

Cloudera DataFlow: Flow Management with Apache NiFi

Code:	APA-NIFI
Length:	3 days
URL:	View Online

This three-day hands-on training course provides the fundamental concepts and experience necessary to automate the ingest, flow, transformation, and egress of data using Apache NiFi. Along with gaining a grasp of the key features, concepts, and benefits of NiFi, participants will create and run NiFi dataflows for a variety of scenarios. Students will gain expertise using processors, connections, and process groups, and will use NiFi Expression Language to control the flow of data from various sources to multiple destinations. Participants will monitor dataflows, examine progress of data through a dataflow, and connect dataflows to external systems such as Kafka and HDFS. After taking this course, participants will have key knowledge and expertise for configuring and managing data ingestion, movement, and transformation scenarios for the enterprise.

Skills Gained

Students who successfully complete this course will be able to:

- Understand the role of Apache NiFi and MiNiFi in the Cloudera DataFlow platform
- Describe NiFi's architecture, including standalone and clustered configurations
- Use key features, including FlowFiles, processors, process groups, controllers, and connections, to define a NiFi dataflow
- Navigate, configure dataflows, and use dataflow information with the NiFi User Interface
- Trace the life of data, its origin, transformation, and destination, using data provenance
- Organize and simplify dataflows
- Manage dataflow versions using the NiFi Registry
- Use the NiFi Expression Language to control dataflows
- Implement dataflow optimization methods and available monitoring and reporting features
- Connect dataflows with other systems, such as Kafka and HDFS
- Describe aspects of NiFi security

Who Can Benefit

This course is designed for Developers, Data Engineers, Data Scientists, and Data Stewards. It provides a no-code, graphical approach to configuring real-time data streaming, ingestion, and management solutions for a variety of use cases. Though programming experience is not required, basic experience with Linux is presumed. Exposure to big data concepts and applications is helpful.

Course Details

Course Outline

- Introduction to Cloudera Flow Management

- Processors
 - Connections
 - Dataflows
 - Process Groups
 - FlowFile Provenance
 - Dataflow Templates
 - Apache NiFi Registry
 - FlowFile Attributes
 - NiFi Expression Language
 - Dataflow Optimization
 - NiFi Architecture
 - Site-to-Site Dataflows
 - Cloudera Edge Management and MiNiFi
 - Monitoring and Reporting
 - Controller Services
 - Integrating NiFi with the Cloudera Ecosystem
 - NiFi Security
-