

Apache Airflow

Course Duration ~50 Hours · 7 Weeks · Live Online · Lab-Heavy · Portfolio of Running Pipelines

Curriculum reviewed: July 2026 — Airflow 3 is evolving across minor releases; this syllabus tracks the current line

[Airflow 3.x](#)
[Assets](#)
[Event-Driven](#)
[DAG Versioning](#)
[Production Ops](#)
[2 → 3 Migration](#)

In short: Airflow 3 isn't a version bump. If you learned Airflow 2, several things you were taught no longer exist — SubDAGs, *execution_date*, XCom pickling, direct database access from task code — and Airflow 2 has reached the end of its open-source support life while a large share of teams still run it. This course teaches the current product: Assets replacing Datasets with data-driven scheduling, event-driven triggering instead of waiting for cron, DAG versioning for debugging and audit, the Task Execution API and Task SDK's service-oriented execution model, and the rewritten UI. **Migration is a first-class track, not a footnote:** you assess a legacy estate, run automated detection and fixes, lint configuration, and plan a staged cutover with rollback. Across 8 modules you ship asset-driven and event-driven pipelines, a 3am-hardened reliability build, a tested CI/CD DAG repository, an orchestrated Spark-to-warehouse-to-BI stack, an ML retraining trigger, a Kubernetes production deployment — and a full Airflow 2 → 3 migration exercise. Written for engineers: fresh IT graduates entering data engineering, and teams still on Airflow 2 who need to migrate. Aligned to Astronomer's Apache Airflow 3 certifications (Fundamentals and DAG Authoring).

Duration	7 weeks · ~50 hours · live + hands-on	Format	Live online · lab-heavy · portfolio of running pipelines
Prerequisite	Working Python, basic SQL, comfort with the command line and Docker basics. No prior orchestration experience required.	Curriculum	8 modules, from orchestration fundamentals to a first-class Airflow 2 → 3 migration track
Certification	Aligned to Astronomer's Apache Airflow 3 certifications (Fundamentals and DAG Authoring)	Ideal for	Fresh IT graduates, data/ML/platform engineers, and teams still on Airflow 2 who need to migrate

When Airflow Is the Wrong Tool

For sub-second streaming, simple cron jobs, or in-application task queues, Airflow's operational weight isn't justified — a stream processor, a crontab, or a job queue is the right answer. Teams also genuinely evaluate alternatives like Dagster and Prefect, and this course covers that comparison fairly. Knowing when *not* to deploy an orchestrator — and which one fits — is a senior-engineer signal, and we teach the boundary explicitly.

Course Highlights

- Every module leads with what you build and ship, not what you "learn about"
- Airflow 3 throughout: Assets, event-driven scheduling, DAG versioning, the Task SDK
- A first-class Airflow 2 → 3 migration track — the removals that bite, the tooling that catches them
- Reliability engineering: idempotency, safe backfills, tasks you can rerun without corrupting data
- Testing and CI/CD for DAGs — pipelines treated as software, not scripts
- Integrations in practice: Spark, Snowflake, dbt, Kubernetes, ML retraining triggers
- Production operations plus honest build-vs-buy on Astronomer, MWAA and Cloud Composer

Hands-on Projects & Portfolio

- **Multi-Source ELT Pipeline** — dynamic task mapping, retries, SLAs and alerting
- **Asset-Driven Pipeline** — downstream DAGs trigger on data updates rather than on a schedule
- **Event-Driven Workflow** — reacts to an external system rather than to cron
- **Orchestrated Multi-Tool Stack** — extract, transform in Spark, load to a warehouse, refresh a BI layer
- **ML Retraining Pipeline** — triggered by a drift signal, with model validation gates
- **Tested, CI/CD-Deployed DAG Repo** — unit-tested DAGs with environment promotion

- **Airflow 2 → 3 Migration Exercise** — a deliberately legacy DAG set, with an assessment report and a cutover plan
- **Capstone** — production orchestration platform with monitoring, RBAC and a documented runbook

Tools & Stack

Apache Airflow 3.x · Python · Docker · Kubernetes · Celery · PostgreSQL · dbt · Apache Spark · Snowflake · Apache Kafka · AWS · GCP · Azure · Astronomer · AWS MWAA · Cloud Composer · Git · GitHub Actions · Prometheus · Grafana

Certification Alignment

- Syllabus aligned to **Astronomer's Apache Airflow 3 Fundamentals** certification — architecture, scheduling, operations
- Modules 2–5 cover the core of **DAG Authoring for Apache Airflow 3** — TaskFlow API, dynamic task mapping, templating, asset-driven scheduling
- Presented as preparation and alignment — no official partnership or accreditation implied, and no exam outcome guaranteed

Career Outcomes

Airflow is the de facto standard for data workflow orchestration and appears in a very large share of data engineering job descriptions — one of the highest-leverage single tools in the data stack. The version transition has created immediate demand for engineers who know the current model and can migrate legacy estates. This course targets:

