

Updated on 23/09/2024

Sign up

Apache Airflow training

3 days (21 hours)

Presentation

Our Apache Airflow training course will enable you to supervise, monitor and schedule hundreds of pipelines. Written in Python, Airflow is an alternative to workflow tools using descriptive languages (XML , JSON , YAML ...) that are difficult to read and write.

By the end of this course, you'll be able to make your workflows dynamic, automated and scalable. You'll also learn how to create your own plug-ins, operators and executors to best fit your big data architecture.

You'll learn about the different concepts that make up Airflow, such as what is a DAG, a task, an executor, an operator and so on. We'll take a tour of the user interface and the command line interface, so that we can quickly develop our first workflow interacting with different data systems.

You'll also learn advanced concepts. We'll use practical examples to show you how to make Airflow scalable through the use of different Executors. A program that will enable you to orchestrate your data pipelines simply and robustly.

As with all our training courses, we'll be using the latest stable version of [Apache Airflow 2.10](#) software.

Objectives

- Discover Apache Airflow features
- Design, plan and monitor workflows
- Understanding the inner workings of Airflow
- Know how to scale Airflow to manage a large number of pipelines
- Learn about Airflow's best practices and limitations

Target audience

- Big data engineer
- Developer
- Tech Lead
- Technical Architect

Prerequisites

- Knowledge of Python
- [Test My Knowledge](#)

Software requirements

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

Recommendations for pre- and post-course reading

- An informative and well-structured article on [best practices](#) for Apache Airflow
- A tutorial [to automate your data pipelines](#), ideal for boosting your productivity!
- 5 minutes of reading to understand [Apache Airflow testing](#)
- A great article on [how to use Airflow](#) properly and avoid common mistakes.

Apache Airflow Training Program

Introduction

- The objectives
- Training organization
- Who am I?
- Introduction to the development environment (tool installation)

Discover Apache Airflow

- Why use it and what you need it for
- Airflow's competitors
- Airflow architecture
- Understand how Airflow works and how workflows are planned
- [Practical] Installing Airflow
- [Practice] Discovering the user interface
- [Practice] Discovering the CLI and HTTP REST API

- Quiz

Coding your first workflow with Airflow

- Defining a DAG (Directed Acyclic Graph)
- [Practical] Coding our first DAG
- The different types of operators
- [Practice] Adding the three main operators to our DAG
- Operator dependencies and bitshift composition
- [Practice] Defining task execution order in our DAG
- How exactly does the Airflow planner work?
- Workflow definition
- Quiz

Executors and their backends

- The different executors and backends
- Advantages and weaknesses
- [How to] Configure a LocalExecutor and run a DAG
- [Practice] Setting up a CeleryExecutor and running a DAG
- Quiz

Airflow's dynamic concepts

- Minimize repetitive DAG patterns with SubDags
- [How to] Simplify a DAG with SubDags
- XCOMS
- [Practice] Exchanging data using XCOMS
- How to execute different tasks according to dynamic criteria
- [Practice] Creating our first conditional tasks
- Control tasks contained in a DAG with SLAs
- [Practice] Defining SLAs for our tasks
- Variable airflow
- [Practical] using variables from our DAG
- Airflow connections
- [Practice] using connections from our DAG
- Jinja Templating, Macros and Variables
- [Practice] Adding data to run time with macros and templates
- Pools
- Quiz

Airflow providers to interact with the rest of the world

- Existing providers and their installation
- Choosing the right operators for your executors and scaling requirements
- Customize an existing operator
- [Practical] Installing a provider and customizing an operator
- Quiz

Administer Airflow: Security, RBAC, Metrics and Logging

- Securing Airflow
- User creation and authentication
- [Practical] Review of different authentication methods
- Discover RBAC (Role-based Access Control)
- [Practice] Creating different users and roles
- Log system discovery and configuration
- [Practice] Log configuration
- [Practical] Airflow monitoring and metrics to watch out for
- Quiz

ADD-ON MODULE (+1 DAY): ADVANCED APACHE AIRFLOW

Creation of an Airflow plugin to interact with ELASTICSEARCH and POSTGRESQL

- Adding functionality to Apache Airflow
- [Practice] Creating a Hook to interact with Elasticsearch
- [How to] Create a TransferOperator to transfer data from PostgreSQL to ElasticSearch
- [Practical] Adding a view to the Airflow user interface
- Quiz

Airflow Advanced administration

- [Practical] Using the Crypto library to secure Airflow
- [Practice] Using InfluxDB and Telegraph to monitor the Airflow
- Quiz

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.