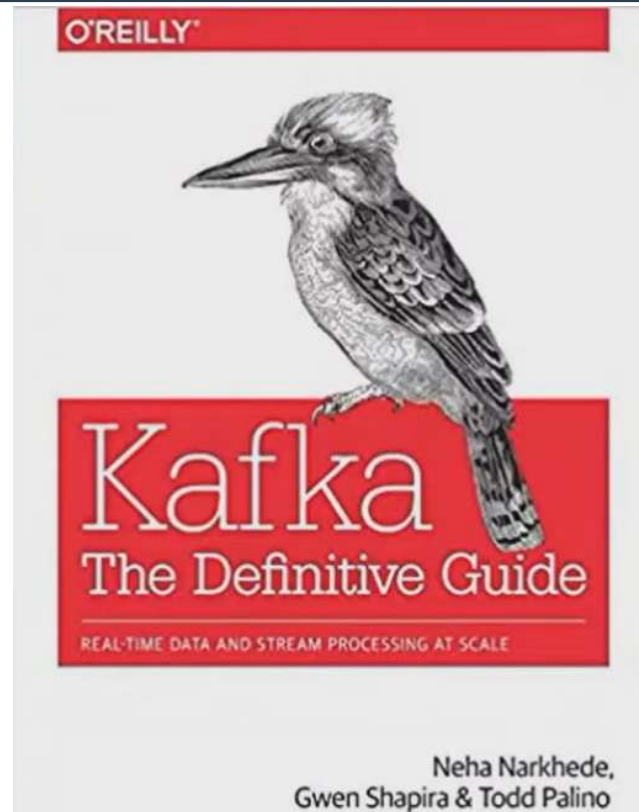




## **Módulo Profesional: Big Data Aplicado**

Kafka

# RECOMENDACIONES



<https://www.confluent.io/resources/ebook/kafka-the-definitive-guide/>

# ÍNDICE

1. Introducción
2. Arquitectura Central
3. Características
4. Kafka Streams
5. Ecosistema

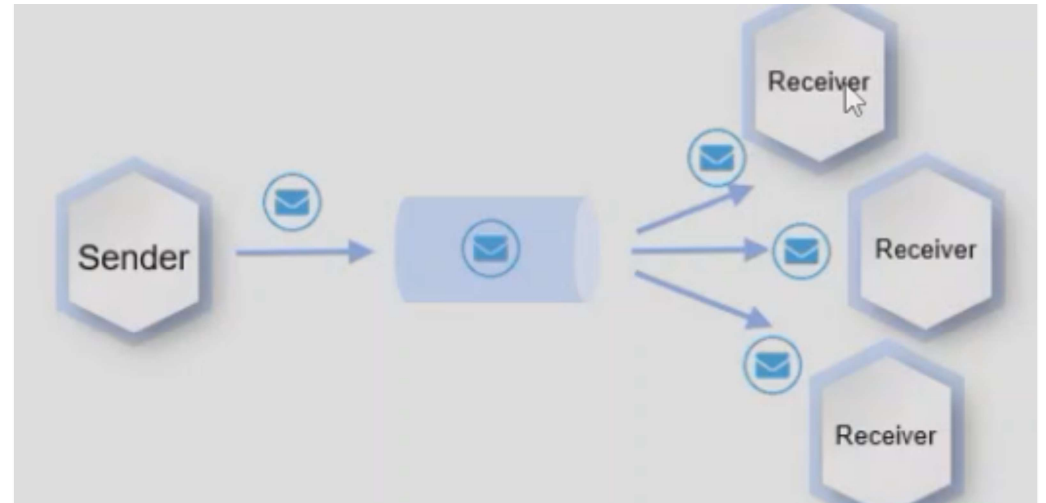
# Introducción

¿KAFKA DE QUÉ TIPO DE SISTEMA DE MENSAJERÍA ES?

