Training: Cloudera
DOPS-242 Ingesting with Cloudera DataFlow

CLOUDERA Training Partner

TRAINING GOALS:

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 at this?

Apache NiFi and 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 dataflow process as well as how to integrate a dataflow within the Cloudera CDP Hybrid ecosystem and external systems.

What you'll learn

During this course, you learn how to:

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

What to expect

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.

CONSPECT:

- Introduction to Cloudera Flow Management
  - Overview of Cloudera Data-in-Motion
  - The NiFi User Interface
  - DataFlow Catalog
  - ReadyFlows

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

- Instructor-Led Demo: NiFi User Interface
- Hands-On Exercise: Build Your First Dataflow

- Processors
  - Overview of Processors
  - Processor Surface Panel
  - Processor Configuration
  - Hands-On Exercise: Start Building a Dataflow Using Processors
- Connections
  - Overview of Connections
  - Connection Configuration
  - Connector Context Menu
  - Hands-On Exercise: Connect Processors in a Dataflow
- Dataflows
  - Command and Control of a Dataflow
  - Processor Relationships
  - Back Pressure
  - Prioritizers
  - Labels
  - Hands-On Exercise: Build a More Complex Dataflow
  - 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 Dataflows 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

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

- 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
- 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
  - Hands-On Exercise: TBD (Creating a function to read and write data?)
- Enriching Record Data
  - ETL Operations

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

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

- NiFi Architecture Overview
  - NiFi Architecture Overview
  - Public Cloud Architecture
  - Private Cloud Architecture
- DataFlow Functions
  - Overview
  - Serverless functions
  - Demo: Deploying a Flow Definition as a Function
- Dataflow Optimization
  - Dataflow Optimization
  - Control Rate
  - Managing Compute
  - Hands-On Exercise: Building an Optimized Dataflow
- 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

REQUIREMENTS:

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.

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

Difficulty level



CERTIFICATE:

The participants will obtain certificates signed by Cloudera (course completion).

Upon completion of the course, attendees are encouraged to continue their study and register for the CDP Data Analyst exam https://www.cloudera.com/about/training/certification/cdp-dataanalyst-exam-cdp-4001.html and/or CDP Data Engineer exam https://www.cloudera.com/about/training/certification/cdp-data-engineer-exam-guide-cdp-3002.html

Certification is a great differentiator. It helps establish you as a leader in the field, providing employers and customers with tangible evidence of your skills and expertise.

TRAINER:

Certified Cloudera Instructor

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007