 UTC
Instance Details
Rendered
Log
XCom
List Instances, all runs
Filter Upstream
Download Log (by attempts):
1
Task Actions
Ignore All Deps
Ignore Task State
Ignore Task Deps
Run
Past
Future
Upstream
Downstream
Recursive
Failed
Clear
Past
Future
Upstream
Downstream
Mark Failed
Past
Future
Upstream
Downstream
Mark Success
Close