

Updated on 08/26/2026

Sign Up

Apache Gravitino Training

2 days (14 hours)

Overview

Apache Gravitino is a federated metadata lake platform that centralizes governance, data access, and catalog management for modern data environments.

This Apache Gravitino training course helps you understand the role of a unified metadata layer in a multi-source, multi-engine data lakehouse architecture. You'll learn how to structure a consistent approach to governance, discovery, and access control.

You'll see how Gravitino integrates with relational sources, file storage systems, message streams, and analytics environments, while exposing APIs and management mechanisms tailored to enterprise use cases. The goal is to learn how to organize metadata in a way that is readable, robust, and actionable by data teams.

The training also covers the concepts of catalogs, unified metadata, centralized security, and interoperability with processing engines. You'll gain the autonomy to set up a technical foundation tailored to production needs.

By the end of this course, you will be able to design a clear Gravitino architecture, manage day-to-day operations, and lay the groundwork for a more governed and scalable data architecture.

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 architecture and positioning of Apache Gravitino.

- Master the logic of catalogs and federated metadata.
- Configure data sources and set up centralized governance.
- Use the APIs and basic administration mechanisms.
- Integrate Gravitino with data lakehouse and multi-engine use cases.
- Apply best practices for security, operations, and data structuring.

Target Audience

- Data Engineer
- Data Architect
- Data Platform Administrator
- Data Governance Engineer
- Data Platform Engineer

Prerequisites

- Basic knowledge of data architecture and governance.
- Experience with lakehouse, SQL, or data catalog environments.
- Understanding of storage, metadata, and access control concepts.
- Previous experience with a Java ecosystem or enterprise data tools.

Technical Prerequisites

- A recent computer with sufficient memory to run the training tools.
- A stable Wi-Fi or Ethernet connection for online exercises.
- Java 17 installed and properly configured in the PATH or via JAVA_HOME.
- A modern web browser and, if needed, a Docker client for demonstrations.
- A suitable code editor or IDE, as well as a headset for the virtual classroom.

Apache Gravitino Training Program

[Day 1 - Morning]

Gravitino Fundamentals

- Understand the role of Apache Gravitino in a modern data architecture.
- Identify the challenges of unified metadata, governance, and traceability.
- Explore the concepts of a metadata lake, a catalog, and a single source of truth.
- Position Gravitino for lakehouse and multi-engine use cases.
- Review the functional architecture, service layers, and key business objects.
- Hands-on workshop: Map a data use case and identify the metadata to be centralized.

[Day 1 - Afternoon] Installation

and Getting Started

- Set up the local environment and verify the necessary dependencies.
- Install and start the core components in a demo environment.
- Explore the administration interface, the CLI, and the initial access points.
- Understand the logic behind REST APIs and client-server interactions.
- Review the structure of a Metalake and its top-level objects.
- Hands-on workshop: Deploy a test instance and run the first inventory commands.

[Day 2 - Morning]

Catalog Management

- Create and manage catalogs based on supported source types.
- Understand the differences between relational, file, messaging, and template catalogs.
- Configure key properties, locations, and associated metadata.
- Organize the discovery, viewing, and updating of managed objects.
- Apply structuring rules to maintain consistent naming conventions.
- Hands-on workshop: Configure multiple catalogs and verify their visibility in Gravitino.

[Day 2 - Afternoon]

Governance, Security, and Advanced Uses

- Implement centralized governance for metadata and access.
- Understand security, control, and audit mechanisms.
- Align Gravitino with data platform requirements and integration with computing engines.
- Understand possible extensions for policies, associated services, and ecosystem developments.
- Prepare a plan for operation, maintenance, and scaling.
- Hands-on workshop: Define a governance strategy and simulate a production deployment scenario.

Target Audience

This training is intended for both individuals and companies—large and small—that wish to train their teams in a new, advanced IT technology or to acquire specific business knowledge or modern methods.

Placement at the Start of Training

The assessment upon enrollment complies with Qualiopi quality standards. Upon final registration, the learner receives a self-assessment questionnaire that allows us to evaluate their estimated proficiency with various types of technology, 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 operation of the training session.

Teaching Methods

Hands-On Training: 60% Practical, 40% Theoretical. 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.