

# CLOUDERA: Introducing Apache Kafka

Código: DENG-157

**Propuesta de Valor:** OTROS CURSOS DE CAPACITACIÓN TECNOLÓGICA

**Duración:** 32 Horas



Este curso comienza presentando Apache Kafka, explicando sus conceptos y arquitectura clave, y discutiendo varios casos de uso comunes. Sobre esta base, aprenderá a planificar una implementación de Kafka y luego obtendrá experiencia práctica al instalar y configurar su propio clúster de varios nodos basado en la nube que ejecuta Kafka en Cloudera Data Platform (CDP).

Luego utilizará este grupo durante más de 20 ejercicios prácticos que siguen, cubriendo una gama de habilidades esenciales, comenzando con cómo crear temas, productores y consumidores de Kafka, y luego continuando con aspectos progresivamente más desafiantes de las operaciones y el desarrollo de Kafka, como los relacionados con problemas de escalabilidad, confiabilidad y rendimiento. A lo largo del curso, aprenderá y utilizará las herramientas recomendadas por Cloudera para trabajar con Kafka, incluido Cloudera Manager, Schema Registry, Streams Messaging Manager y Cruise Control.



## AUDIENCIA

- Este curso está diseñado para administradores de sistemas, ingenieros de datos y desarrolladores.



## PRE REQUISITOS

- Se espera que todos los estudiantes tengan experiencia básica en Linux y se recomienda un dominio básico del lenguaje de programación Java.
- No se necesita experiencia previa con Apache Kafka.



## OBJETIVOS

- Planifique, implemente y opere clústeres de Kafka.
- Crear y administrar temas.
- Desarrollar productores y consumidores.
- Utilice la replicación para mejorar la tolerancia a fallas.
- Utilice el particionamiento para mejorar la escalabilidad.
- Solucionar problemas comunes y problemas de rendimiento.



## CERTIFICACIÓN DISPONIBLE

- Certificado oficial de **COGNOS**.



## CONTENIDO

### 1. DESCRIPCIÓN GENERAL DE KAFKA

- 1.1. ARQUITECTURA DE ALTO NIVEL
- 1.2. CASOS DE USO COMUNES
- 1.3. DISTRIBUCIÓN DE CLOUDERA DE APACHE KAFKA

### 2. IMPLEMENTACIÓN DE APACHE KAFKA

- 2.1. REQUISITOS Y DEPENDENCIAS DEL SISTEMA
- 2.2. FUNCIONES DE SERVICIO
- 2.3. PLANIFICACIÓN DE LA IMPLEMENTACIÓN IMPLEMENTACIÓN DE LOS SERVICIOS DE KAFKA
- 2.4. EJERCICIO: PREPARACIÓN DEL ENTORNO DEL EJERCICIO
- 2.5. EJERCICIO: INSTALACIÓN DEL SERVICIO KAFKA CON CLOUDERA MANAGER
- 2.6. EJERCICIO (OPCIONAL): CREAR PANELES DE MÉTRICAS
- 2.7. EJERCICIO (OPCIONAL): USO DE LA API DE CM

### 3. CONCEPTOS BÁSICOS DE LA LÍNEA DE COMANDOS DE KAFKA

- 3.1. CREAR Y ADMINISTRAR TEMAS
- 3.2. EJECUTANDO PRODUCTORES Y CONSUMIDORES

### 4. USO DE STREAMS MESSAGING MANAGER (SMM)

- 4.1. DESCRIPCIÓN GENERAL DE STREAMS MESSAGING MANAGER
- 4.2. PRODUCTORES, TEMAS Y CONSUMIDORES
- 4.3. EXPLORADOR DE DATOS
- 4.4. CORREDORES
- 4.5. GESTIÓN DE TEMAS
- 4.6. EJERCICIO: GESTIÓN DE TEMAS MEDIANTE LA CLI
- 4.7. EJERCICIO: CONEXIÓN DE PRODUCTORES Y CONSUMIDORES DESDE LA LÍNEA DE COMANDOS

### 5. CONCEPTOS BÁSICOS DE LA API JAVA DE KAFKA

- 5.1. DESCRIPCIÓN GENERAL DE LAS API DE KAFKA
- 5.2. GESTIÓN DE TEMAS DESDE LA API DE JAVA
- 5.3. EJERCICIO (OPCIONAL): ADMINISTRACIÓN DE TEMAS DE KAFKA MEDIANTE LA API DE JAVA
- 5.4. USO DE PRODUCTORES Y CONSUMIDORES DE LA API DE JAVA
- 5.5. EJERCICIO: DESARROLLO DE PRODUCTORES Y CONSUMIDORES CON LA API DE JAVA

## 6. MEJORA DE LA DISPONIBILIDAD A TRAVÉS DE LA REPLICACIÓN

- 6.1. REPLICACIÓN
- 6.2. EJERCICIO: OBSERVACIÓN DEL TIEMPO DE INACTIVIDAD DEBIDO A LA FALLA DEL CORREDOR
- 6.3. CONSIDERACIONES PARA EL FACTOR DE REPLICACIÓN
- 6.4. EJERCICIO: ADICIÓN DE RÉPLICAS PARA MEJORAR LA DISPONIBILIDAD

