Interact with an External Service

Airflow's power comes from interacting with and coordinating work in other services.

Prerequisites

To interact with an external service, you need to have set up your local Astro project as described under the section "Install a Provider" and have a connection as described under the section "Install the HTTP Provider"

You should also have the DAG `extract_stars.py` in your folder `dags/`. If not, take a look at the previous activity `Start coding your first DAG…`

Where are you at

In the file `extract_stars.py` in your folder `dags/`, your DAG should look like this:

```
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), cat
chup=False) as dag:

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

Find the Operator

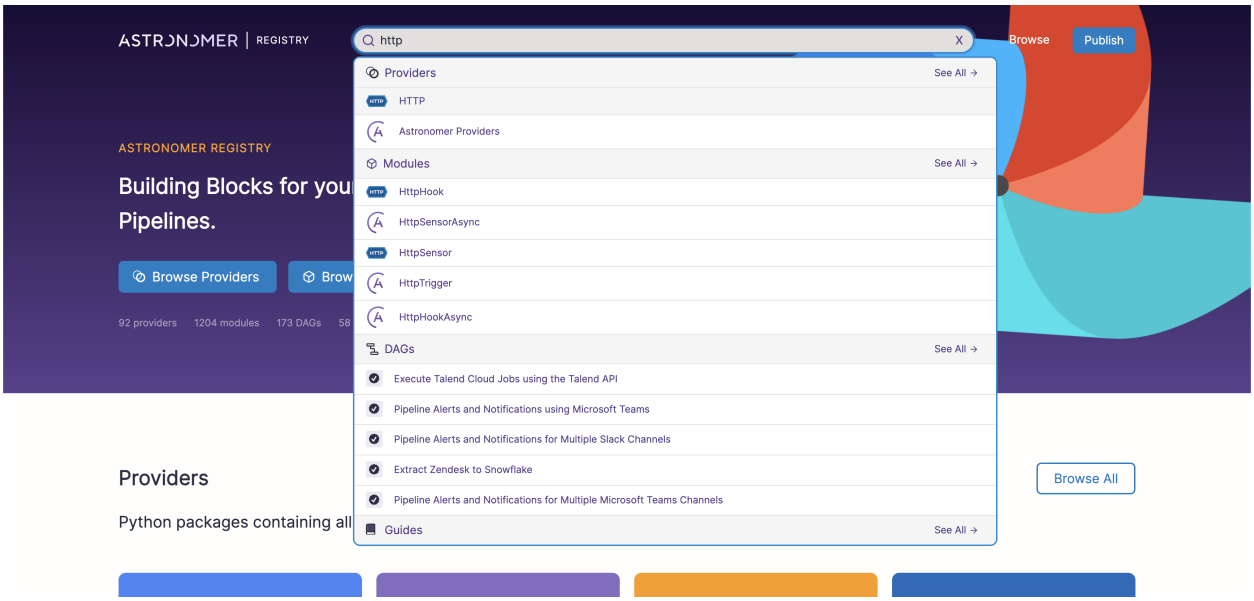
