ADMIN 336

Airflow Administration and Operation in Cloudera Data Engineering

Course Overview

Course Type

Instructor-Led Training

Level

Intermediate

Duration

4 days

Platform

Cloudera on Premise
Cloudera on Cloud

Topics Covered

- Airflow Architecture & Core Concepts
- DAG Development & Workflow Orchestration
- Scheduling, Automation & Task Management
- Airflow Configuration & Administration
- CDE Integration & Airflow Job Orchestration
- Kubernetes-Based Airflow Deployment in CDE
- User Management, RBAC & Security
- Airflow Services & Operational Management
- Monitoring, Metrics, Logging & Alerting
- Troubleshooting & Standard Operating Procedures
- Airflow CLI & REST API
- Airflow 2 vs. Airflow 3
- Airflow 3 Assets, Dynamic Task Mapping & Human-in-the-Loop
- External Airflow Integration with CDE
- Standalone Airflow Monitoring & Alert Reporting

About This Training

This four-day hands-on training provides a practical introduction to Apache Airflow and its integration with Cloudera Data Engineering (CDE). Participants will learn how to develop and schedule DAGs, orchestrate CDE jobs, configure and manage Airflow services, understand its Kubernetes-based deployment, and implement security, monitoring, alerting, and operational procedures. The course also covers the key differences between Airflow 2 and Airflow 3, including new scheduling, asset, task execution, and workflow management capabilities. Through hands-on exercises, participants will gain the skills needed to develop, operate, monitor, and troubleshoot Airflow workflows in a CDE environment.

What Skills You Will Gain

During this course, you learn:

- Develop, schedule, and manage Airflow DAGs and workflows.
- Orchestrate Spark and other CDE jobs using Airflow.
- Configure and administer Airflow services and environments.
- Understand Airflow architecture and Kubernetes-based deployment in CDE.
- Configure connections, variables, pools, APIs, and authentication.
- Implement monitoring, metrics, logging, and alerting for Airflow.
- Troubleshoot failed, stuck, and performance-critical workflows.
- Apply security, RBAC, and DAG-level access controls.
- Integrate external Airflow environments with CDE.
- Understand key differences between Airflow 2 and Airflow 3 and apply new Airflow 3 features.

Who Should Take This Course?

This course is primarily designed for Airflow and CDE administrators, data platform administrators, and DevOps professionals responsible for deploying, configuring, managing, monitoring, and troubleshooting Airflow environments. It is also suitable for data engineers and developers who need to understand Airflow administration and production operations.

© 2026 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries.
All other trademarks are the property of their respective companies.
Information is subject to change without notice.

ADMIN 336

Airflow Administration and Operation in Cloudera Data Engineering

Day-1

Chapter 1: Introduction to Airflow

- What Airflow Is and Is Not
- Foundational Core Concepts
- Airflow 1.x vs 2.x Major Differences
- Platform Limitations and Expectations

Chapter 2: Architecture and components

- High-Level Architecture Overview
- Core Components
- How Components Interact
- How XCom Works Internally
- Executor Types
- High Availability (HA) Considerations
- Hand-on Exercise: Exploring Airflow Concepts
- Hand-on Exercise: Passing Data Between Tasks with XComs

Chapter 3: Introduction to Cloudera Data Engineering (CDE)

- What is CDE?
- CDE vs. Standalone Airflow
- Key CDE Concepts
- Hand-on Exercise: Setting Up Your CDE Environment

Chapter 4: Writing Your First DAG

- DAG File Structure
- Common Operators
- Task Dependencies
- Airflow Hooks
- Airflow Variables
- Jinja Templating
- Hand-on Exercise: Build and Deploy Your First CDE DAG
- Hand-on Exercise: Implementing Branching Logic
- Hand-on Exercise: Orchestrating CDE Jobs with Airflow
- Hand-on Exercise: Jinja Templating

Chapter 5: Scheduling and automation

- Cron Expressions & Timetables
- Catchup & Backfill Behavior
- BranchPythonOperator
- Trigger Rules
- Dataset-Driven Scheduling (Airflow 2.4+)
- SLA Definitions & Miss Callbacks
- Hand-on Exercise: Scheduling and Catchup Behavior
- Hand-on Exercise: Dynamic Task Generation
- Hand-on Exercise: Dataset-Driven Scheduling Exercise
- Hand-on Exercise: Multi-Step Spark Pipeline

Day-2

Chapter 6: Configuring Airflow

- airflow.cfg Sections & Environment Variable Overrides
- Connections, Variables, and Pools
- Metadata Database Setup & Migration
- Log Destinations (Local, S3, GCS)
- SMTP for Alerting

Chapter 7: Anatomy of Airflow's Kubernetes Deployment in CDE

- Connecting to CDE K8s virtual cluster
- Understanding Airflow Components in K8s cluster
  - Deployment
  - Configmap & Secret
  - Role & Rolebinding
- Understanding components in monitoring (Prometheus) & DEX base namespace (Grafana)
  - Deployment & Daemonset in monitoring namespace

© 2026 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries.
All other trademarks are the property of their respective companies.
Information is subject to change without notice.

