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

Updated on 03/20/2026

Sign up

Apache Atlas Training

3 days (21 hours)

Overview

Apache Atlas is an open-source data governance solution designed for Big Data environments. It enables you to centralize metadata, visualize data lineage, and strengthen control over data assets in distributed architectures.

Our Apache Atlas training will help you understand the principles of data cataloging, traceability, and classification in a modern governance environment.

You will learn how to install, configure, administer, and automate the platform, while integrating it into an ecosystem that uses Hadoop, Hive, Spark, or Ranger.

By the end of the training, you will be able to structure a data catalog, implement data lineage, secure access, and deploy a governance approach tailored to your organization's needs.

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 fundamental concepts of data governance.
- Install and configure Apache Atlas in a Big Data environment.
- Create and manage metadata, entities, and classifications.
- Leverage data lineage to analyze data flows and impacts.
- Implement Atlas security, compliance, and automation.

Target audience

- Data engineers
- Data architects
- Big Data administrators
- Data DevOps profiles and technical governance managers

Prerequisites

- Basic knowledge of Big Data architectures
- Understanding of distributed processing and metadata
- General experience in technical administration or data environments

Apache Atlas Training Program

[Day 1 - Morning]

Fundamentals of data governance with Apache Atlas

- Understanding the challenges of data governance
- Introduction to the Big Data ecosystem and the role of Apache Atlas
- Mastering key concepts: metadata, data lineage, classification
- Exploring the overall architecture of Apache Atlas
- Identifying the main use cases in the enterprise
- Hands-on workshop: exploring the Atlas interface and navigating the catalog.

[Day 1 - Afternoon]

Installation, configuration, and getting started

- Review technical prerequisites and the target environment
- Install Apache Atlas in a Hadoop environment
- Configure the platform’s initial settings
- Getting familiar with the user interface and main features
- Create the first metadata objects
- Hands-on workshop: installation and initial configuration of an Atlas environment.

Modeling and organizing metadata

- Create types, entities, and relationships
- Managing classifications, labels, and catalog structure
- Organizing a data catalog that teams can use
- Using search and exploration features
- Applying best practices for business and technical modeling
- Hands-on workshop: modeling a business metadata repository.

[Day 2 - Morning]

Traceability, lineage, and impact analysis

- Understanding the principles of data lineage
- Visualizing data flows between systems and processes
- Leveraging integration with Apache Hive and Apache Spark
- Performing an impact analysis on transformations
- Implementing reliable traceability in a data chain
- Hands-on workshop: analyzing a data flow and visualizing the lineage.

[Day 2 - Afternoon]

Security, advanced governance, and compliance

- Defining security policies for metadata
- Integrating Apache Ranger for access governance
- Managing roles, permissions, and access controls
- Addressing compliance and GDPR challenges
- Leveraging audit and traceability of operations
- Hands-on workshop: Implementing security and governance rules.

APIs, automation, and integration into the data ecosystem

- Using the Apache Atlas REST API
- Automating metadata ingestion and updates
- Integrating Atlas with third-party tools and existing workflows
- Synchronize metadata across multiple environments
- Set up platform monitoring and maintenance
- Hands-on workshop: Automating Atlas actions via API.

[Day 3 - Morning]

Optimization, industrialization, and best practices

- Optimizing platform performance and scalability
- Managing large volumes of metadata
- Defining a sustainable data governance strategy
- Standardizing administration and update processes
- Exploring advanced architectures based on Atlas
- Hands-on workshop: optimizing an existing Atlas environment.

[Day 3 - Afternoon]

Advanced use cases and multi-source governance

- Deploying Apache Atlas in a Data Lake context
- Implementing governance across multiple data sources
- Using Atlas as a data catalog and data discovery tool
- Fostering collaboration between data, business, and governance teams
- Analyzing feedback and real-world business cases
- Hands-on workshop: comprehensive case study on multi-source governance.

Overarching project and operational simulation

- Define a business use case based on a data catalog
- Set up a comprehensive catalog with entities and classifications
- Implement lineage and traceability mechanisms
- Secure access and structure the target governance
- Validate architectural choices and present results
- Hands-on workshop: supervised final project on deployment and governance with Apache Atlas.

Target Audience

This training is intended for both individuals and companies, large or small, seeking to train their teams in new advanced IT technologies or to acquire specific business knowledge or modern methodologies.

Assessment upon enrollment

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 in 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 allows us to anticipate certain connection or internal security issues within the company (intra-company or virtual classroom) that could pose challenges for monitoring and ensuring the smooth running of the training session.

Teaching Methods

Practical Training: 60% practical, 40% theoretical. Training materials will be 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 questionnaire is used to verify that the skills have been properly acquired.

Certification

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