Kafka Operations + Security + Monitoring + Kafka Connect + ksqlDB + Kafka Streams

Prerequisites: Basic knowledge of java

Duration: 15 Days (8 Hrs./Day)

Course Objective: This comprehensive course covers the fundamentals and advanced topics in Big Data, Apache Kafka, and related technologies. Participants will begin with an introduction to Big Data and Kafka, understanding its architecture, core concepts, and real-world applications. The course progresses into practical labs, including Kafka cluster setup, Java-based Kafka development, performance tuning, and Kafka Connect. Advanced modules explore Kafka security, monitoring, ksqlDB, Kafka Streams, and achieving exactly-once semantics. By the end of the course, students will have hands-on experience in developing, managing, and securing Kafka ecosystems, as well as leveraging Kafka Streams and ksqlDB for real-time data processing.

Apache Kafka version: Latest.

Lab Requirement: Koenig DC/Linux (CentOS 9) (customizable).

Module 1 - Introduction to Big Data and Apache Kafka

Introduction to Big Data

Big Data Customer Scenarios

What is Kafka?

Need for Kafka

Core Concepts of Kafka

Kafka Architecture

Where is Kafka Used?

Module 2 - About Kafka Cluster

Understanding the Components of Kafka Cluster

Producer of Kafka

Consumer of Kafka

Lab: Installation of Kafka Cluster

Lab: Configuring Kafka Cluster

Lab: Producer and Consumer in Action

Module 3 - Kafka with Java

Understanding the Components of Kafka Cluster

Kafka SDK List

Lab: Creating Kafka Project with Java

Lab: Java Producer

Lab: Java Producer Callbacks

Lab: Java Producer with Keys

Lab: Java Consumer

Lab: Java Consumer with Graceful Shutdown

Lab: Java Consumer in Consumer Group

Lab: Java Consumer with Incremental Cooperative Rebalance and Static Group Membership

Lab: Java Consumer Auto Offset Commit Behavior

Module 4 - Kafka Operations and Performance Tuning

Offset

Design

Hardware

Kafka Monitoring and Issues

Kafka Performance Tuning

Module 5 - Kafka Wikimedia Producer and Advanced Producer Configurations

Introduction to Kafka Producer

Lab: Wikimedia Producer Project Setup

Lab: Wikimedia Producer Implementation with Java

Lab: Wikimedia Producer Run with Java

Lab: Producer Config Intros

Lab: Producer Acknowledgements

Lab: Producer Retries

Lab: Idempotent Producer

Lab: Safe Kafka Producer Settings

Lab: Wikimedia Producer Safe Producer Implementation

Lab: Kafka Message Compression

Module 6 - OpenSearch Consumer and Advanced Consumer Configurations

Introduction to OpenSearch

Lab: OpenSearch Consumer with Java

Lab: Setting Up OpenSearch on Docker

Lab: OpenSearch Consumer Implementation with Java

Lab: Consumer Delivery Semantics

Lab: OpenSearch Consumer Implementation Idempotence

Lab: Consumer Offsets Commit Strategies

Lab: OpenSearch Consumer Implementation Delivery Semantics

Lab: OpenSearch Consumer Implementations Batching Data

Lab: Consumer Offset Reset Behavior

Lab: OpenSearch Consumer Implementation Replaying Data

Module 7 - Kafka Monitoring Setup with Prometheus

Overview of Monitoring

Lab: Setting up Prometheus + Kafka Broker 1

Lab: Broker 2 & 3 Prometheus Setup

Lab: KRaft & Prometheus Setup

Module 8 - Kafka Monitoring Setup with Grafana

Overview of Monitoring

Lab: Setup Grafana on Administration Server

Lab: Setup Kafka Dashboard on Grafana

Lab: Adding Graphs to Grafana

Lab: Adding more graphs to Grafana

Lab: Observing Grafana when we kill a broker

Module 9 - Introduction to Kafka Security

Kafka Authentication Basics

Kafka Authentication with SSL and SASL_SSL

Authorization

Encryption

Securing Kraft

Audit Logs

Security Recommendations

Module 10 – Kafka security hands on

Lab: Creating a Certificate Authority (CA)

Lab: SSL Setup in Kafka

Lab: SSL Setup for Clients

Lab: SSL Authentication

Lab: Hands-On Kerberos - Part 1: Setup Azure Virtual Machines

Lab: Hands-On Kerberos - Part 2: Principals & Keytabs

Lab: Hands-On Kerberos - Part 3: Kafka Configuration

Lab: Hands-On Kerberos - Part 4: Client Configuration

Lab: JAAS file / config

Lab: Hands-On ACL demo

Module 11 - Introduction to Kafka Connect

What is Kafka Connect?