- Data Engineer · Analytics Engineer · MLOps Engineer · Data Platform Engineer
- Workflow / Orchestration Engineer · DevOps Engineer (data platforms) · Backend Engineer owning scheduled systems · Data Architect

Who This Course Is For

- Final-year students and fresh IT graduates targeting data engineering roles
- Python developers moving into data engineering
- Data analysts and analytics engineers automating manual, scheduled work
- ML engineers orchestrating training and retraining pipelines
- DevOps and platform engineers who now operate someone else's Airflow
- **Teams still on Airflow 2 who need to migrate** — this course was built with you in mind

Module 1: Orchestration Fundamentals & Airflow 3 Architecture

6 Hours

YOU SHIP

- A running Airflow 3 environment stood up from scratch, with an annotated architecture map — scheduler, API server, workers, triggerer — and a defended executor choice for a given workload

TOPICS

- What an orchestrator is actually for; DAGs as code
- The scheduler, API server, workers, metadata database and triggerer; the service-oriented execution model
- Executor choices (Local, Celery, Kubernetes, Edge) and the trade-offs behind each

Assessment: Quiz + Lab

Module 2: Authoring DAGs

7 Hours

YOU SHIP

- A multi-source ELT DAG with dynamic task mapping, task groups and branching — written in the modern SDK namespace, readable at fifty tasks

TOPICS

- The modern SDK namespace and the TaskFlow API; tasks and dependencies; operators, sensors and hooks
- XComs and their limits; connections and variables; templating and macros

- Dynamic task mapping; task groups; branching and trigger rules; DAG design patterns that stay readable

Assessment: Quiz + Lab

Module 3: Scheduling & Assets

6 Hours

YOU SHIP

- An asset-driven pipeline where downstream DAGs trigger on data updates, plus an event-driven workflow that reacts to an external system instead of cron

TOPICS

- Schedules and timetables; catchup and backfill done properly; data intervals versus the newer non-interval model
- Assets and data-driven scheduling — among the most-used features in practice
- Cross-DAG dependencies; event-driven triggering from external systems

Assessment: Quiz + Lab

Module 4: Reliability Engineering for Pipelines

6 Hours

YOU SHIP

- A pipeline hardened for 3am: idempotent tasks, retries with backoff, SLAs, alerting callbacks, and a safe backfill run that doesn't corrupt downstream data

TOPICS

- Idempotency as a design requirement; retries and exponential backoff; timeouts and SLAs
- Alerting and callbacks; partial failure and recovery
- Backfilling safely; tasks you can rerun without corrupting downstream data

Assessment: Quiz + Lab

Module 5: Testing, Versioning & CI/CD

6 Hours

YOU SHIP

- A tested, CI/CD-deployed DAG repository with validation gates and environment promotion — and DAG version history used to debug a regression

TOPICS

- Unit-testing DAGs and tasks; DAG validation in CI; local development environments
- DAG versioning and what it changes about debugging and audit
- Deployment strategies; managing dependencies across environments

Assessment: Quiz + Lab

Module 6: Integrations in Practice

6 Hours

YOU SHIP

- An orchestrated multi-tool stack — extract, transform in Spark, load to a warehouse, refresh a BI layer — plus an ML retraining DAG triggered by a drift signal with validation gates

TOPICS

- Providers and the operator ecosystem; orchestrating Spark jobs, Snowflake and warehouse loads, dbt runs
- Kafka-adjacent event workflows; cloud services on AWS/GCP/Azure; containerised and Kubernetes-native tasks
- ML pipelines including training and retraining triggers

Assessment: Quiz + Lab

Module 7: Running Airflow in Production

7 Hours

YOU SHIP

- A production deployment on Kubernetes — scaled schedulers and workers, secrets backend, RBAC for two teams, remote logs, metrics — plus a written build-vs-buy recommendation across the managed options

TOPICS

- Deploying on Kubernetes and with Docker; scaling schedulers and workers; resource isolation; secrets backends
- RBAC and multi-team tenancy; logging and remote log storage; monitoring and metrics; performance tuning and cost
- Managed options — Astronomer, AWS MWAA, Google Cloud Composer — with honest build-vs-buy reasoning

Assessment: Quiz + Lab**Module 8: Migration & Capstone**

6 Hours

YOU SHIP

- A full Airflow 2 → 3 migration of a deliberately legacy DAG set — assessment report, automated fixes, staged cutover and rollback plan — and your capstone: a production orchestration platform, defended

TOPICS

- Assessing an existing estate; the removals and breaking changes that bite — SubDAGs, execution_date, schedule handling, XCom pickling, direct database access
- Automated detection and fixes; configuration linting; staged cutover, rollback planning and validation
- Capstone design, build and defense

Assessment: Capstone Project + Review