

Updated on July 15, 2026

Sign Up

Apache Polaris Training

2 days (14 hours)

Overview

Apache Polaris is an open-source Iceberg catalog that centralizes metadata, secures access, and simplifies interoperability between SQL engines, Spark, Flink, and other REST-compatible environments.

Our Apache Polaris training will help you go beyond the basics to master the key challenges of a modern, reliable, and multi-engine-ready data catalog.

You'll learn to understand Polaris's architecture, structure catalogs, namespaces, and Iceberg tables, and then leverage the service's governance, authentication, and automation capabilities.

By the end of the training, you will be able to deploy and administer an Apache Polaris platform, manage access rights, prepare integrations with major processing engines, and ensure the reliability of read and write operations on data.

This training also covers best practices for configuration, security, monitoring, and operations to build a consistent technical foundation for data engineering architectures in real-world environments.

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 Apache Polaris in an Apache Iceberg ecosystem.
- Master the creation and organization of catalogs, namespaces, and tables.

- Configure authentication, authorization, and governance mechanisms.
- Prepare for the integration of Polaris with various query and processing engines.
- Adopt best practices for operations, security, and maintenance.

Target Audience

- Data Engineer
- Data Architect
- Data Platform Engineer
- Data Administrator
- Data Developer

Prerequisites

- Basic knowledge of data architecture and data modeling.
- Prior experience with a processing or query engine such as Spark, Flink, or Trino.
- Familiarity with Apache Iceberg or next-generation table format concepts.
- Understanding of REST API principles and client-server communication.
- Basic experience in system administration or operating a data platform.

Technical Prerequisites

- A recent computer with sufficient RAM to run the demo tools and local environments.
- A stable Wi-Fi or Ethernet connection to participate in the workshops and access training resources.
- A development environment with Java 21+, Docker 27+, and an up-to-date web browser.
- An IDE or code editor compatible with Python 3.x and the configuration files provided during the training.
- For virtual classes, a working microphone and headset, as well as the necessary permissions to run Docker locally.

Apache Polaris Training Program

[Day 1 - Morning] Polaris

Fundamentals

- Apache Polaris's role within the Apache Iceberg ecosystem.
- The role of a REST catalog in governance and table access.
- Understanding business objects, catalogs, namespaces, and tables.
- Overview of use cases, multi-engine support, and interoperability.
- Overview of the overall architecture and service components.

- Hands-on workshop: mapping a data use case and identifying where Polaris fits into the processing chain.

[Day 1 - Afternoon] Installation

and Getting Started

- Setting up the local environment and required dependencies.
- Starting the service and exploring the available interfaces.
- Creating the first catalogs and namespaces.
- Managing basic settings and naming conventions.
- Introduction to storage concepts and related configuration.
- Hands-on workshop: Install Polaris locally, create a catalog, and publish an initial data structure.

[Day 2 - Morning]

Security and Governance

- Principles of authentication and authorization in Polaris.
- Managing roles, permissions, and access control.
- Best practices for securing sensitive operations.
- Organizing governance around catalog resources.
- Action traceability and operational logic in a shared environment.
- Hands-on workshop: Defining a permissions model for multiple teams and testing differentiated access.

[Day 2 - Afternoon] Integration

and Operations

- Connecting Polaris to REST-compatible engines.
- Reading and writing tables via an Iceberg workflow.
- Managing routine operations and operational scenarios.
- Monitoring, diagnostics, and key considerations in production.
- Best practices for designing a robust and maintainable architecture.
- Hands-on workshop: Run a complete end-to-end scenario using a query engine and validate the catalog's behavior.

Target Audience

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

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 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 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.