**Sistema de mensajería punto a punto**



**Sistema de mensajería Publish-Subscribe**



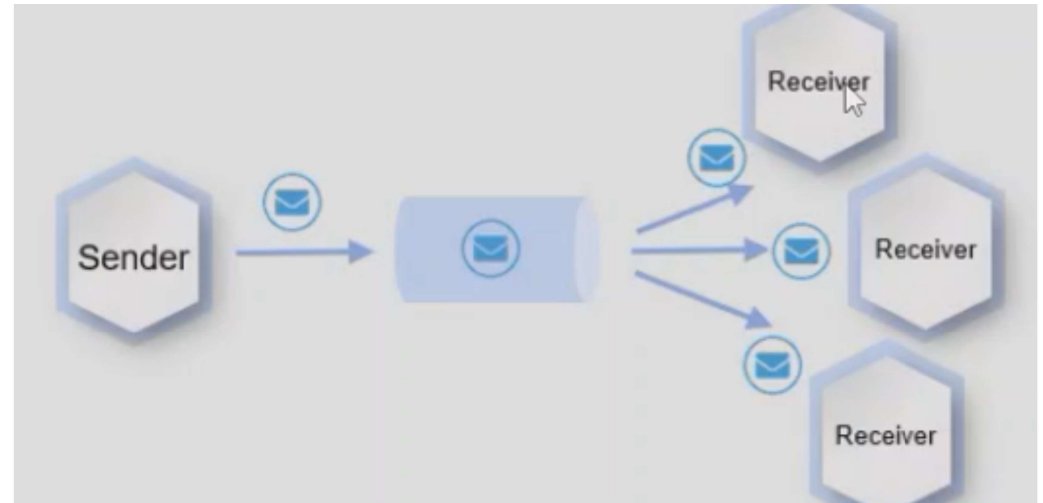
# Introducción

**Sistema de mensajería punto a punto**



**Sistema de mensajería Publish-Subscribe**

**Kafka es de este tipo**



# Apache Kafka

- Kafka como un sistema de mensajería Publicación/Suscripción
- Kafka como un sistema de almacenamiento
- kafka es un sistema distribuido

## Características

- Rápido, baja latencia
- Escalable
- Duradero
- Tolerante a fallos
- Confiable
- Replicación

# Apache Kafka

Consumers

## PRODUCERS

Allows an application to publish a stream of records to one or more Kafka topics.

## CONNECTORS

Allows building and running reusable producers or consumers that connect Kafka topics to existing applications or data systems

## STREAMS

Allows an application to act as a stream processor, consuming an input stream from one or more topics and producing an output stream to one or more output topics, effectively transforming the input streams to output streams

## CONSUMERS

Allows an application to subscribe to one or more topics and process the stream of records produced to them

API

# Kafka como Streaming Platform

- Publicar y suscribirse a flujos de registros.
- Almacenar flujos de registros de forma tolerante a fallos.
- Procesar flujos de registros en tiempo real o desde cualquier punto en el tiempo.

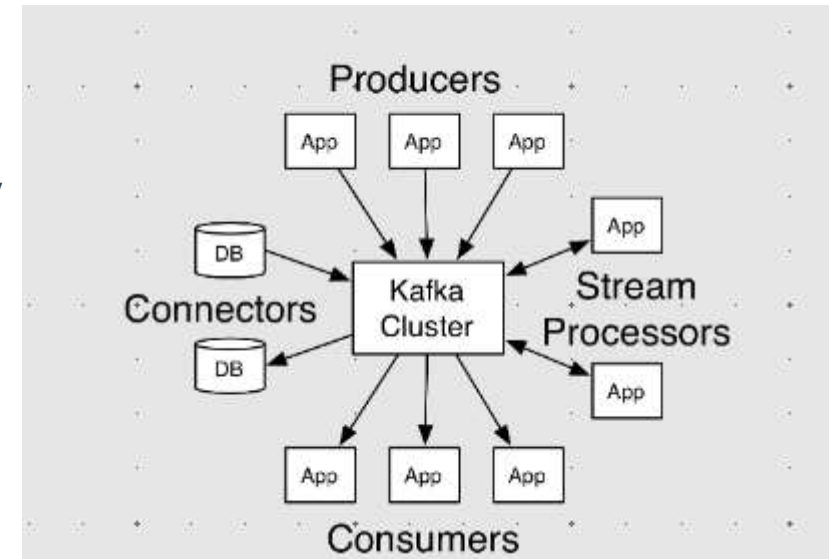
## **Cuatro APIs principales:**

Producer/Consumer API – Permite publicar flujos de registros y suscribirse a uno o más topics.

Streams API – API de nivel superior basada en Producer y Consumer API para procesar y transformar datos en tiempo real.

Connector API – Facilita la integración con sistemas externos mediante productores y consumidores reutilizables.

KSQL – Motor de procesamiento basado en SQL sobre Streams API, que permite consultas declarativas sobre flujos de datos en Kafka





# Confluent Kafka

<https://www.confluent.io/>

Confluent es para Apache Kafka lo que Cloudera es para Hadoop.  
También ofrece Confluent Cloud, es decir, Apache Kafka como servicio.

## Apache Kafka

Licencia Apache v2.0  
Kafka Brokers (es decir, los demonios)  
Producer y Consumer API  
Streams API  
Connect API

## Confluent Open

Cientes (Java, C, C++, Python, .)  
Conectores (JDBC, Elasticsearch, HDFS, etc.)  
Schema Registry  
REST Proxy  
KSQL

## Confluent Enterprise

Control Center para monitoreo  
Replicación en múltiples centros de datos  
Balanceo automático de datos  
Soporte 24/7

## 2. Arquitectura Central. Kafka Cluster

- El clúster de Kafka está compuesto por múltiples brokers.
- Todos los brokers son iguales, no hay maestro/esclavo.
- Los brokers generalmente se ejecutan de forma nativa como un servicio del sistema operativo.
- Los brokers los registros en disco y replicación.
- Zookeeper se encarga de mantener el estado del clúster. Gestiona la configuración, la elección de líderes y la coordinación del clúster.

