

**Advanced Apache Airflow**

**Course Number:** PYTH-222  
**Duration:** 2 days

**Overview**

This Advanced Airflow training course goes beyond the fundamentals to teach attendees how to create more sophisticated DAGs (Directed Acyclic Graphs) and apply security practices to Apache Airflow. In addition, students learn how to scale Airflow within Kubernetes.

**Prerequisites**

All students should have taken [Introduction to Apache Airflow](file:////training/apache-airflow-introduction) or have the equivalent knowledge.

**Materials**

All Airflow training students receive comprehensive courseware.

**Software Needed on Each Student PC**

* Python 3.5 or later
* Airflow 2.1 or later

**Objectives**

* Secure your Apache Airflow installation
* Create highly concurrent DAGs in Kubernetes
* Leverage most of the new functionality Airflow 2.x brings

**Outline**

* Introduction
* Creating DAGs
  + Secrets, connections, and variables
  + Creating connections on startup
  + Using Pools for long-running and demanding tasks
  + Simulating long-running tasks
  + DAG serialization
  + DAG versioning
  + Testing DAGs
  + CI/CD in Airflow
* Modularizing DAGs
  + TaskGroups vs subDAGs
  + TaskFlowAPI and XComs
  + Modularizing
  + Dynamic and Functional DAGs
  + SmartSensors and timeouts
* Airflow Security
  + RBAC in Airflow
  + Setting up OAuth authentication
  + Add Google OAuth
  + Adding SSL certs
  + Default Roles and custom roles
  + Creating a custom role
* Airflow in Kubernetes
  + The Helm chart
  + Deploying Airflow with Helm
  + Deploying single tasks to Kubernetes: KubernetesPodOperator
  + Adding a task in Kubernetes
  + Scaling Airflow with Kubernetes executor
  + Changing the Helm charts values
  + KEDA autoscaler
  + Preparing DAGs for Kubernetes
  + Creating a DAG fully in Kubernetes
  + The CeleryKubernetes executor for extreme scalability
* Upgrading from Airflow 1.10
* Conclusion