Recap

Time to sum up!

Summary

- Docker allows to run a Docker container which is an instance of a Docker image.
- A Docker image is defined from a Dockerfile specifying how your application should be installed and run.
- Docker-compose is used to run a multi-containers application very easily and create a network shared by all the containers within this application.
- You can modify the configuration file of Airflow by using environment variables. For example, if we want to change the value of the parameter dags_folder under the core section, we just need to create the environment variable AIRFLOW__CORE__DAGS_FOLDER and set the value
- A FERNET_KEY is related to the crypto package from python in order to crypt the connections saved into the metadatabase increasing the security.

Summary

- With Sequential Executors and Docker you can share a common SQLite file database if the path of the SQL_ALCHEMY_CONN points to the same binded volume for both containers (the webserver and the scheduler).
- It is also possible to run the scheduler and the webserver from the same container which is actually considered as a best practice. Your metadabase should be run from a separate container as we did in the example with Local Executors.
- Celery Executors and Docker allow you to quickly add and remove workers to scale out Airflow as your needs increase or decrease.
- The Celery Flower user interface is used to monitor your workers and the tasks they are processing.

What's Next?

Well, you have just finished the course! Congratulation! I hope you have learned a lot of new and exciting things about Apache Airflow. It is really a great tool customizable as much as you want, offering you almost limitless possibilities to create your own DAGs.

It was a pleasure to be your teacher during this course and I would be very happy to see you soon. Feel free to visit my website if you want to see more advanced concepts about Apache Airflow.

Do not hesitate to ask me any questions or suggestions.

Have a great day!

Marc Lamberti