To illustrate interacting with an API, we will pull the ⭐ Stars from Github's API for the Apache Airflow Open Source project.

For that, the first step is to create an HTTP connection. That's what you've done in the previous activity.

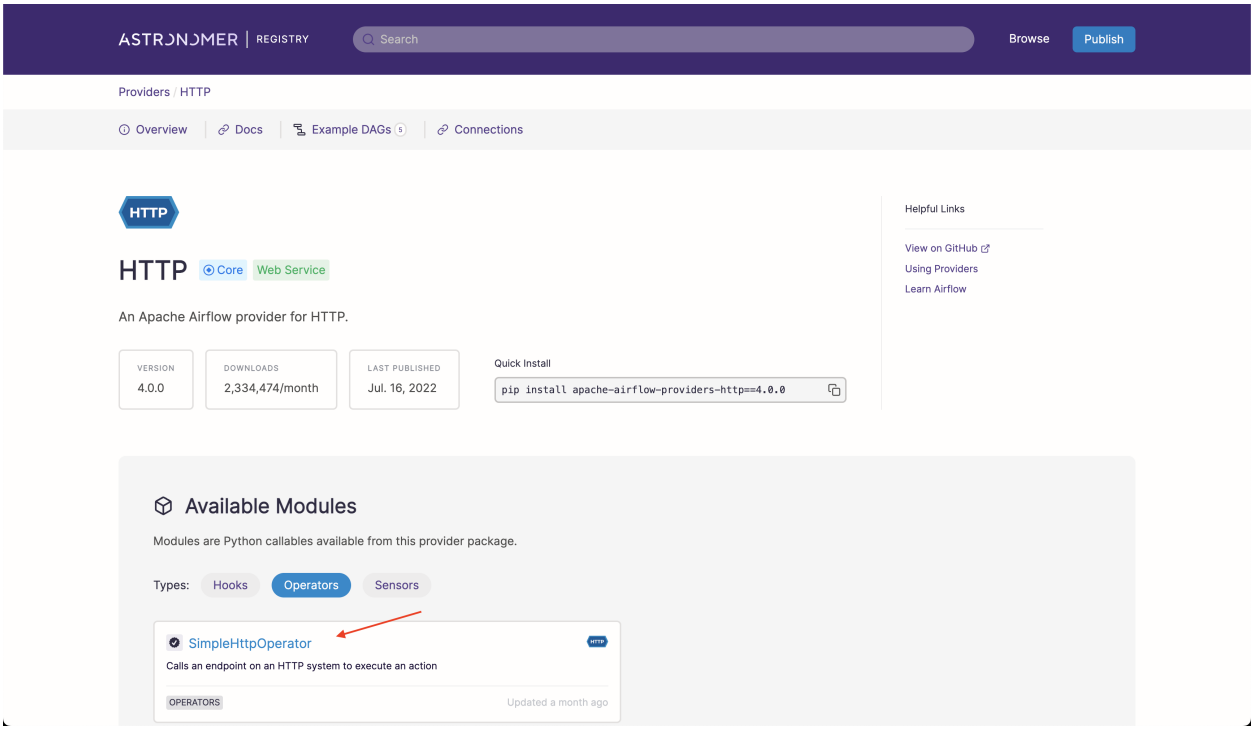
Now, it's time to use an Operator that downloads the content of an HTML page.

Back to the registry.astronomer.io

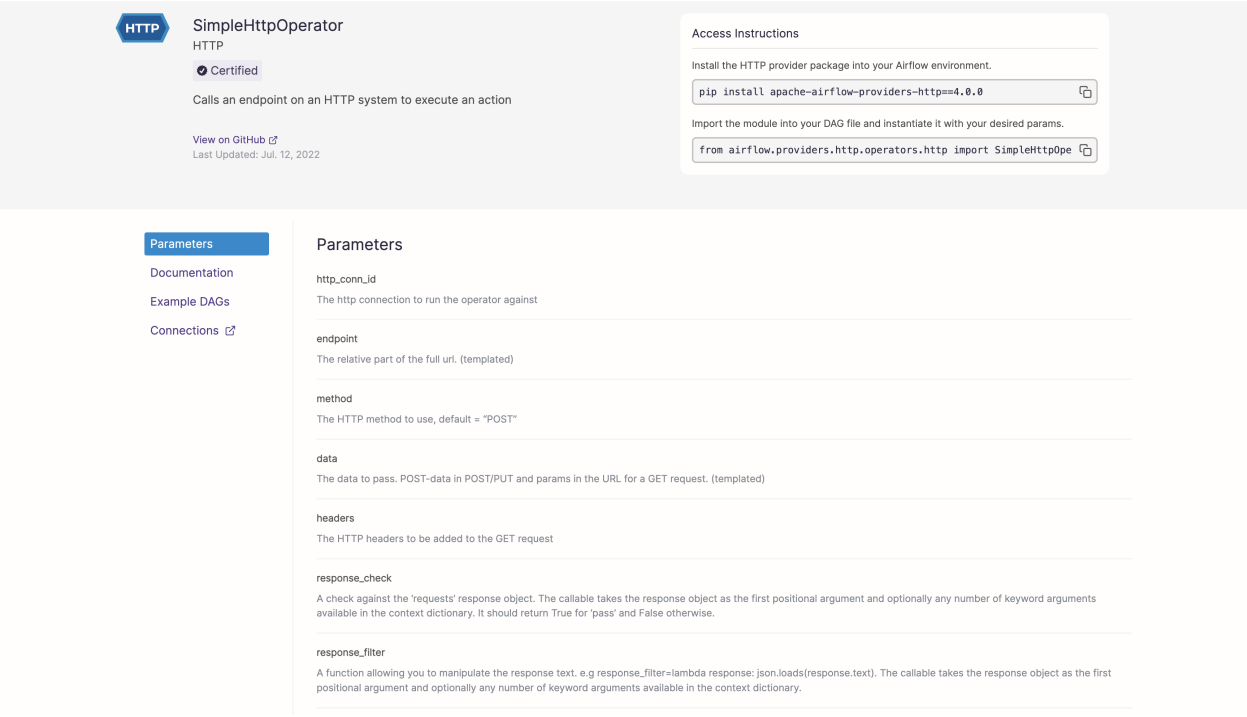
Look for `HTTP` in the search bar



Select the `HTTP` provider and click on `Operators` under `Available Modules`, then select the `SimpleHttpOperator`



The `SimpleHttpOperator` calls an endpoint on an HTTP system to execute an action. Typically, the endpoint is an HTML page. There are a couple of parameters to define:



Feel free to take a look at each of these. In this example, we focus only on:

