How to Interact With External Sources Using Hooks?

Interfaces to interact with your external sources

Use Case

- Let's suppose we have a PostgreSQL database and we would like to fetch data from a table named "source" using Apache Airflow. How could we do that?
- If you use directly PostgreOperator and put the following request as sql parameter: "SELECT * FROM source", you will end up with nothing since PostgreOperator does not return any result from the sql request.
- You could try to use BashOperator or PythonOperator to do the job but it's not actually the best solution as you gonna need to use libraries and so on.
- Apache Airflow helps you by providing Hooks.

Definition

A Hook is simply an interface (a set of functions) to interact with external systems such as HIVE, PostgreSQL, Spark, SFTP and so on. Apache Airflow provides you many different hooks in order to make your life easier.

For instance, PostgreOperator uses PostgresHook to interact with your PostgreSQL database and execute your request. It handles the connection and allows you to execute SQL like if you were logged into the PostgreSQL command interface.

Practice Time

Time to practice!