Cloud
Contraptions
TRAINING & CONSULTING

To discuss this course and customizations:
Call: 434-509-5680 or Email: sales@cloudcontraptions.com

Apache Airflow Administration: Scalable Workflow Automation and Orchestration

Duration

14 hours of intensive training with live instruction delivered over two to four days to accommodate varied scheduling needs.

Target Audience

- All students should have practical experience writing Python programs.

Executive Summary

This intensive two-day course delivers a deep dive into Apache Airflow's architecture and core components—DAGs, operators, schedulers and executors—while contrasting it with Cron Jobs and Celery. Through hands-on labs in installation, Python/PostgreSQL and Kubernetes (EKS/Helm) deployment, custom container image building, and monitoring with logs and Grafana, participants will master the skills to configure, scale and optimize production-grade workflow automation solutions.

Description

This course provides a deep dive into Apache Airflow, a powerful workflow automation platform for managing complex data pipelines. Participants will explore the architecture of Airflow, including Directed Acyclic Graphs (DAGs), operators, and schedulers. The course covers installation, configuration, and integration with Kubernetes, AWS EKS, and Helm. Attendees will gain hands-on experience deploying Airflow, optimizing workflows, customizing container images, and monitoring performance using logging and metrics. Designed for professionals, this course ensures participants can build scalable, reliable, and efficient workflow automation solutions.

Objectives

- Understand Apache Airflow's architecture and how it compares to Cron Jobs and Celery.
- Learn the fundamentals of DAGs, operators, tasks, variables, and schedulers.

To discuss this course and customizations:
Call: 434-509-5680 or Email: sales@cloudcontraptions.com

- Install and configure Apache Airflow using Python environments, PostgreSQL, and Kubernetes.
- Gain hands-on experience deploying Airflow on Kubernetes, including EKS and Helm.
- Configure Airflow's executors, logs, and advanced settings for scalability.
- Build and use custom Airflow container images with additional dependencies.
- Implement monitoring solutions using logs, Grafana, and external storage.
- Apply best practices for workflow reliability, scaling, and automation in production environments.

Training Materials

Students receive comprehensive courseware, including reference documents, code samples, and lab guides.

Software Requirements

Students will need a free, personal GitHub account to access the courseware. Students will need permission to install Python and Visual Studio Code on their computers. Also, students will need permission to install Python Packages and Visual Studio Extensions. If students are unable to configure a local environment, a cloud-based environment can be provided.

Training Topics

What is Apache Airflow?
- Distributed Task Automation

- Compared to Cron Jobs

- Compared to Celery

- Scalability and Reliability

- Directed Acyclic Graphs (DAGs)

- Workflows as Code

Workflows as Code (no programming)

- Anatomy of a DAG

- Directed Acyclic Graphs

- Operators

- Tasks

- Variables

- XComs

- Providers

- Connections

- Explore how DAG parts connect to the UI

- DAG Serialization

- Listeners

- Schedulers

- Pools

Installation and Configuration

- Python Virtual Environment

- Install Airflow

- Airflow Constraints File

- Standalone Mode

- Run the Webserver and Scheduler Independently

- SQLite vs PostgreSQL

- Configure with PostgreSQL

- Airflow and Kubernetes (with Minikube)

- Airflow and AWS Elastic Kubernetes Service (EKS)

- Airflow Helm Chart

Hands-On Kubernetes (K8s)
- Containerization and Orchestration

- Kubectl

- Helm

- Nodes

- Namespaces

- Pods, Containers, and Services

- Connect to the Internet (EKS)

- Keda Autoscaler

- Pod Logs

- SSH into Pods/Containers

- Live Upgrading Airflow

Airflow Configuration
- Airflow Configuration File Location

- Airflow Executor Configuration

To discuss this course and customizations:
Call: 434-509-5680 or Email: sales@cloudcontraptions.com

- Airflow Log Levels

- Helm Chart Configuration

- Learn How to Configure Airflow and K8s Pods

- Local Executor

- Celery Executor

- K8s Pod Executor

Airflow Custom Image

- Airflow Container Image

- Why Create a Custom Image?

- Create a Custom Image

- Install Software with Apt

- Install Software with PyPi

- Install Providers and Custom Software

- Use the Custom Container Image

Monitoring

- Logging

- Log File Structure

- Log Levels

- Review Task Logs in the Web UI

- External Log Storage

To discuss this course and customizations:
Call: 434-509-5680 or Email: sales@cloudcontraptions.com

- Metrics Configuration

- Monitor with Grafana

- Notifications