Hortonworks HDP Operations Hortonworks Data Flow

Course Summary

Description

This course is designed for Data Stewards or Data Flow Managers who are looking forward to automate the flow of data between systems. Topics Include introduction to NiFi, installing and configuring NiFi, detail explanation of NiFi user interface, explanation of its components and elements associated with each. How to build a dataflow, NiFi expression language, understanding NiFi clustering, data provenance, security around NiFi, monitoring tools, and HDF best practices is also covered.

Objectives

By the end of this course, students will be able to:

- Describe HDF, Apache NiFi and its use cases.
- Describe NiFi Architecture
- Understand Nifi Features and Characteristics.
- Understand System requirements to run Nifi.
- Understand Installing and Configuring NiFi
- Understand NiFi user interface in depth.
- Understand how to build a DataFlow using NiFi
- Understand Processor and its Elements
- Understand Connection and its Elements

- Understand Processor Group and its elements
- Understand Remote Processor Group and its Elements
- Learn how to optimize a DataFlow
- Learn how to use NiFi Expression language and its use.
- Learn about Attributes and Templates in NiFi
- Understand Concepts of NiFi Cluster
- Explain Data Provenance in NiFi
- Learn how to Secure NiFi
- Learn How to effectively Monitor NiFi
- Learn about HDF Best Practices

Topics

- Manual Installation of NiFi
- Building a WorkFLow
- Working with Processor Groups
- Working with Remote Processor Groups
- Using NiFi Expression Language.
- Understanding and using Templates.
- Installing and Configuring NiFi Cluster

- Securing NiFi
- Monitoring NiFi
- End Of the Course Project.
- Getting Familiar to NiFi User Interface
- Anatomy of a Processor
- Anatomy of a Connection
- Data Provenance

Audience

This course is designed for data engineers, integration engineers, and architects who are looking to automate Data flow between systems.

Prerequisites

Before taking this course, students should be familiar with programming principles and have previous experience in software development. Experience with Linux and a basic understanding of DataFlow tools would be helpful. No prior Hadoop experience required, but is very helpful.

Duration

Three days