VERSION 2.0.0

ADMIN 336

Airflow Administration and Operation in Cloudera Data Engineering

- Configmap & Service
- Checking which Pods back the Service in monitoring namespace
- Exposing the Airflow Web UI via Ingress with ELB
- Enable enabling Grafana login

Chapter 8: Managing Airflow services

- Starting, Stopping, and Restarting Services
- Health-Check Endpoints
- Upgrading Airflow
- DAG Deployment Strategies
- Managing Python Dependencies & Provider Packages
- Deployment and CI/CD Pipelines

Chapter 9: User and role authorizations

- RBAC Model (Role-Based Access Control)
- Built-in Roles
- Custom Roles
- Integration with LDAP, OAuth2, or SSO Providers
- Audit Logging
- DAG-Level Access Control

Day-3

Chapter 10: Reporting

- Using the Airflow UI (Run History & SLA Miss Reports)
- Exporting Metrics to Prometheus via StatsD Exporter
- Building Grafana Dashboards on Metadata DB
- Scheduled Email Digests via Callbacks
- Configuring Monitoring and Metrics Reporting for Airflow in CDE

Chapter 11: Installing client

- Installing the Airflow CLI and Required Providers
- Installing the Airflow CLI and Required Providers
  - Configuring Core Environment Variables
  - Authenticating Against the Webserver REST API
  - Verifying Connectivity with CLI Commands
- Installing the Airflow CLI in CDE
- Configuring REST API in CDE to access to Airflow

Chapter 12: Monitoring and alerting

- Key Metrics to Watch
- Setting Up Email and Slack Callbacks
- Integrating with Grafana and PagerDuty
- Reading the Gantt View for Bottleneck Identification
- Setting Up Airflow Alerts in CDE
- Hand-on Exercise: Debugging a Broken DAG
- Hand-on Exercise: Deferrable Operator + Trigger pattern
- Hand-on Exercise: Failure Handling

Chapter 13: Standard operating procedure

- Manually Triggering a DAG Run
- Clearing Failed or Stuck Tasks
- Pausing and Unpausing DAGs
- Handling Zombie Tasks
- Escalation Matrix
- Change Freeze Procedures & On-Call Handoff
- Post-Incident Log Collection & RCA Preparation
- Hand-on Exercise: Connecting External Airflow to CDE

Day-4

Chapter 14: Airflow 3 vs. Airflow 2

- Release & Lifecycle
- Architecture
- UI & Authentication
- DAG Authoring

© 2026 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries.
All other trademarks are the property of their respective companies.
Information is subject to change without notice.

ADMIN 336

Airflow Administration and Operation in Cloudera Data Engineering

- Scheduling Semantics
- REST API & CLI
- Breaking Changes
- Task Execution Model
- Databases & Metadata
- Operations & Monitoring
- Platform & Ecosystem
- Migration Path
- Essential Points
- Hand-on Exercise: Asset-Driven Scheduling
- Hand-on Exercise: Extras & Pull-Based Watchers
- Hand-on Exercise: TaksGroup
- Hand-on Exercise: Asset-Driven Scheduling
- Hand-on Exercise: Extras & Pull-Based Watchers
- Hand-on Exercise: TaksGroup
- Hand-on Exercise: Deadline Alerts
- Hand-on Exercise: Time Sensor
- Hand-on Exercise: Dynamic Task Mapping
- Hand-on Exercise: Xcom with Task_ID
- Hand-on Exercise: Human-in-the-Loop
- Hand-on Exercise: Parition
- Hand-on Exercise: Rollup
- Hand-on Exercise: Connecting External Airflow3 to CDE

Appendix: Building Standalone Airflow Monitoring & Alert Reporting

© 2026 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries.
All other trademarks are the property of their respective companies.
Information is subject to change without notice.

VERSION
2.0.0