

Apache NiFi training: Automate your data flows

2 days (14 hours)

Presentation

Our Apache NiFi training course will enable you to process and distribute data easily efficiently.

In fact, close an ETL, Apache NiFi is a tool that has proven its effectiveness, beyond the fact that it is highly configurable so that users benefit from guaranteed delivery, low latency, high throughput, dynamic prioritization and functionality for modifying flows at runtime.

Naturally component-oriented, NiFi's paradigm is flow-based programming (FBP). Messages of up to several gigabytes can be processed automatically in your "data factory". NiFi's highly productive, intuitive interface lets you create data flows with a simple drag-and-drop operation, without to write a single line of code.

What's more, the program is open-source and free of charge. Documentations referring to Apache NiFi are numerous. Software users benefit from customized reporting, a seamless interface for data processing with DAG technology, and rapid prototyping.

NiFi also supports secure protocols such as SSL, HTTPS and SSH. Our Apache NiFi training course is based on the latest version of the software, [Apache NiFi 2.3](#).

Objectives

- Install and configure Apache NiFi
- Understanding the architecture and fundamentals of data processing with Apache NiFi
- Mastering data flow management
- Integrating Kafka with NiFi

Target audience

Developers, Big Data professionals, Architects, System administrators, DevOps

Prerequisites

- Basic knowledge of Unix systems
- Basic knowledge of Java

Apache NiFi training program

Introduction to NiFi and the Data Ecosystem

- Overview of the Big Data ecosystem and NiFi's role
- Apache NiFi use cases and positioning
- Fundamentals of Flow-Based Programming
- Introduction to key NiFi concepts: FlowFile, Processor, Connection, Attributes

NiFi Installation, Configuration and Getting Started

- Detailed NiFi installation and initial configuration procedure (Standalone, Unix-based)
- Navigating and using NiFi's user interface (UI)
- Understanding processor and connection states
- Add, connect and configure your first processors (GenerateFlowFile, LogAttribute)
- Essential introduction to NiFi Expression Language (EL)
- Introduction to Controller Services: Role and management
- Understanding Execution Engines (Traditional vs Stateless - Conceptual)

Database and Flow Handling

- Working with attributes: UpdateAttribute, ExtractText
- File management: GetFile, PutFile, ListFile
- Basic routing of FlowFiles: RouteOnAttribute
- Introduction to Record Processors
- Configuration of Record Readers/Writers (CSV, JSON)
- Basic record processing : UpdateRecord, SplitRecord
- Specific JSON processing: EvaluateJsonPath, SplitJson
- Simple error handling and 'failure' relationship

In-depth integration with external systems (Focus Kafka & API)

- Review of key Apache Kafka concepts (Topic, Partition, Broker, Consumer Group)

- NiFi-Kafka integration (Consumption) : Configuring and using ConsumeKafkaRecord
- NiFi-Kafka integration (Production) : Configuring and using PublishKafkaRecord
- Best practices for NiFi-Kafka integration (Keys, Group ID, Formats)
- Interaction with REST APIs (Client): InvokeHTTP
- Exposing a simple API with NiFi: ListenHTTP, HandleHttpRequest/HandleHttpResponse
- Interaction with databases (SQL): QueryDatabaseTableRecord, PutDatabaseRecord

Administration, Monitoring and Optimization of NiFi Flows

- NiFi status monitoring via UI (Dashboard, Status, Bulletins)
- Data Provenance Analysis : Follow-up and investigation
- Advanced debugging techniques and error handling (queues, retry strategies)
- User and access management: Basic policies (Users, Groups, Global policies)
- Using Templates to reuse flows
- Introduction to Parameter Contexts for configuration management
- Performance optimization: Back Pressure, Yield Duration, Concurrent Tasks
- Introduction to Reporting Tasks for external monitoring
- Introduction to Flow Analysis Rules for flow validation

Advanced Management, Security and NiFi Ecosystem

- Introduction to NiFi Registry: Role and architecture
- Stream version management with NiFi Registry (Register, Import, Deploy)
- Integration with a Schema Registry (e.g. Confluent) for centralized schema management
- Security management: Key concepts (Authentication, Authorization)
- SSL/TLS certificate management for secure communication (Controller Service StandardSSLContextService)
- Introduction to NiFi Command Line Interface (CLI) for administration and automation
- Clustering management in NiFi: Concepts and benefits (High Availability, Scalability)
- (Optional) Introduction to MiNiFi for data collection at source (Edge/IoT)

Further information

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

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

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.