

Apache Kafka using Python

Duration: 4 days (8hrs/day)

Prerequisites: Basic knowledge of Linux and Python.

Course Objective: This Apache Kafka course introduces the fundamentals of Apache Kafka and its core concepts such as topics, partitions, replication, and message queues. It also covers Kafka's architecture, configuration, integration, and security. Additionally, the course explores the Kafka Producer API, Consumer API, Streams API, and Connect API. Upon completion of this course, you will have a solid understanding of Kafka and its capabilities and will be able to use it to stream data in real time.

Apache Kafka Version: Latest

Lab Requirement: Koenig DC

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

Module 3 - Kafka Admin API

Admin API Introduction

Lab: CLI to Manage Topics

Lab: CLI to Manage Topics Continued (Advance Configurations)

Lab: Python to Create Topics

Lab: Python Create and Alter Topics with Advanced Configurations

Module 4 - Kafka Producer API

Producer API Introduction

Producers at a High Level

Producers and their Influence on Message Partition Assignment

Lab: CLI Tools for Producing Messages to Kafka

Lab: Basic Producer in Python

Detailed Overview of Kafka Producer

Changing Partitions Change Ordering

Lab: Advanced Producer in Python

Module 5 - Kafka Consumer API

Consumer API Introduction

Consumer Group Offsets and Progress Tracking

Consumer Group Rebalances

Lab: Basic Consumer in Python

Auto Offset Commits and At Least Once Processing

Manual Offset Commits and At Least Once Processing

Manual Offset Commits and At Most Once Processing

Manual Offset Commits and Exactly Once Processing

Lab: Advanced Consumer in Python

Module 6 – Schema Registry

Schema Registry Introduction

What is Confluent Schema Registry

Why use Confluent Schema Registry

How Schema Registry Fits into Kafka Architecture

Quick Overview of Apache Avro

Schema Registry Compatibility Settings and Schema Evolution Checks

Lab: Integrating Avro and Schema Registry in a Producer

Lab: Integrating Avro and Schema Registry in a Consumer

Module 7 – Stream Processing with Faust in Python

What is Faust

Key Data Constructs of Faust Library

Types of Streaming Computations

Faust Channels, Topics, Streams and Agents

Lab: Install and Setup Faust & Agents and Topics

Lab: Faust Tasks and Timers

Lab: Simple Faust Producer Consumer

Lab: Producing and Consuming Complex Types

Lab: Calculating Aggregates using Faust Tables