+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated on 08/26/2026

Sign Up

OpenLineage Training

2 days (14 hours)

Overview

OpenLineage is an open standard for data traceability, lineage, and pipeline observability. With its Run, Job, and Dataset object model, extensible facets, and integrations with data engineering tools, OpenLineage helps ensure the reliability of modern data processing pipelines.

Our OpenLineage training will go beyond an introduction to the standard, enabling you to master the practical challenges of instrumenting and monitoring data processing in a real-world environment.

You'll learn to understand the object model, leverage facets, structure lineage events, and integrate OpenLineage with data observability and governance practices.

By the end of the training, you'll be able to design a coherent traceability strategy, better diagnose dependencies between jobs, improve metadata quality, and enhance the transparency of data flows.

This training also covers the OpenLineage ecosystem, best practices for integration, production-oriented use cases, and implementation methods tailored to data engineering teams.

Like all our training courses, this one will introduce you to **the latest stable version** of the technology and its new features.

Objectives

- Understand the role of OpenLineage in the traceability of data processing.

- Master the conceptual model based on Run, Job, and Dataset.
- Identify and leverage facets to enrich lineage events.
- Structure an observability approach for pipelines tailored to a production environment.
- Implement best practices for instrumentation and integration with the data ecosystem.
- Analyze dependencies, incidents, and metadata quality to ensure the reliability of data flows.

Target Audience

- Data Engineer
- Data Engineer
- Data Architect
- Data Platform Manager
- ETL Developer

Requirements

- Basic knowledge of data engineering and pipeline orchestration.
- Understanding of metadata, processing flows, and dependencies between jobs.
- Experience with a cloud environment or a data platform such as Airflow, dbt, or equivalent.
- Proficiency in reading structured formats such as JSON and an understanding of API interactions.

Technical Prerequisites

- A recent computer with sufficient memory to run the data manipulation tools and the provided examples.
- A stable Wi-Fi or Ethernet connection to access training resources and remote environments.
- An up-to-date browser and a compatible code editor, such as VS Code.
- Depending on the workshops, a local environment with Java or Python 3.x installed, as well as Git.
- For remote sessions, a working microphone and headset for discussions and guided exercises.

OpenLineage Training Program

[Day 1 - Morning]

Foundations and Key Concepts

- Understand the role of OpenLineage in lineage data and observability.
- Identify the problems solved by an open standard for processing traceability.
- Discover the functional scope, integrations, and main use cases.
- Master the basic vocabulary: run, job, dataset, and events.
- Connect OpenLineage to governance, quality, and diagnostic challenges.
- Hands-on workshop: mapping an existing pipeline and identifying the lineage information to capture.

[Day 1 - Afternoon]

Object Model and Events

- Explore the OpenLineage object model and the relationships between entities.
- Understand the structure of an event and the minimum required information.
- Analyze the distinction between static metadata and runtime metadata.
- Study the principles of object identification, naming, and consistency.
- Discover the concepts of the backend and event transport to a target.
- Hands-on workshop: Manually build a simple OpenLineage event based on a processing scenario.

[Day 2 - Morning]

Facets and Enrichment

- Understand the role of facets as an extension of the base model.
- Distinguish between run facets, job facets, and dataset facets.
- Discover how facets can be used to enrich production metadata.
- Learn how to read a facet specification and interpret its structure.
- Relate facets to governance, audit, and observability needs.
- Hands-on workshop: Analyze several facets and select the ones to use for a given business case.

[Day 2 - Afternoon]

Integration and Deployment

- Explore instrumentation strategies for jobs and pipelines.
- Understand best practices for integration with data engineering tools.
- Structure a deployment approach focused on production and operations.
- Address quality control, dependency tracking, and incident diagnosis.
- Identify key considerations to ensure the reliability of metadata collection and usage.
- Hands-on workshop: Design an OpenLineage integration plan for a target data architecture.

Relevant Companies

This training program is designed for both individuals and businesses—large and small—that wish to train their teams in new, advanced IT technologies or to acquire specific industry knowledge or modern methods.

Placement Assessment at the Start of Training

The pre-training assessment complies with Qualiopi quality standards. Upon final registration, the learner receives a self-assessment questionnaire that allows us to evaluate their estimated proficiency level across various types of technologies, as well as their expectations and personal goals for the upcoming training, within the limits imposed by the selected format. This questionnaire also enables us to anticipate certain connection issues or

internal security issues within the company (in-house or virtual classroom) that could hinder the monitoring and smooth operation of the training session.

Teaching Methods

Hands-On Course: 60% Hands-On, 40% Theory. Training materials distributed in digital format to all participants.

Organization

The course alternates between theoretical input from the instructor—supported by examples and reflection sessions—and group work.

Assessment

At the end of the session, a multiple-choice quiz is administered to verify that participants have successfully acquired the skills.

Certification

A certificate will be issued to each trainee who has completed the entire training program.