Cognixia

Introduction to Apache Kafka

join@cognixia.us

www.cognixia.com/us

COURSE OVERVIEW

Overview

Apache Kafka has emerged as a popular messaging component for streaming platforms. This course will introduce Kafka to students; we will learn about Kafka features and learn how to use it for streaming data.

What You Will Learn

- Streaming technologies and architecture
- Kafka concepts and architecture
- Programming using Kafka API
- Kafka Streams API
- Kafka Connect
- Confluent Platform
- Kafka Administration
- Monitoring Kafka
- Tuning / Troubleshooting Kafka
- Best practices
- Use cases
- Designing real world solutions using Kafka

Skill Level - Introductory – Intermediate

DURATION

4 Days

TARGET AUDIENCE

Developers, Architects

PRE-REQUISITES

- Recommended: Comfortable with Java programming language and Java development tools (Eclipse, Maven) - programming exercises are in Java
- Nice to have: Comfortable in Linux environment (be able to navigate Linux command line, run commands)

COURSE CONTENT

Introduction to Streaming Systems

- Understanding fast data
- Streaming terminologies
- Understanding at-least-once / at-most-once / exactly-once processing patterns
- Popular streaming architectures
- Lambda architecture
- Streaming platforms overview

Introducing Kafka

- Comparing Kafka with other queue systems (JMS / MQ)
- Kafka Architecture
- Kaka concepts: Messages, Topics, Partitions, Brokers, Producers, commit logs
- Kafka & Zookeeper
- Producing messages
- Consuming messages
- Consumers, Consumer Groups
- Message retention
- Scaling Kafka
- Kafka across multiple data centers and disaster recovery
- Labs:

– Getting Kafka up and running
– Using Kafka utilities

Using Kafka APIs

- Configuration parameters
- Producer API - sending messages to Kafka
- Consumer API - consuming messages from Kafka
- Producer send modes
- Message compression
- Commits , Offsets, Seeking
- Managing offsets - auto commit / manual commit
- Labs:

– Writing Producer / Consumer
– Benchmarking Producer send modes
– Comparing compression schemes
– Managing offsets
– Clickstream processing

Kafka Streams API

- Introduction to Kafka Streams library
- Features and design
- Streams concepts: KStream / KTable / KStore
- Streaming operations (transformations, filters, joins, aggregations)
- Using Streams API: foreach / filter / map / groupby
- Labs:

– Kafka Streaming APIs

Kafka Connect

- Connect ecosystem
- Popular connectors
- Sample configurations

Kafka Administration

- Hardware selection
- Software prerequisites
- Kafka specific settings for OS, Disks, memory, CPU and Network
- Installing and setting up Kafka
- Verifying installation
- Securing Kafka
- Capacity planning
- Best practices

Monitoring and Instrumenting Kafka

- Monitoring Kafka metrics
- Introduction to Metrics library

- Instrumenting Kafka applications with the Metrics library
- Using Grafana to visualize metrics
- Labs

– Monitor Kafka cluster
– Instrument Kafka applications with the metrics library

Kafka Best Practices

- Avoiding common mistakes
- Hardware selection
- Cluster sizing
- Partition sizing
- Zookeeper settings
- Compression and batching
- Message sizing
- Monitoring and instrumenting
- Troubleshooting

Kafka Case Studies

- This section will feature case studies from various companies using Kafka solve real world problems

Kafka Design Exercises

- In section, students will work as a group and design end-to-end scenarios with Kafka

Final workshop (time permitting)

- Students will build an end-to-end application using Kafka