

Apache NiFi for Data Flow Automation and Integration

Course Description This course provides a comprehensive, hands-on introduction to Apache NiFi, focusing on building, managing, and optimizing data flows for enterprise integration. Participants will learn how to design pipelines, handle real-time streaming, ensure security, and integrate NiFi with modern data platforms.

Duration 5 Days (Instructor-led, hands-on training)

Pre-requisites

- Basic understanding of data integration concepts
- Familiarity with databases and SQL
- Exposure to distributed systems or cloud platforms (preferred)
- Working knowledge of Linux commands

Learning Objectives

- Understand NiFi architecture and core concepts
- Design, deploy, and monitor data flows
- Implement processors, controllers, and templates effectively
- Secure NiFi flows with authentication and authorization
- Integrate NiFi with Kafka, Hadoop, and cloud services
- Optimize performance and troubleshoot common issues

Day-wise Module Coverage

Day 1: Introduction and Fundamentals

Module 1: Overview of Apache NiFi

- Data flow automation concepts
- NiFi history and ecosystem
- Key use cases in enterprises
- NiFi vs. other ETL/streaming tools
- NiFi architecture overview

Module 2: NiFi Installation and Setup

- System requirements
- Installing NiFi on Linux/Windows
- NiFi UI walkthrough
- Configuring NiFi properties

- Starting and stopping NiFi services

Module 3: NiFi Basics

- FlowFile, Processor, and Connection concepts
- NiFi Registry introduction
- Creating your first data flow
- Using processors for ingestion
- Hands-on lab: Simple file ingestion

Day 2: Core Processors and Flow Design

Module 4: Working with Processors

- Processor categories and functions
- Configuring processor properties
- Scheduling and execution strategies
- Error handling and retries

Module 5: Connections and Queues

- Data provenance tracking
- Back pressure and prioritization
- Queues and load balancing
- Monitoring data lineage

Module 6: Templates and Reusability

- Creating and exporting templates
- Importing and versioning flows
- NiFi Registry integration
- Best practices for reusable flows

Day 3: Advanced Flow Management

Module 7: Controllers and Reporting Tasks

- Controller services overview
- Configuring database connections
- Reporting tasks for monitoring

Module 8: Data Transformation and Enrichment

- Using ExecuteScript processors

- JSON, XML, and CSV transformations
- Schema validation and enrichment

Module 9: Real-time Streaming Integration

- NiFi with Apache Kafka
- NiFi with Hadoop/HDFS
- NiFi with cloud storage (AWS S3, Azure Blob)

Day 4: Security and Administration

Module 10: NiFi Security Fundamentals

- User authentication and authorization
- Role-based access control
- SSL/TLS configuration

Module 11: Monitoring and Troubleshooting

- NiFi logs and diagnostics
- Provenance data analysis
- Troubleshooting common errors

Module 12: Performance Optimization

- Tuning NiFi properties
- Scaling NiFi clusters
- Load balancing strategies

Day 5: Enterprise Use Cases and Best Practices

Module 13: Enterprise Integration Patterns

- NiFi for IoT data ingestion
- NiFi for cloud data pipelines
- NiFi in hybrid architectures

Module 14: Governance and Compliance

- Data lineage and audit trails
- Compliance considerations (GDPR, HIPAA)

Module 15: Capstone Project

- End-to-end data flow design
- Integration with multiple systems
- Performance and security validation