- http_conn_id: we will use the connection id of the connection you've created before
- endpoint: the relative part (page) of the full url
- method: the HTTP method to use, GET
- log_response: allows checking the response in the Task logs

The SimpleHttpOperator

In your `extract_stars.py` DAG file, just under the `get_task` task.

Import the SimpleHttpOperator

Add a new task named `query_github_stats` with the SimpleHttpOperator

Define the following arguments:

- task_id: `query_github_stats`

- endpoint: `repos/apache/airflow` (the Apache Airflow repository)
- method: `GET`
- http_conn_id: `github_api`
- log_response: `True`

Save the file and go to the Airflow UI (localhost:8080) to see if you don't get any errors.

Try to do it, and I'll see you for the correction below

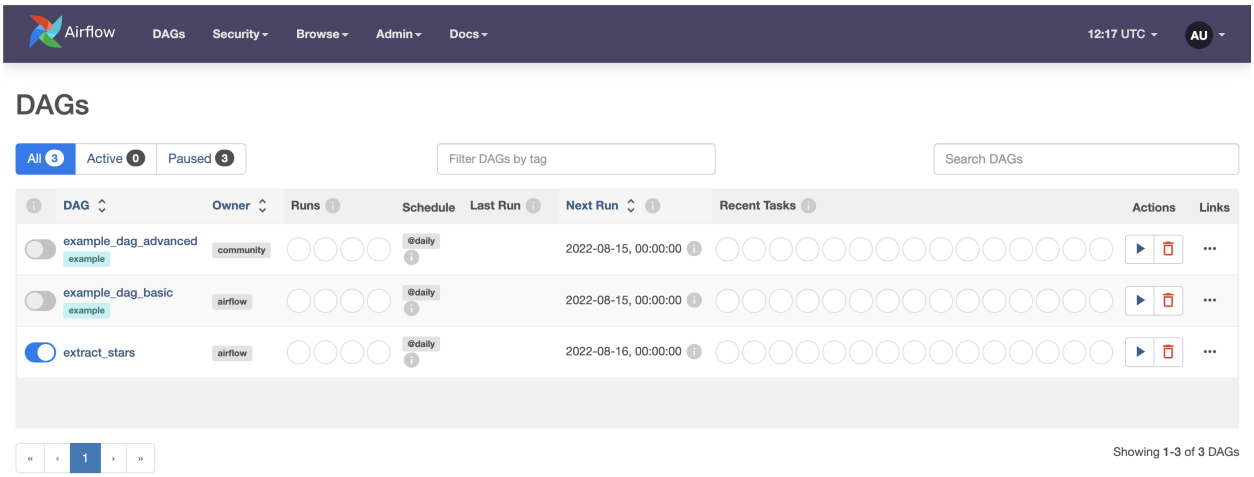
Solution

