Faysal Aziz

Apache Pulsar Fundamentals

Messaging architecture and a guided producer–consumer workflow

FA-0581 · 1 day

Data & Analytics · DevOps, Cloud & Infrastructure · Software Development

Explore Pulsar architecture and practise a small messaging workflow, with an overview of operational and advanced streaming features.

Introduction

This one-day course introduces developers, data engineers and architects to Apache Pulsar's messaging architecture and core operations. A guided exercise covers namespaces, topics, message production and consumption using a selected Java or Python client.

Multi-tenancy, geo-replication, tiered storage, functions and production management are introduced through discussion or selected demonstrations. The focus is a small learning deployment and an application sketch, not independent administration of a resilient production cluster or a complete real-time platform.

Learning outcomes

The course teaches participants to:

- Explain the roles of brokers, BookKeeper, metadata stores, topics and subscriptions.
- Set up or connect to a prepared Pulsar learning instance.
- Create a namespace or topic and produce and consume sample messages.
- Describe multi-tenancy, geo-replication, tiered storage and Pulsar Functions.
- Identify monitoring, recovery and integration considerations for a larger deployment.

Prerequisites

- Basic knowledge of distributed systems and messaging.

- Working familiarity with Java or Python; the exercise uses one selected language.
- Containerisation or orchestration knowledge is helpful but not essential for the basic client exercise.
- Access to a prepared learning instance or a suitable local environment.

Training outline

1. Module 1 — Messaging and Pulsar Concepts

- Overview of Apache Pulsar and its place in the data ecosystem
- Core features, potential uses and limitations of Pulsar.
- Comparison with other messaging and streaming platforms

2. Module 2 — Architecture

- Brokers, BookKeeper bookies and metadata-store coordination; compare ZooKeeper with other backends supported by the chosen release.
- The role of topics, subscriptions, and producers/consumers in Pulsar
- How the architecture separates serving, persistent storage and coordination; relate durability to configuration and failure assumptions.

3. Module 3 — Guided Messaging Exercise

- Set up or connect to a prepared learning instance and review basic configuration.
- Basic operations: Creating namespaces, managing topics, and configuring retention policies
- Producing and consuming messages: Hands-on examples

4. Module 4 — Advanced Capability Overview

- Multi-tenancy and geo-replication concepts, with an illustrative multi-region design.
- Tiered storage concepts and relevant configuration trade-offs.
- Demonstrate a small Pulsar Function where the prepared environment supports it.

5. Module 5 — Deployment and Operations Overview

- Review production deployment considerations rather than build a complete production cluster.
- Monitoring and managing cluster health and performance
- Investigate a selected common issue and discuss performance-tuning evidence.

6. Module 6 — Application Design and Review

- Design considerations for real-time streaming applications
- Integrating Apache Pulsar with external systems and services
- Discuss clearly illustrative application scenarios and trade-offs.
- Review the sample producer–consumer workflow and identify reliability, data handling and integration decisions still required.
- Summarise which concepts were practised and which need deeper deployment work.

© 2026 Faysal Aziz · Kuala Lumpur