Start coding your first DAG

A DAG (Directed Acyclic Graph) is the core concept of Airflow, collecting Tasks together, organized with dependencies and relationships to say how they should run. A DAG is a data pipeline.

Starting from this first Activity, we are going to build a data pipeline together that downloads and displays Github stars of the Apache Airflow project.

Simple but powerful enough to show you the most important features of Airflow.

So, fasten your seatbelt and let's go 🚀

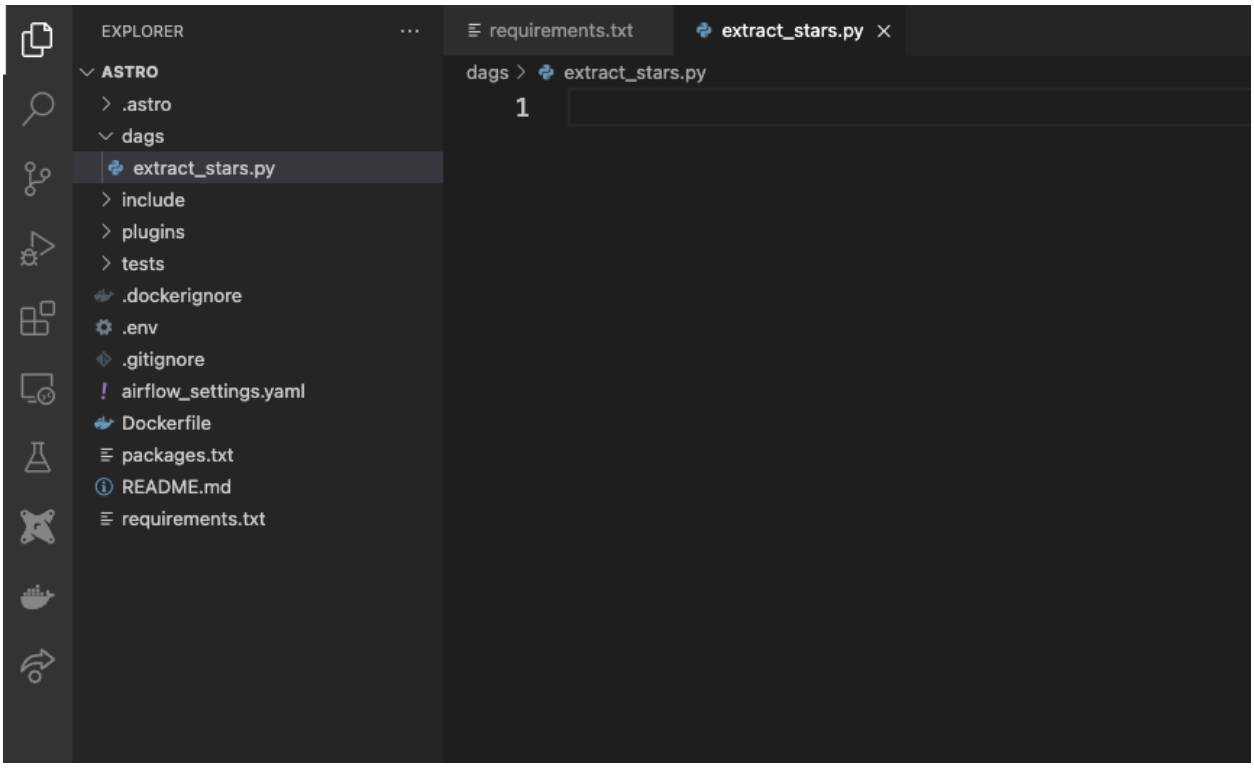
Prerequisites

To create a DAG you need have set up your local Astro project as described under the section "Set up your local environment"

The Skeleton

There are basic parameters that you will always define to create a DAG.

First, create a new file `extract_starts.py` in the folder `dags/`



Add the following imports

```
from airflow import DAG

from datetime import datetime
```

Next, define a DAG object with

- dag_id: `extract_stars`
- schedule_interval: `@daily`
- start_date: `2022-01-01`
- catchup: `False`

Try to implement it as shown in the `DAG minimum requirements` video https://academy.astronomer.io/onboarding-course/1287017

Once you are done, look at the solution below.

Solution

```
from airflow import DAG

from datetime import datetime

with DAG('extract_stars', schedule_interval='@daily', start_date=datetime(2022,1,1), catch
up=False) as dag:
  None
```

⚠️ Be careful if you copy and past the code below. Make sure you only have tabs and not a mix of spaces and tabs otherwise you will get an error.

Add your first task

After having your DAG defined, it's time to add the first task.

This task prints the current date on the standard output by executing the `date` bash command.

First, go to https://registry.astronomer.io/

Look for the `BashOperator`

Try to implement the task with following requirements:

- task_id: `get_date`
- bash_command: `date`

Once you are done, look at the solution below.

💡 Don't forget to save your file

Solution

```
from airflow import DAG
from airflow.operators.bash import BashOperator
```

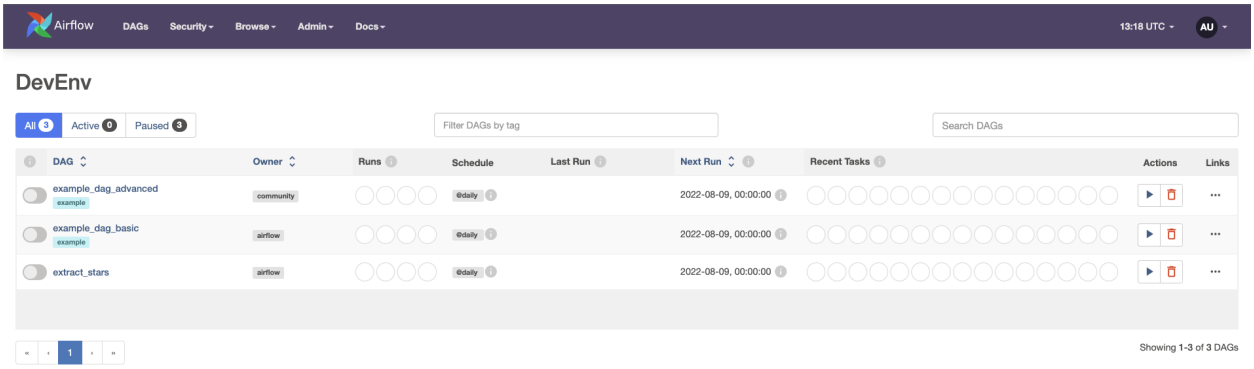
```
from datetime import datetime

with DAG('extract_stars', schedule_interval='@daily', start_date=datetime(2022,1,1), catch
up=False) as dag:

  get_date = BashOperator(
    task_id="get_date",
    bash_command="date"
  )
```

⚠️ Be careful if you copy and past the code below. Make sure you only have tabs and not a mix of spaces and tabs otherwise you will get an error.

Airflow UI

Go on the Airflow UI `localhost:8080`

If you didn't make any typo, you should see your DAG as shown below:



⚠️ You need to wait up to 5 mins before getting your DAG on the UI

Well done! You have successfully created your first DAG! But… we are not done wait 😉