



## Самостоятельное задание.

### ETL пайплайн в Airflow.

#### Цель

Вы научитесь выстраивать пайплайны в Airflow и пользоваться механизмом catchup. Это позволит обращаться к историческим данным и научит базовым инструментам Airflow.

#### Описание задания

В видео вы видели пример построения пайплайна с рандомными user, вам необходимо построить подобный пайплан с другими данными. Задание выполняется самостоятельно, **не требует проверки**, потренируйтесь создавать пайплан самостоятельно, прежде чем вы приступите к заданию по проекту.

Необходимо развернуть инстанс Apache Airflow: [Документация по установке Airflow](#) либо воспользоваться пошаговой инструкцией:

1. Создать директорию для инстанса Airflow
2. Поместить в неё файл docker-compose  
(скачивается по ссылке из статьи [установка Airflow из docker-compose](#))
3. Оставляем расширение .yaml
4. Открыть терминал или CMD и перейти в эту директорию
5. Создать файл .env
6. Внести в него переменную  
AIRFLOW\_UID=50000
7. В отдельном окне терминала выполнить команду  
docker-compose up -d
8. После развертывания ожидается результат healthy у всех контейнеров (нужно подождать несколько минут) при выполнении команды:  
docker-compose ps
9. Адрес локального Airflow на localhost: 8080
10. Чтобы приостановить инстанс, используйте команду  
docker-compose down

Необходимо создать в нём новый DAG: получающий данные из API, преобразовывающий (то есть произвести нормализацию JSON) и сохраняющий данные в таблицу. Данные необходимо сохранить за 1 предыдущий месяц, то есть DAG должен срабатывать каждый день предыдущего месяца. Для данной задачи данные можно сохранять в Metastore.

#### Инструменты, которые пригодятся для выполнения:

- Docker docker-compose
- Apache Airflow 2
- Для выполнения задания используйте API одного из ресурсов:



<https://www.boredapi.com/> апи с видами активности на свободное время

**По итогу выполнения задания у вас должен быть вывод из** таблицы PostgreSQL с данными за 30 дней, которые вы скачали из API

Как вам урок?



Изучил, далее >>

Слёрм ©

[+7 \(495\) 248-05-80](tel:+7(495)248-05-80)

[Лицензия №ДЛ-1368 от 22.08.2019](#)

[Политика конфиденциальности](#)

[Публичная оферта](#)