# **ЛР 1. Airflow + docker compose**

### ▼ Задача

Развернуть Apache Airflow с помощью docker-compose и разработать свой DAG

## ▼ Описание

# Dockerfile:

- указать образ FROM apache/airflow:2.7.1
- указание workdir дефолтного для Airflow
- осуществить копирование директории/файла с DAGom в workdir

Для непосредственно деплоя можно использовать композ файл с этой страницы: <a href="https://airflow.apache.org/docs/apache-airflow/stable/howto/docker-compose/index.html">https://airflow.apache.org/docs/apache-airflow/stable/howto/docker-compose/index.html</a>

### Подготовка

#### docker-compose.yml:

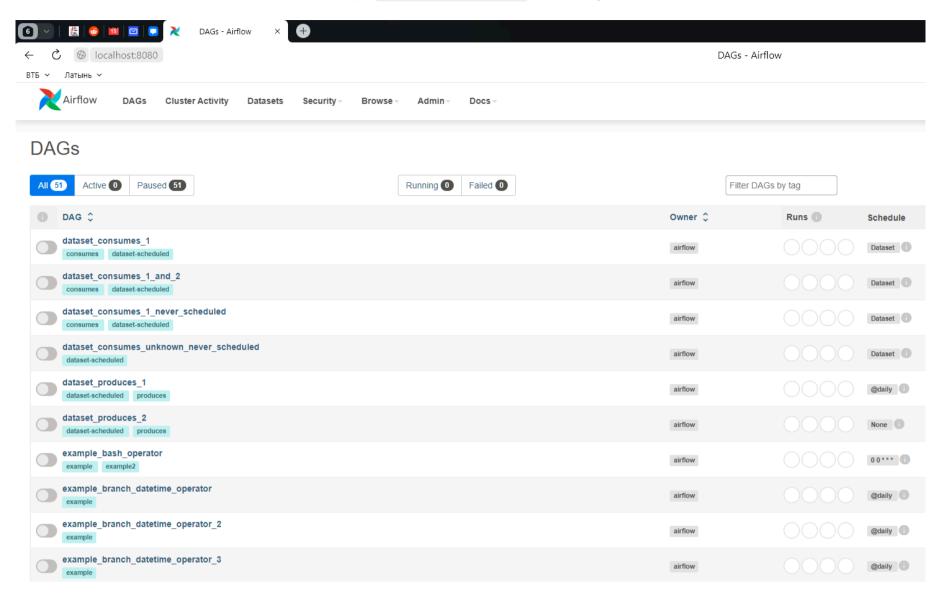
- вырезать сервисы redis, airflow-worker, airflow-triggerer, airflow-cli, flower главное, чтоб были webserver, scheduler, init и БД к ним (postgres)
- поменять дефолтный образ на свой кастомный (заменить *image* ЛИБО прописать переменную <u>AIRFLOW\_IMAGE\_NAME</u> ЛИБО использовать вместо <u>image</u> директиву <u>build</u>)
- поменять значение переменной AIRFLOW\_core\_executor с CeleryExecutor на LocalExecutor т.к. мы отключаем сервис airflow-worker
- убрать из блока &airflow-common-depends-on две строчки, касающийся redis т.к. мы его отключаем, нам не нужна такая зависимость контейнеров друг от друга
- <u>при желании</u> можно заполнить значения переменных \_airflow\_www\_user\_username и \_airflow\_www\_user\_password , тогда для подключения через браузер будут использоваться эти креды (вместо дефолтных airflow:airflow)

