Seminar

Apache Airflow Application Development

Homework 3 – Airflow setup

Homework 4 – Your first Airflow Project

Deadline: April 26, 2024

Homework 3: Airflow setup

Prerequisites

- Walk through the presentations of Lectures on Airflow and brush up on knowledge
- Check the official documentation at airflow.apache.org

Tasks

- **Install Apache Airflow** in any way described in the presentation 2 or in your own way (e.g., see recommended readings)
- **Create a DAG with more than 3 Tasks** using different operators (e.g. BashOperator, PythonOperator, EmailOperator)
- **Schedule and test it** using your *Airflow instance*.

Homework 3: Airflow setup

- Create a report on Homework as a collection of screenshots to illustrate steps of the application development and testing process (see slide 7 on report content)
- Add a short comments to slides explaining your actions

- **Deadline for HW3 submission: Friday, April 26th, 2024**

Homework 3: Airflow setup - reading

Luís Oliveira, **How to Setup a Lightweight Local Version for Airflow**. With Docker and Docker Compose. 17 Sep 2022.

Source: https://datatalks.club/blog/how-to-setup-lightweight-local-version-for-airflow.html

Homework 4: Airflow mini-project

Prerequisites

- Check the official documentation at airflow.apache.org/
- Find a dataset (e.g use Kaggle datasets)

Tasks

- **Create a DAG or several DAGs with more than 3 Tasks** using different operators (e.g. BashOperator, PythonOperator, EmailOperator) and sensors.
- **Make a connection to DB** and apply Hook to insert the data*
- **Schedule and test it** using your *Airflow instance*.

- *extra points

Homework 4: Airflow mini-project

- Create a report on Homework as a collection of screenshots to illustrate steps of the application development and testing process (see next slide for recommendation on report content)
- Add a short comments to slides explaining your actions
- Make a presentation with your experience and results

- **Deadline for HW4 submission: Friday, April 26th, 2024**

Homework 4: Airflow mini-project - Reporting

1. **Provide a general description of the project.**
   - What do your DAG/DAGs intend to do?
   - For what purpose were they created?
2. **Describe the data used in project.***
   - What data source did you use?
   - What is the selected dataset?
3. **Describe DAG/DAGs and attach a picture of the graph.**
   - What Tasks do they consist of?
   - What category do Tasks belong to?
   - What is the sequence of their execution?
4. **Describe the features of your project (any Decorators, Sensors, Hooks?)**
5. **Demonstrate and explain how your DAG/DAGs work**

- *extra points

Homework 4: Airflow mini-project - Readings

- https://medium.com/@SaiParvathaneni/apache-airflow-for-beginners-build-your-first-dag-542affef6192
- https://medium.com/@SaiParvathaneni/building-an-end-to-end-data-pipeline-with-airflow-and-python-3bf60fb6986
- https://medium.com/@SaiParvathaneni/building-an-end-to-end-data-pipeline-part-2-with-aws-services-96b1f88a9270
- https://betterprogramming.pub/a-simple-airflow-design-pattern-to-avoid-a-cardinal-sin-cebeafd19a99