

DOPS-242

Ingesting with Cloudera Data Flow and Apache NiFi



Course Overview

Course Type

Instructor-Led Training

Level

Intermediate

Duration

4 days

Platform

Cloudera on cloud

Topics Covered

- Apache NiFi
- Flow Definitions
- NiFi Registry
- NiFi Expression Language
- ETL Operations
- NiFi Security
- Cloudera Data Flow Service

About This Training

One of the most critical functions of a data-driven enterprise is the ability to manage ingest and data flow across complex ecosystems. Does your team have the tools and skill sets to succeed?

This four-day course provides the fundamental concepts and experience necessary to automate the ingress, flow, transformation, and egress of data using NiFi. The course also covers tuning, troubleshooting, and monitoring the Data Flow process as well as how to integrate a data flow within the Cloudera hybrid ecosystem and external systems.

What Skills You Will Gain

During this course, you learn how to:

- Define, configure, organize, and manage data flows
- Transform and trace data as it flows to its destination
- Track changes to data flows with NiFi Registry
- Use the NiFi Expression Language to control data flows
- Optimize data flows for better performance and maintainability
- Connect data flows with other systems, such as Apache Kafka, Apache Hive, and HDFS
- Utilize the Cloudera Data Flow Service

Who Should Take This Course?

This course is designed for developers, data engineers, administrators, and others with an interest in learning NiFi's innovative no-code, graphical approach to data ingest. Although programming experience is not required, basic experience with Linux is presumed, and previous exposure to big data concepts and applications is helpful.

DOPS-242

Ingesting with Cloudera Data Flow and Apache NiFi

Introduction to Cloudera Flow Management

- Overview of Cloudera Data-in-Motion
- The NiFi User Interface
- Data Flow Catalog
- ReadyFlows
- Instructor-Led Demo: NiFi User Interface
- Hands-On Exercise: Build Your First Data Flow

Processors

- Overview of Processors
- Processor Surface Panel
- Processor Configuration
- Hands-On Exercise: Start Building a Data Flow Using Processors

Connections

- Overview of Connections
- Connection Configuration
- Connector Context Menu
- Hands-On Exercise: Connect Processors in a Data Flow

Data Flows

- Command and Control of a Data Flow
- Processor Relationships
- Back Pressure
- Prioritizers
- Labels
- Hands-On Exercise: Build a More Complex Data Flow
- Hands-On Exercise: Creating a Fork Using Relationships
- Hands-On Exercise: Set Back Pressure Thresholds

Process Groups

- Anatomy of Process Group
- Input and Output Ports
- Hands-On Exercise: Simplify Data Flows Using Process Groups

FlowFile Provenance

- Data Provenance Events
- FlowFile Lineage
- Replaying a FlowFile
- Hands-On Exercise: Using Data Provenance

Parameters

- Parameter Contexts
- Referencing Parameters
- Managing Parameters
- Migrating from Variables
- Hands-On Exercise: Creating, Using, and Managing Parameters

Flow Definitions and Templates

- Flow Definition Overview
- Creating a Flow Definition
- Importing and Deploying a Flow
- Using (migrating from) Templates
- Hands-On Exercise: Creating, Using, and Managing Flow Definitions

Apache NiFi Registry

- Apache NiFi Registry Overview
- Using the Registry
- Hands-On Exercise: Versioning Flows Using NiFi Registry

DOPS-242

Ingesting with Cloudera Data Flow and Apache NiFi

FlowFile Attributes

- FlowFile Attribute Overview
- Routing on Attributes
- Hands-On Exercise: Working with FlowFile Attributes

NiFi Expression Language

- NiFi Expression Language Overview
- Syntax
- Expression Language Editor
- Setting Conditional Values
- Hands-On Exercise: Using the NiFi Expression Language

Controller Services

- Controller Services Overview
- Common Controller Services
- Hands-On Exercise: Adding Apache Hive Controller

Record-based Components

- Record-oriented data
- Record-based Processors
- Avro Schema Registry
- Schema Format

Reading and Writing Record Data

- Querying Record Data
- QueryRecord Processor
- Writing Record Data

Enriching Record Data

- ETL Operations
- Split and Join Processor
- Update Record Processors
- Wait and Notify Processors

NiFi Architecture Overview

- NiFi Architecture Overview
- Public Cloud Architecture
- Private Cloud Architecture

Data Flow Functions

- Overview
- Serverless functions
- Demo: Deploying a Flow Definition as a Function

Data Flow Optimization

- Data Flow Optimization
- Control Rate
- Managing Compute
- Hands-On Exercise: Building an Optimized Data Flow

Monitoring, Reporting, and Troubleshooting

- Monitoring from NiFi
- Reporting
- Examples of Common Reporting Tasks
- Hands-On Exercise: Monitoring and Reporting

NiFi Security

- NiFi Security Overview
- Securing Access to the NiFi UI
- Metadata Management

Integrating NiFi

- NiFi Integration Architecture
- Available ReadyFlows
- A Closer Look at NiFi and Apache Hive