Если все ок, то после выполнения docker-compose up -d и docker ps должен показаться список из двух контейнеров airflow и одного postgres. Если у контейнеров состояние (health: starting), то необходимо дождаться (healthy)

```
$ nano docker-compose.yaml
time="2023-09-14T23:37:38+03:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
time="2023-09-14T23:37:38+03:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
          Network llidd_default
          Volume "llidd_postgres-db-volume"
          Container llidd-postgres-1
         Container llidd-airflow-init-1
Container llidd-airflow-webserver-1
         Container llidd-airflow-scheduler-1 Started
$ docker ps
CONTAINER ID
                                                                                                                                                                                                                       CREATED
                                                                                                                                 COMMAND
                                                                                                                                                                                                                                                                                                STATUS
                                                                                                                                 "/usr/bin/dumb-init ..."
"/usr/bin/dumb-init ..."
"docker-entrypoint.s..."
                                                  apache/airflow:2.7.1
                                                                                                                                                                                                                                                                                                                                                                                                                         0.0.0.0:8080->8080/tcp
404c09e565de
                                                                                                                                                                                                                      53 seconds ago
                                                                                                                                                                                                                                                                                                Up 29 seconds (health: starting)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             llidd-airflow-webserver-1
                                                 apache/airflow:2.7.1
                                                                                                                                                                                                                                                                                                Up 29 seconds (health: starting)
85647195cde1
                                                                                                                                                                                                                      53 seconds ago
                                                                                                                                                                                                                                                                                                                                                                                                                         8080/tcp
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               llidd-airflow-scheduler-1
 764e40204bb6
                                                                                                                                                                                                                      About a minute ago
                                                                                                                                                                                                                                                                                              Up 52 seconds (healthy)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              llidd-postgres-1
                                                  postgres:13
                                                                                                                                                                                                                                                                                                                                                                                                                         5432/tcp
$ docker ps
CONTAINER ID
                                                                                                                                                                                                                                                                                                                                                                                                                                                           NAMES
                                                                                                                                  COMMAND
                                                                                                                                                                                                                        CREATED
                                                                                                                                                                                                                                                                              STATUS
                                                                                                                                                                                                                      2 minutes ago
2 minutes ago
2 minutes ago
                                                                                                                                                                                                                                                                             Up 2 minutes (healthy)
Up 2 minutes (healthy)
                                                                                                                                                                                                                                                                                                                                                                    0.0.0.0:8080->8080/tcp
                                                   apache/airflow:2.7.1
                                                                                                                                    "/usr/bin/dumb-init ..."
404c09e565de
                                                                                                                                                                                                                                                                                                                                                                                                                                                         llidd-airflow-webserver-1
                                                                                                                                  "/usr/bin/dumb-init ..."
                                                                                                                                                                                                                                                                                                                                                                    8080/tcp
                                                                                                                                                                                                                                                                                                                                                                                                                                                           llidd-airflow-scheduler-1
85647195cde1
                                                  apache/airflow:2.7.1
                                                                                                                                   "docker-entrypoint.s..."
764e40204bb6
                                                  postgres:13
                                                                                                                                                                                                                                                                              Up 2 minutes (healthy)
                                                                                                                                                                                                                                                                                                                                                                    5432/tcp
                                                                                                                                                                                                                                                                                                                                                                                                                                                           llidd-postgres-1
time="2023-09-14T23:50:25+03:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string." time="2023-09-14T23:50:25+03:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
                                                                                                 _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/_/ | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/// | _/
                                                                                                                                                                                                                                                                               117} INFO - Loaded executor: LocalExecutor
                                                                                                   WARNING!!!: AIRFLOW_UID not set!
                                                                                                   If you are on Linux, you SHOULD follow the instructions below to set
                                                                                                   AIRFLOW_UID environment variable, otherwise files will be owned by root.
                                                                                                   For other operating systems you can get rid of the warning with manually created .env file:
                                                                                                                 See:\ https://airflow.apache.org/docs/apache-airflow/stable/how to/docker-compose/index.html \#setting-the-right-airflow-user for the following the following properties of the following properties
```

ЛР 1. Airflow + docker compose

Попасть на Airflow через браузер можно по адресу <a href="http://localhost:8080/">http://localhost:8080/</a>. Там уже будет множество примеров DAG-ов, если они не нужны, то на этапе редактирования композ файла надо поставить переменную <a href="http://localhost:8080/">AIRFLOW\_CORE\_LOAD\_EXAMPLES</a> в значение false



Основное требование касательно DAGa: состоять из нескольких шагов/тасков, быть сложнее чем 'Hello world' (в идеале, что-то должно считаться)

# ▼ Отчетность

Ссылка на Gitlab-репозиторий от команды, где содержится

- Dockerfile
- docker-compose.yml
- Директория или файл, содержащая DAG
- README.md с описанием содержимого и работы DAGa, а также кратким описанием, как задеплоить сервис у себя локально
- при желании можно добавить CHANGES.md, где отражать крупные изменения в репозитория (на момент 1-ой лабораторной их, вероятно, не будет, но дальше...)

ЛР 1. Airflow + docker compose