

Curso de Administración y Operación de Kafka

Duración: 5 días - 25 horas

Descripción del curso:

El curso trata de proporcionar los conceptos y habilidades necesarias para que los alumnos puedan administrar un cluster de brokers de Kafka, conozcan las características de la gestión de topics, como aplicar seguridad y fiabilidad a un sistema de mensajería como Kafka. El alumno conocerá como realizar tareas como lanzar crear y mantener topics, lanzar y consumir mensajes, o crear grupos de consumidores.

Dirigido a:

Principalmente a desarrolladores, sobre todo para aquellos que tengan conocimientos y experiencia sobre el lenguaje de programación a usar: Scala o Python.

Contenido:

1. Introducción

- Conceptos básicos de Kafka
- Kafka vs otros sistemas de mensajería y motivación
 - o Complejidad de los sistemas
 - o Procesamiento en tiempo real
 - o Plataforma de streaming de datos

2. Diseño y arquitectura de Kafka

- Como almacena Kafka los datos en disco
 - o Logs, índices y checkpoints
 - o Gestión de Offset
- Componentes de Kafka
 - o Producers
 - o Brokers
 - Roles de los brokers: leaders, followers
 - o Consumers
 - o Uso de Zookeeper por parte de Kafka
- Protocolo de comunicaciones de Kafka
- Integración con Kafka connect
 - o Motivación
 - o Tipos de conectores
 - o Implementación de Kafka connect
- Multi-cluster Kafka
 - o Mirror maker
 - o Kafka connect
- Diferencias entre las principales versiones de Kafka (0.9, 1.0, 1.1)
 - Migraciones e incompatibilidades
- AVRO, Schema Management en Kafka

3. Diseño de detalle de Kafka

- Ciclo de vida de las peticiones de los productores
- Ciclo de vida de las peticiones de los consumidores
- Kafka controller

4. Diseño de soluciones

- Arquitectura de referencia de Kafka
- Dimensionamiento
- Diseñando para alta disponibilidad (HA)
- Kafka en múltiples datacenters:
 - o Distribución de brokers, réplicas
 - o Disaster recovery
 - o Replicación intra-cluster
- Consideraciones de despliegue

5. Seguridad

- Securizando Kafka
 - o SSL
- Securizando Zookeeper
- Autenticación, autorización
 - o Kerberos, Active Directory
 - o ACLs y Cuotas
 - o SASL

6. Operación

- Herramientas en línea de commando y operativa de administración
- Administración de Logs, retención y compactación
- Recuperación ante fallo de un bróker
- Añadir nuevos brokers a un cluster
- Realojamiento de un tópico/replica
- Configuración de Brokers
- Configuración hardware recomendada

7. Rendimiento

- Tuning de rendimiento
 - o Mejora de rendimiento de los brokers
 - o Mejora del tiempo de recuperación ante caída de un bróker
 - o Mejora de rendimiento de los productores
- Balanceo de carga
- Pruebas de rendimiento

8. Monitorización

- Monitorización y alertas
- Visualización e interpretación de las métricas del cluster

9. Identificación de errores

- Análisis de logs de error
- Análisis de logs de peticiones
- Análisis de Logs de controllers
- Troubleshooting y best practices
- Diagnóstico de problemas de latencia en el cluster

10. Distribuciones habituales

- Apache Kafka
- Confluent
- Hortonworks
- Cloudera
- Tibco
- Diferencias entre distribuciones