```
from airflow.providers.http.operators.http import SimpleHttpOperator

query_github_stats = SimpleHttpOperator(
      task_id="http",
      endpoint="repos/apache/airflow",
      method="GET",
      http_conn_id="github_api",
      log_response=True
  )
```

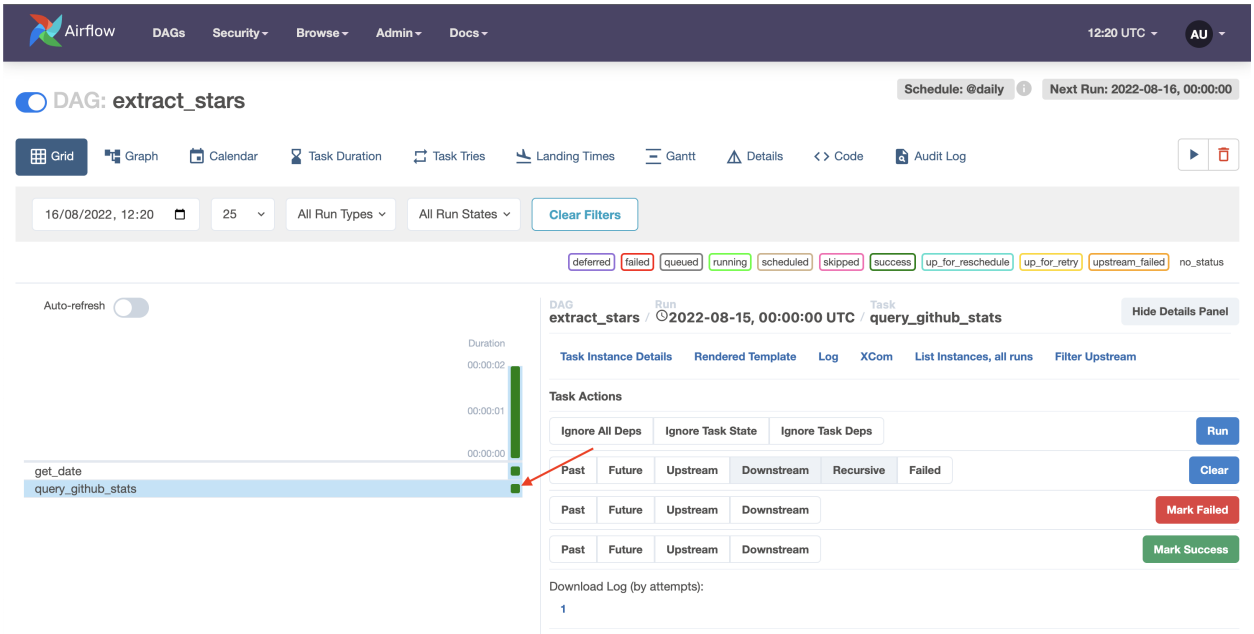
Get the repo info

Go to the Airflow UI (localhost:8080) to see if you don't get any errors.



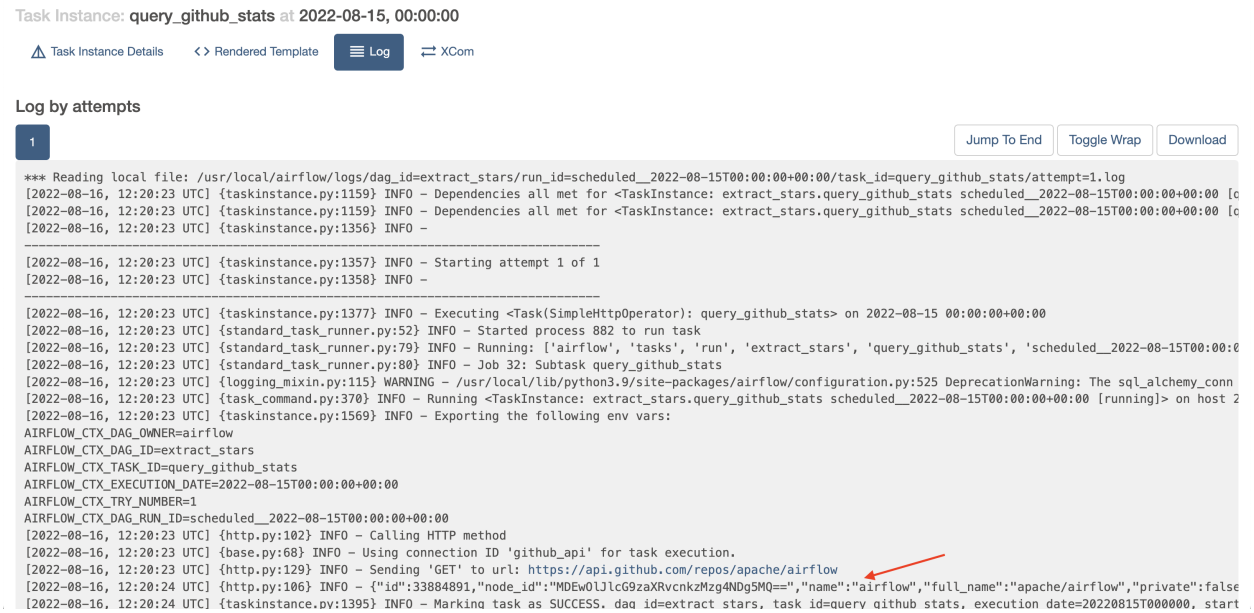
Trigger your DAG by clicking the `lecture` button ▶️ and `Trigger DAG`

Click on your DAG and the square corresponding to the `query_github_stats` task



Go to the `Log`

You should see the following output



Well done! You are now able to download HTTP data from your data pipelines

Additional resources

SimpleHttpOperator:
https://registry.astronomer.io/providers/http/modules/simplehttpoperator