Certified Kafka Course – Curriculum

The Certified Kafka course by DataFlair is a perfect blend of in-depth theoretical knowledge and strong practical skills via implementation of real life projects to give you a headstart and enable you to bag top Big Data jobs in the industry.

Course duration: 24 Hours

Module 1: Demystifying Apache Kafka

This module introduces you to the rudiments of Big Data, what makes it so important, some problems and solutions, and why we need Kafka.

- Introduction to Big data
- Importance for messaging queue
- Need for distributed messaging queue
- Conventional solutions and associated problems
- Why we need Apache Kafka

Module 2: Kafka Introduction

We then introduce you to the basics of Kafka, its features and architecture, and also take some real-life case studies.

- What is Apache Kafka?
- Kafka Features and terminologies
- High-level kafka architecture
- Real-life Kafka Case Studies

Module 3: A Deep Dive into the Kafka Architecture

This module deals with the internals of architecture and some core concepts. It also talks about different components of Kafka. You will also learn about its different versions.

- Internals of architecture and core concepts
- Kafka components - Broker, Producer, Consumer, Topics, Partitions
- Different versions of Kafka

Module 4: Understanding Brokers

Next, we discuss brokers in Kafka and how to deploy them. We'll also see how to deploy multiple brokers on a single machine.

- Working of a broker
- Broker Deployment
- Multiple brokers on a single machine
- Decommisioning brokers

Module 5: Understanding Producers

After finishing this module, you will be comfortable with producers. We will discuss their architecture, partitions, types, and configurations.

- Basics of producers
- The producer architecture
- Producer partition- Custom, Round Robin, Field-Based Partition
- Producer Java API
- Types of producers - sync, async
- Different producer configurations
- Sync and async producer hands-on

Module 6: Understanding Consumers

This module teaches you about consumers, queuing, groups, broadcast, and the consumer Java API.

- Basics of consumers
- Consumer queuing, consumer groups, and consumer broadcast
- The Consumer Java API
- Producer and Consumer Hands-on

Module 7: Mirroring Kafka

We now begin talking mirroring. We talk of how it works, the mirror maker, and producers and consumers for mirroring.

- What is mirroring?
- How mirroring works
- Mirror Maker and its role
- Kafka Mirroring Hands-on
- Mirror Maker producers and consumers

Module 8: Topics and Partitions

Halfway through the course now, we begin to explore other concepts like topics and partitions. This module teaches you performance optimization, using and distributing partitions, and partition reassignment.

- Working with topics and updating them
- Performance optimization and configurations
- Using Partitions and distribution of partitions
- Partition reassignment - Automatic migrating
- Partition reassignment - Manual migrating

Module 9: Replication in Kafka

We move on to replication. In this module, we talk of replication and its various types.

- High availability and reliability using replication
- ISR - In Sync Replication
- Primary Replication and Quorum Replication
- Partition reassignment - Increasing Replication
- Topic, Partition and Replication Hands-on

Module 10: Introduction to Zookeeper

With Kafka by your side, we now talk about Zookeeper. This module discusses its architecture, applications, installation, and troubleshooting.

- What is Zookeeper?
- Leader Election by Zookeeper
- Applications using Zookeeper
- Zookeeper Architecture
- Zookeeper configuration and installation
- Zookeeper troubleshooting

Module 11: Installation of Kafka

Now, we learn to install Kafka. We observe its setup, configuration, deployment, and scaling.

- Setup environment and pre-requisites
- Configuration of different components
- Deploying Apache Kafka on single node cluster
- Deploying Apache Kafka on multi-node cluster
- Balancing Leadership
- Scaling Kafka cluster and troubleshooting

Module 12: Kafka Advanced Operations – Part 1

This module focuses on some advanced operations like buffering data and consumer grouping. We also consider some best practices.

- Working with multiple topics for different type of data

- Developing applications using high and low level APIs
- Buffering data in Apache Kafka
- Best practices of data distribution
- Different types of consumers high level & low level
- Multiple consumers grouping

Module 13: Kafka Advanced Operations – Part 2

Now, we take a brief introduction to multiple partitions and monitoring. Then, we consider some more best-practices.

- Multiple partitions and multi-threaded consumers
- Best practices for consumers
- Configuring multiple producers & consumers together
- Log Segment and Data Retention
- Monitoring and management of Kafka Cluster
- JMX Monitor and Kafka administration tools

Module 14: Real-life Kafka Project

We conclude this course with a real-life, live Kafka project to prepare you for the industry. Here, we make use of various Kafka components to solve real- world problems in Big Data Analytics.