Kafka Connect Architecture Design

Connectors, Configuration, Tasks, Workers

Standalone vs Distributed Mode

Distributed Architecture in Details

Errors and Dead Letter Queues

Lab: Starting Kafka Connect Cluster using Docker

Module 12 - Kafka Connect Source

Kafka Connect Source Architecture Design

Lab: FileStream Source Connector - Standalone Mode

Lab: FileStream Source Connector - Distributed Mode

Lab: List of Available Connectors

Module 13 - Kafka Connect Sink

Kafka Connect Sink Architecture Design

Lab: ElasticSearch Sink Connector - Distributed Mode

Lab: Kafka Connect REST API

Lab: JDBC Sink Connector - Distributed Mode

Module 14 - Kafka Connect Connector programming

Lab: Core Concepts and APIs

Lab: Developing a Simple Connector

Lab: Dynamic Input/Output Streams

Lab: Configuration Validation

Lab: Working with Schemas

Module 15 - ksqlDB and KSQL Basics

Lab: KSQL Setup

Lab: Our first KSQL Stream

Lab: Create a Stream with JSON

Lab: KSQL Datagen - Generating Streams

Lab: Manipulate a Stream

Lab: Streams from streams and functions

Lab: ksqlDB Tables

Module 16 - ksqlDB and KSQL Intermediate

Lab: KSQL Joins

Lab: Pull Queries

Lab: Kafka Connect with ksqlDB

Lab: Data Encodings

Lab: CSV Delimited Data

Lab: JSON Data

Lab: Avro Data

Lab: Avro Schema Evolution

Lab: Nested JSON

Lab: Build a rekeyed table

Lab: Repartition a Stream

Lab: Merging Streams

Lab: Windowing

Lab: Geospatial

Lab: Extending KSOL- UDF / UDAF

Lab: using the UDF / UDAF

Module 17 - ksqlDB and KSQL in production

Lab: Moving to Productions-Headless for KSOL

Lab: Scaling and Load Balancing

Lab: Configuration Settings

Lab: State Stores

Lab: Testing ksqlDB applications

Module 18 - Introduction to Kafka Streams

What is Kafka Streams?

Use Case Examples

Lab: Running your first Kafka Streams Application: WordCount

Kafka Streams vs other stream processing libraries (Spark Streaming, NiFI, and Flink)

Module 19 - Kafka Streams core concepts

Kafka Streams Basic Operations

KTable

Serialization

Joins

Stateful Operations

Windowing

Time Concepts

Processor API

Testing

Error Handling

Internals

Stateful Fault Tolerance

Interactive Queries

Module 20 - Build your first Apache Kafka Streams application (WordCount) from scratch

Lab: Project setup

Lab: Kafka Streams Application Properties

Lab: Java 8 Lambda Functions - quick overview

Lab: Word Count Application Topology

Lab: Printing the Kafka Streams Topology

Lab: Kafka Streams Graceful Shutdown

Lab: Running Application from IntelliJ IDEA

Lab: Debugging Application from IntelliJ IDEA

Lab: Internal Topics for our Kafka Streams Application

Lab: Packaging the application as Fat Jar & Running the Fat Jar

Lab: Scaling our application

Module 21 - KStreams and KTables Simple Operations (Stateless)

KStream & KTables

Stateless vs Stateful Operations

MapValues / Map

Filter / FilterNot

FlatMapValues / FlatMap

Branch

SelectKey

Reading from Kafka

Writing to Kafka

Streams Marked for Re-Partition

Refresher on Log Compaction

KStream and KTables Duality

Transforming a KTable to a KStream

Transforming a KStream to a KTable

Module 22 - KStreams and KTables Advanced Operations (Stateful)

K Table groupBy

KGroupedStream / KGroupedTable Count

KGroupedStream / KGroupedTable

KGroupedStream / KGroupedTable Reduce

KStream peek

KStream Transform / TransformValues

What if I want to write to an external System?

Module 23 - Exactly Once Semantics

What's Exactly Once?

Exactly Once in Kafka 0.11

What's the problem with at least once anyway?

How to do exactly once in Kafka Streams

Lab: Achieving Exactly once semantics using kafka producer

Lab: Achieving Exactly once semantics using kafka streams

Module 24 - Joins - KStream to GlobalKTable

What are joins in Kafka Streams?

Join Constraints and GlobalKTables

The different types of joins: Inner Join, Left Join & Outer Join

Lab: Creating a join with UserEnrich Kafka Streams App