## 7. MEJORA DE LA ESCALABILIDAD DE LAS APLICACIONES

- 7.1. FRACCIONAMIENTO
- 7.2. CÓMO SE DIVIDEN LOS MENSAJES
- 7.3. EJERCICIO: OBSERVAR CÓMO LA PARTICIÓN AFECTA EL RENDIMIENTO
- 7.4. GRUPOS DE CONSUMIDORES
- 7.5. EJERCICIO: IMPLEMENTACIÓN DE GRUPOS DE CONSUMIDORES
- 7.6. REEQUILIBRIO DEL CONSUMIDOR
- 7.7. EJERCICIO: USO DE UNA CLAVE PARA CONTROLAR LA ASIGNACIÓN DE PARTICIONES

## 8. MEJORA DE LA CONFIABILIDAD DE LAS APLICACIONES

- 8.1. SEMÁNTICA DE ENTREGA
- 8.2. DEMOSTRACIÓN (OPCIONAL): ISR FRENTE A ACK
- 8.3. ENTREGA AL PRODUCTOR
- 8.4. EJERCICIO: PRODUCTOR IDEMPOTENTE
- 8.5. ACTAS
- 8.6. EJERCICIO: PRODUCTORES Y CONSUMIDORES TRANSACCIONALES
- 8.7. MANEJO DE FALLAS DEL CONSUMIDOR
- 8.8. GESTIÓN DE COMPENSACIONES
- 8.9. EJERCICIO: DETECCIÓN Y SUPRESIÓN DE MENSAJES DUPLICADOS
- 8.10. EJERCICIO: MANEJO DE REGISTROS NO VÁLIDOS
- 8.11. MANEJAR LA FALLA DEL PRODUCTOR

## 9. ANÁLISIS DE CLÚSTERES DE KAFKA CON SMM

- 9.1. LATENCIA DE EXTREMO A EXTREMO
- 9.2. NOTIFICADORES
- 9.3. POLÍTICAS DE ALERTA
- 9.4. CASOS DE USO

## 10. MONITOREO DE KAFKA

- 10.1. DESCRIPCIÓN GENERAL DE LA SUPERVISIÓN
- 10.2. SUPERVISIÓN CON CLOUDERA MANAGER
- 10.3. GRÁFICOS E INFORMES EN CM
- 10.4. RECOMENDACIONES DE MONITOREO
- 10.5. MÉTRICAS PARA LA RESOLUCIÓN DE PROBLEMAS
- 10.6. DIAGNÓSTICO DE FALLAS EN EL SERVICIO
- 10.7. EJERCICIO: MONITOREO DE KAFKA

## 11. GESTIÓN DE KAFKA

- 11.1. ADMINISTRAR EL ALMACENAMIENTO DE TEMAS DE KAFKA
- 11.2. DEMOSTRACIÓN (OPCIONAL): PERÍODO DE RETENCIÓN DE MENSAJES
- 11.3. LIMPIEZA Y RECOPIACIÓN DE REGISTROS
- 11.4. REEQUILIBRIO DE PARTICIONES
- 11.5. CONTROL DE CRUCERO
- 11.6. EJERCICIO: INSTALACIÓN DEL CONTROL DE CRUCERO
- 11.7. EJERCICIO: SOLUCIÓN DE PROBLEMAS DE TEMAS DE KAFKA
- 11.8. ELECCIÓN DE LÍDER INMUNDO
- 11.9. EJERCICIO: ELECCIÓN DE LÍDER INMUNDO
- 11.10. ADICIÓN Y ELIMINACIÓN DE AGENTES
- 11.11. EJERCICIO: ADICIÓN Y ELIMINACIÓN DE AGENTES
- 11.12. MEJORES PRÁCTICAS

## 12. ESTRUCTURA, FORMATO Y CONTROL DE VERSIONES DEL MENSAJE

- 12.1. ESTRUCTURA DEL MENSAJE
- 12.2. REGISTRO DE ESQUEMAS
- 12.3. DEFINICIÓN DE ESQUEMAS
- 12.4. EVOLUCIÓN DEL ESQUEMA Y CONTROL DE VERSIONES
- 12.5. CLIENTE DE REGISTRO DE ESQUEMA
- 12.6. EJERCICIO: USO DE UN ESQUEMA DE AVRO

## 13. MEJORA DEL RENDIMIENTO DE LAS APLICACIONES

- 13.1. TAMAÑO DEL MENSAJE
- 13.2. PROCESAMIENTO POR LOTES
- 13.3. COMPRESIÓN

13.4. EJERCICIO: OBSERVAR CÓMO AFECTA LA COMPRESIÓN AL RENDIMIENTO

## 14. MEJORA DEL RENDIMIENTO DEL SERVICIO DE KAFKA

14.1. ESTRATEGIAS DE AJUSTE DEL RENDIMIENTO PARA EL ADMINISTRADOR

14.2. TAMAÑO DEL CLÚSTER

14.3. EJERCICIO: CAPACIDAD DE PLANIFICACIÓN NECESARIA PARA UN CASO DE USO

## 15. PROTECCIÓN DEL CLÚSTER DE KAFKA

15.1. CIFRADO

15.2. AUTENTICACIÓN

15.3. AUTORIZACIÓN

15.4. REVISIÓN DE CUENTAS



## BENEFICIOS

- Al finalizar el curso, los participantes podrán planificar, implementar y operar clústeres de Kafka. Solucionar problemas comunes y problemas de rendimiento