Big Data Streaming for Developers

Course Overview

This course is applicable to software versions 10.2.1 and forward. Gain the skills necessary to execute end-to-end big data streaming use cases. Learn to prepare, process, enrich, and maintain streams of data in real time using Informatica, Edge, Kafka, and Spark.

Objectives

After successfully completing this course, students should be able to:

- Discuss streaming
- Describe Kappa architecture
- List the types of streaming data
- Create an EDS Service
- Create, deploy, and monitor a data flow
- List the BDS key features
- Describe the BDS component architecture
- Describe Kafka data objects
- Create Kafka connections
- Discuss and list sources, and targets in a streaming mapping
- Discuss lookup sources
- Execute a streaming mapping
- Monitor logs and troubleshoot streaming mappings

Target Audience

- Developer

Prerequisites

- None

Agenda

Module 1: Streaming Overview

- Key differences between batch and streaming
- Streaming Data Management use cases
- Streaming architecture
- Kappa architecture
- End-to-end Streaming Data Management
- Types of streaming data
- Benefits of streaming

Module 2: Edge Data Streaming (EDS) Overview

- EDS architecture
- EDS key features
- EDS Data flow process
- EDS UI
- Create an EDS Service
- Create a data flow
- Deploy a data flow
- Monitor the data flow
- Lab1: Create Edge Data Streaming Service
- Lab2: Create and Deploy a Data Flow

Module 3: Big Data Streaming Overview

- Big Data Streaming overview
- Stream Data Processing with Spark streaming
- BDS component architecture
- BDS key features

Module 4: Kafka Overview

- Kafka Concepts
- Kafka core APIs
- Topics in Kafka

- Kafka models
- Kafka Use cases

Lab1: Create a Kafka connection

Module 5: Streaming Mappings

- Sources in a streaming mapping
- Targets in streaming Mapping
- Lookup sources
- Kafka Data Object Properties
- Lab1: Create a Mapping with Kafka Source and HDFS Target
- Lab 2: Create a Mapping with Kafka Source and Kafka Target
- BDS Transformations
- Lab3: Enhance Mapping Using Filter and Expression Transformations
- Lab4: Enhance Mapping Using Window and Aggregator Transformations
- Lab5: Enhance Mapping Using Sorter and Rank Transformations

Module 6: Monitoring Logs and Troubleshooting

- Spark Monitoring
- Viewing Logs
- Troubleshooting
- Lab1: Monitor an EDS Data Flow
- Lab2: Monitor a BDS Mapping

Module 7: Performance Tuning and Best Practices

- Tune performance of Spark jobs
- List some best practices while working with streaming data

Module 8: End-to-End Use Case

- Use Case
- EDS and BDS – Final Goal
- Lab 1: Convert Unstructured Streaming Data into Structured Data
- Lab 2: Ingest Data from EDS to BDS and Execute a Mapping in BDS