Airflow Lab

DS 5559

Kaia Lindberg (pkx2ec)

Link to Git repo: https://github.com/kaiarl/DS5559\_airflow\_lab

Steps:

* Create a directory from where you'll be working from and execute next steps from inside that directory.
* Copy or save this file to the directory.  Link--> [docker-compose.yamlLinks to an external site.](https://gist.github.com/EfrainOlivaresUVA/43edf07973a222c52e35921dc1267fb4)
* Create a sub-directory called 'dags', (I believe Airflow may create that for you but just be aware you'll need this folder to add tasks).
* Run 'docker-compose up' from the command line.

Text

Description automatically generated

A picture containing text, newspaper, screenshot, document

Description automatically generated

* After that, you will see the download of 3 docker images, and console should start flowing for the web server.
* Open a browser to localhost:8080
* Log into the running Airflow server with user=airflow, password=airflow

Graphical user interface, application

Description automatically generated

* You should now have a running Airflow system with example tasks.
* It should allow you to explore and follow along sections of the Data Pipelines with Apache Airflow book.

Graphical user interface, text, application

Description automatically generated

What you will turn in (please read it to the end before you start so you can get screenshots along the way):

* Create a new repo (yes, another repo, sorry, but this just makes it so much easier to review and grade :) ).  Name it DS5559\_airflow\_lab
* This repo should contain your file that will be placed into the 'dags' folder.
* Copy this file -> [LINKLinks to an external site.](https://gist.github.com/EfrainOlivaresUVA/6643ef661d09b0b2067b394da80cf06e" \t "_blank)  into a file named user\_processing.py into your 'dags' folder
* Now when you bring up Airflow again you should see the user\_processing dag at the bottom of your DAG view.
* You DO NOT, I say again,  **DO NOT,** have to go through all the steps of rebuilding this file as described here: [LINKLinks to an external site.](https://github.com/UVADS/data_engineering/blob/main/09_lte1_airflow/Airflow%20Lab.ipynb" \t "_blank) .   This file is already fully built and working.  We are now using the lab more to explore rather than build the DAG, so I'll update the text in the next generation.  For now it should be a good walkthrough of how it was put together and why.
* You DO however, **DO** need to do two steps of manual setup so that it will actually work end to end.  the steps are mentioned in the next section.

Text

Description automatically generated

Please turn in the following.

1.  **Screenshot** of the DAG view showing this dag having loaded

Graphical user interface, application

Description automatically generated

2.  **Screenshot** after you trigger it and failed showing in the DAG page.

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email, Teams

Description automatically generated

Graphical user interface, application

Description automatically generated

3.  Now manually set up the connections mentioned in the lab instructions, steps **1.3, and 1.7.**

Graphical user interface, application, website

Description automatically generated

4.  **Screenshot** of the dag triggered again, and it should now show a 'passing' circle too

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated