TRAINING OFFERING

STREAM APPLICATIONS WITH APACHE NIFI, KAFKA, STORM & SAM

SUBJECT MATTER EXPERT

This course is designed for developers who need to create real-time applications to ingest and process streaming data sources using Hortonworks Data Flow (HDF) environments. Specific technologies covered includes: Apache NiFi, Apache Kafka, Apache Storm and Hortonworks Schema Registry and Streaming Analytics Manager.

PREREQUISITES

Students should be familiar with programming principles and have experience in software development. **Java programming experience is required.** SQL and light scripting knowledge is also helpful. No prior Hadoop knowledge is required.

TARGET AUDIENCE

Developers and data engineers who need to understand and develop real-time / streaming applications on Hortonworks Data Flow (HDF).

FORMAT

50% Lecture/Discussion
50% Hands-0n Labs

COURSE OBJECTIVES

- Define Enterprise Data Flow
- Describe the Purpose and Function of HDF
- List HDF 3.0's NiFi Architecture and Features
- List the Three Key Concepts of Apache NiFi
- Describe the Components of the NiFi User Interface
- Describe the Purpose and Function of Draggable Components
- Describe the Purpose of the Summary and History
- Define the Anatomy of Processors and Connections
- Build a NiFi Data Flow
- Describe How to Start and Stop a Component

- Describe How to Start and Stop a Component
- Describe the Anatomy of a Processor Group
- Describe the Anatomy of a Remote Processor Group
- Describe Remote Processor Group Transmission
- Define NiFi Site-to-Site Communication
- Recognize Use Cases for Apache Kafka
- Explain the Concept of a Topic Leader and Followers
- Describe the Publication and Consumption of Kafka Messages
- Define Apache Storm Terms Tuple, Stream, Topology, Spout, Bolt, Nimbus and Supervisor
- Diagram the Relationship Between a Supervisor, Worker Process, Executor and a Task
- Diagram how Storm Components Interact to Provide Scalable, Distributed and Parallel Computation of Real-time Data
- Given the Java Code for a Topology, Diagram the Spout and Bolt Connections
- Define the Purpose, and Types, of a Stream Grouping
- List the Differences Between Storm Local Mode and Distributed Mode
- Describe How to Integrate Apache Storm with Apache Kafka
- List Tools Used to Manage Apache Storm
- Display Online Help Using the Storm Command-line Client
- Idenfity How to Open the Storm UI Console
- Interpret the Metrics Displayed in the Apache Storm UI Console
- Idenfity the Differences Between Reliable and Unrealiable Operation
- Diagram a Tuple Tree and Identify its Branches
- List the Two Requirements for Reliable Operation
- Given a Diagram, Describe the Operation of an Acker Task
- Describe the Responses to Various Apache Storm Component Failures
- List Three Methods to Disable Reliable Operation
- Introduce Hortonworks Schema Registry (SR) and Streaming Analytics Manager (SAM)

HANDS-ON LABS

- Demonstration: NiFi User Interface
- Building a NiFI DataFlow
- Working with Processing Groups
- Working with Remote Processor Groups
- Creating and Managing Apache Kafka Topics
- Integrating HDF with HDP
- Creating a Word Count Topology
- Submitting Topologies in Local Mode
- Performing a Kafka Word Count
- Demonstration: SAM in Action