Updated on 02/09/2026

Calista Framework Training

2 days (14 hours)

Overview

Calista is an open source, industrial-grade Data Quality (DQ) framework that enables the validation of massive volumes of data within Big Data environments.

Our Calista training will enable you to master the "Data Quality as Code" approach to automate your data checks. You will learn how to configure the engine, define complex validation rules, and integrate these tests directly into your ETL or Spark pipelines.

The goal of our training is to ensure that your data is reliable, accurate, and ready for decision-making analysis.

Like all our training courses, this one focuses on a practical and operational approach based on concrete production use cases.

Objectives

- Understand Calista's philosophy and architecture.
- Set up data profiles and control rules.
- Isolate and process data anomalies in an industrial manner.
- Integrate Calista into a CI/CD chain or orchestrator.
- Generate quality reports and actionable KPIs.

Target audience

- Data Engineers
- Data Quality Managers
- Data Architects

Prerequisites

- Basic knowledge of SQL
- Basic knowledge of Scala or Java
- Familiarity with the Spark/Hadoop ecosystem

Technical prerequisites

- Minimum 8 GB RAM, 16 GB recommended
- Docker environment or local installation of Spark
- An IDE (preferably IntelliJ IDEA)
- Access to a terminal (Bash or PowerShell)

Calista training: Data Quality & Industrialization

[Day 1 - Morning]

Fundamentals and

Architecture

- Understanding the challenges of Data Quality: the 6 key dimensions
- Introduction to Calista: "DQ as Code" philosophy and ecosystem
- Technical architecture: Spark engine, configurations, and metadata
- Comparison with other tools on the market (Great Expectations, Deequ)
- Installation and setup of the development environment
- Hands-on workshop: Initializing a Calista project and first scan of a dataset.

[Day 1 - Afternoon]

Implementation of controls

- Profiling: automatic analysis of structure and distribution
- Definition of Rules & Checks: types, formats (Regex), nullity, uniqueness
- Management of tolerance thresholds and alert levels
- Configuring outputs and execution logs
- Hands-on workshop: Creation of a comprehensive control plan for a business dataset.

[Day 2 - Morning]

Advanced Functions and Customization

- Cross-checks: inter-source consistency
- Calculation of custom KPIs and extension of the framework
- Management and isolation of anomalies (rejected data)
- Generation of non-compliance reports

- Hands-on workshop: Complex reconciliation between an SQL source and a CSV file.

[Day 2 - Afternoon] Industrialization

and monitoring

- CI/CD integration (Jenkins, GitLab CI)
- Using results for reporting and dashboards
- Optimizing performance on large volumes (Spark tuning)
- Production deployment and monitoring strategies
- Hands-on workshop: Complete pipeline from injection to final quality report.

Target companies

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

Positioning at the start of training

The positioning at the start of the training complies with Qualiopi quality criteria. Upon final registration, the learner receives a self-assessment questionnaire that allows us to assess their estimated level of proficiency in different types of technologies, as well as their expectations and personal objectives 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 be problematic for the monitoring and smooth running of the training session.

Teaching methods

Practical training: 60% practical, 40% theory. Training materials distributed in digital format to all participants.

Organization

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

Validation

At the end of the session, a multiple-choice questionnaire is used to verify that the skills have been correctly acquired.

Certification

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