Work with APIs in Apache Airflow

Dayo Bamikole

Data Architect

@dayo_bamikole | www.dayobam.com

Overview

How to use HttpOperator

Understanding API calls and process responses

Integrating an API call into a DAG

How to Use HttpOperators

What Are HttpOperators?

HttpOperator is an Apache Operator

Used to make HTTP requests within a DAG

Goal

It allows workflows to interact with web APIs

Making it possible to retrieve or send data during workflow execution

HttpOperator Features

HTTP methods

GET, POST, PUT, DELETE

Customizable

Headers, parameters, and body can be specified

Response handling

Can log responses and process them further in the DAG

Install Http Provider

```
pip install 'apache-airflow-providers-http'
```

Including HttpOperator in DAG

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

Parameters to Include Using SimpleHttpOperator

```
task_id:
method:
http_conn_id:
endpoint:
headers:
```
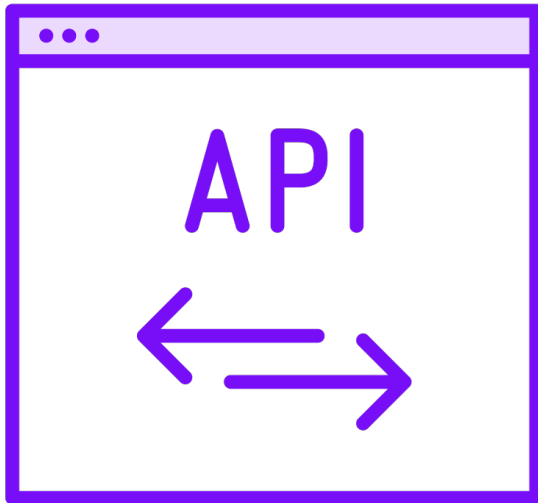
Understanding API Calls and Process Responses

API Calls and Process Responses

PythonOperator

- Executes Python functions or callable objects as tasks

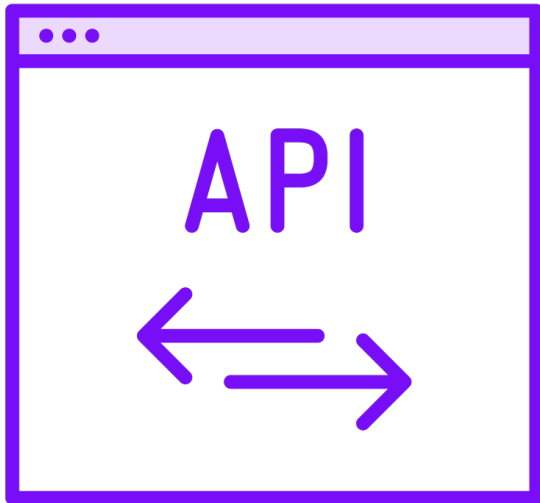- Ideal for custom data processing, transformations, or integrations

Using the PythonOperator

```python
from airflow.operators.python import PythonOperator
```

API Calls and Process Responses

XComs

- Short for cross-communication
- Allow tasks to exchange data
- Used to pass messages or data between tasks within a DAG

Push Data Using XCom

`do_xcom_push=True`

Integrating an API Call into a DAG

Integrate an API Call into a DAG

□ → ○ → ◇

HttpSensor

- Ensures the API is available before making requests
- Helps in identifying issues with API availability early

Using the PythonOperator

```python
from airflow.providers.http.sensors.http import
HttpSensor
```

Best Practices Using APIs in Airflow

Always check API availability before making requests

Handle errors and retries gracefully

Use logging to monitor and debug API calls and responses

Keep your DAGs modular and maintainable

Summary

Understanding and using SimpleHttpOperator for API integration

Utilizing XCom, PythonOperator for data processing

HttpSensor to identify issues

Best practices for using APIs in Airflow