

Programas Asíncronos

SÍLABO

I. DATOS GENERALES

Curso	:	Introducción a Data Streaming con Apache Kafka
Área	:	Big Data
Docente	:	Guillermo Suxo
Duración	:	3 horas académicas

II. SUMILLA

La calidad de los datos que se emplean para la inteligencia de negocios y la analítica, es directamente proporcional a la calidad de las decisiones a las que conllevan. Tal calidad de datos está sujeta a la evaluación de una serie de cualidades, entre las que se encuentra la 'Oportunidad', referida a la disponibilidad de los datos en el momento correcto. En pro de ello, han surgido dos grupos de tecnologías, por un lado, Hardware como los sensores y otros dispositivos que emiten datos en tiempo real, y por el otro Software en capacidad de leerlos y distribuirlos al mismo ritmo a otros sistemas que lo requieran, como es el caso de Apache Kafka, una solución basada en eventos para trabajo en Streaming.

Por ello, DMC presenta el Curso Asíncrono "Introducción a Data Streaming con Apache Kafka", el cual te permitirá comprender las características generales de Apache Kafka en el procesamiento de datos de Streaming y su implementación en casos específicos de negocio.

III. REQUISITOS

- Ninguno

IV. DIRIGIDO A

- Profesionales y estudiantes de las áreas de ciencias de la computación, informática y afines interesados en comprender los fundamentos de Apache Kafka.

V. OBJETIVOS – LOGROS

El curso le permitirá al participante:

- Comprender las características generales de Apache Kafka y cómo se integra con un pipeline de datos.
- Comprender cómo emplear las librerías KafkaProducer y KafkaConsumer en casos específicos de negocio.

VI. METODOLOGÍA

Los contenidos se desarrollarán siguiendo la secuencia teoría – práctica con evaluaciones modulares, utilizando videos grabados y lecturas como recursos didácticos.

VII. CONTENIDOS

Total horas	3 horas
Sesión 1	
• Tema 1: Pipeline. Definición, arquitectura básica: Data source, ETL, Data storage.	15 min.
• Tema 2: Batch Processing vs. Streaming Processing. Momento de ejecución, Data sources tradicionales, métricas de performance, casos de uso.	10 min.
• Tema 3: Apache Kafka. Características generales, arquitectura básica: Producers vs. Consumers.	10 min.
• Tema 4: Apache Kafka. Arquitectura interna: ZooKeeper, Topic, Broker.	10 min.
• Tema 5: Descarga y secuencia de puesta en marcha. Ejemplo desde la consola Windows.	20 min.
Lectura complementaria 1: https://book.huihoo.com/pdf/confluent-kafka-definitive-guide-complete.pdf	
Sesión 2	
• Tema 6: Librería KafkaProducer. Lista de parámetros, configuración básica.	10 min.
• Tema 7: Creación de Producer con Python I: Caso Enviar a Kafka datos digitados por el usuario.	15 min.
• Tema 8: Creación de Producer con Python II: Caso Enviar a Kafka datos contenidos en un CSV.	15 min.
• Tema 9: Creación de Producer con Python III: Caso Enviar a Kafka datos a partir de la visita a una página web.	20 min.
• Tema 10: Librería KafkaConsumer: Lista de parámetros, configuración básica.	10 min.
• Tema 11: Creación de Consumer con Python: Caso Leer de Kafka y almacenarlo en DataFrame.	15 min.
• Tema 12: Creación de Consumer con MongoDB: Instalación y configuración de “MongoDB conector for Apache Kafka”.	20 min.
• Tema 13: Apache Kafka en entornos Big Data, posibilidades de su integración con Hadoop y Apache Spark.	10 min.

VIII. SISTEMA DE EVALUACIÓN

La nota final del curso se obtendrá a partir de los siguientes rubros de evaluación y pesos asociados:

Rubro	Peso
Evaluación	100%

IX. Docente

Guillermo Suxo

Resumen

- Ingeniero de Sistemas y magister en Gestión de Tecnología de la Información y Comunicaciones (c).
- Tengo más de 10 años de experiencia en el campo de tecnología de la información para el desarrollo de sistemas.
- Actualmente trabaja como Data Engineer en el Banco de Crédito del Perú.