

Updated 03/14/2024

Sign up

Advanced Apache Airflow training

1 day (7 hours)

Presentation

Our advanced Apache Airflow training course will enable you to master one of the leading MLOps tools. This will enable you to manage your data workflows efficiently.

Our course will start with a focus on dynamic DAG generation and its associated actions. You'll learn the fundamentals of scheduling and advanced templating techniques with Jinja.

We'll go on to learn about custom operators, hook manipulation and API integration. Various operators such as [PythonVirtualenvOperator](#) or [KubernetesPodOperator](#) will no longer hold any secrets for you.

Finally, we'll look at case studies, ModelOps and best practices.

As with all our training courses, we'll be teaching you the latest stable version: [Apache Airflow 2.8](#).

Objectives

- Create dynamic DAGs using external configurations
- Customize operators and sensors for specific needs
- Integrate tasks with external environments such as PythonVirtualenv, Docker, and Kubernetes
- Applying best practices in ModelOps and resource management for the engineering of data
- Understanding Airflow's limitations

Target audience

- Big data engineer
- Developer
- Tech Lead
- Technical Architect
- Data engineer
- MLOps
- Data Scientist

Prerequisites

- Knowledge of the Python language
- At least 2 years' experience in data science
- Basic Airflow skills
- Preferably have taken our [Airflow training course](#)

Software requirements

- Latest version of Docker installed
- Latest version of Docker Compose installed
- Latest version of Python installed
- The latest version of Airflow installed
- Minimum 8GB RAM and SSD

Apache Airflow advanced training program

Advanced DAG models

- Dynamic DAG generation
 - Task groups
 - Triggering DAG execution
 - Dynamic task mapping
 - Delayed operator
 - Installation
- Why not use SubDAGs?
- Datasets and data aware scheduling
 - Reminder of classic scheduling and timetables
- Advanced templating with Jinja
- Generate DAGs dynamically based on external configurations
- Practical :
 - Task groups
 - Triggering DAG execution
 - Dynamic task mapping
 - Deferred operators to organize complex workflows

Custom operators

- Reminder of available Airflow operators
- Create custom operators and sensors
- Creating hooks for external systems
- Create a custom operator to interact with an API
- Create a sensor that expects a specific condition

Integration with external environments

- ExternalPythonOperator
- PythonVirtualenvOperator
- DockerOperator
- KubernetesPodOperator
- Implementing a task using the PythonVirtualenvOperator

Case studies, best practices and limitations

- ModelOps
- Dynamic resource allocation
- Data engineering workflow
- Best practices
 - Idempotence
 - Avoid high-level code
 - Decide where to carry out data processing tasks
 - Do not pass large amounts of data through Xcom
- The limits
- Resources for improving your knowledge

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.