

Formación senior en Kafka

PARTE DE LA FORMACIÓN BIG DATA ENGINEER DE BIG DATA
ACADEMY PERÚ

Concepto

BIG DATA
ACADEMY

Kafka

Es una plataforma para la **construcción de pipelines de datos real time** que implementa el **patrón “publicador/suscriptor”**. Es altamente **escalable** y **paralelizable**. Se ejecuta **sobre un clúster** y es **tolerante a fallos**.



Objetivo fundamental

Objetivo fundamental de Kafka

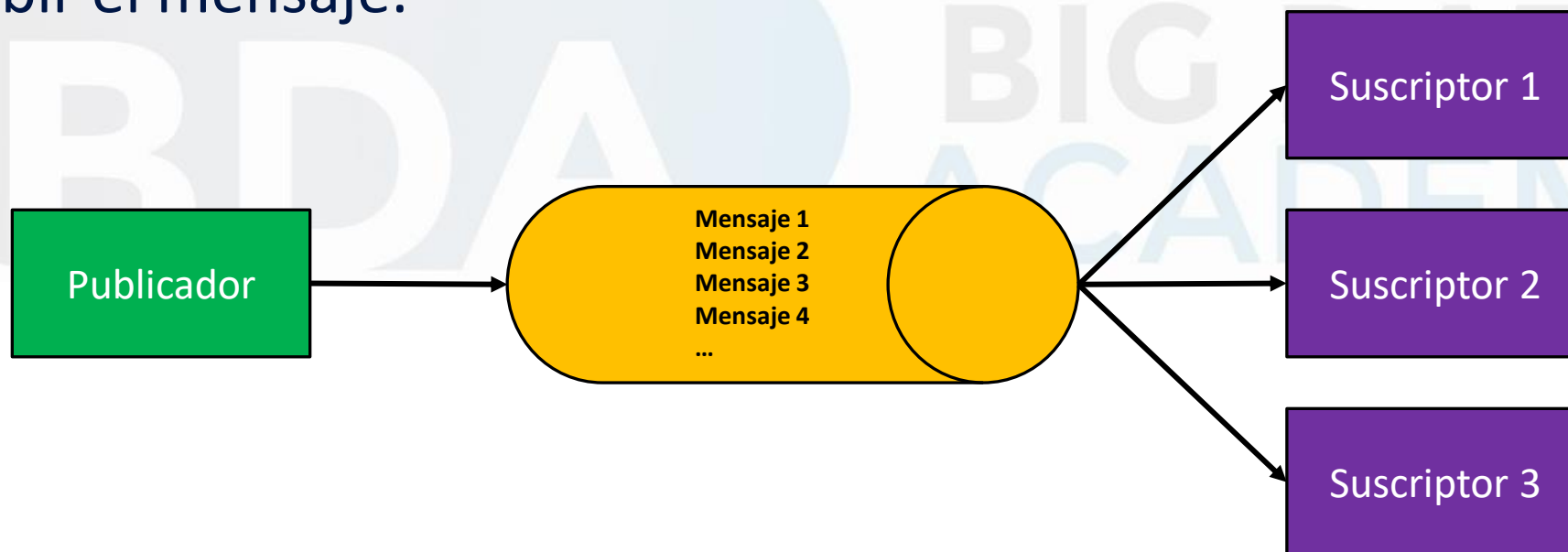
1. Almacenar con baja latencia **datos de manera temporal**
2. Distribuir los datos a uno o varios sistemas

Naturaleza de funcionamiento

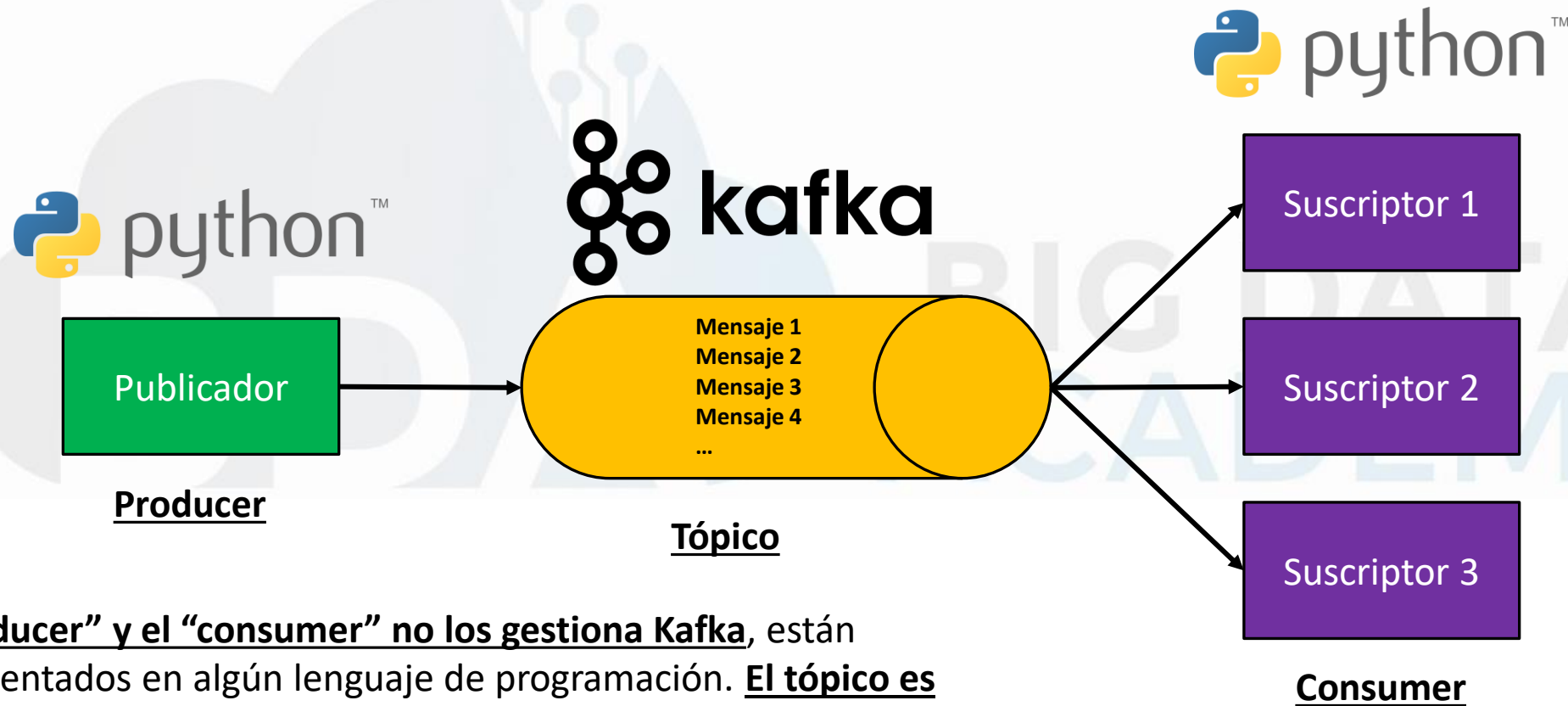


El patrón publicador/suscriptor

Indica que existe un componente que envía uno o varios mensajes, estos mensajes son almacenados en un punto centralizado y luego distribuido a todos aquellos componentes que estén suscritos para recibir el mensaje.

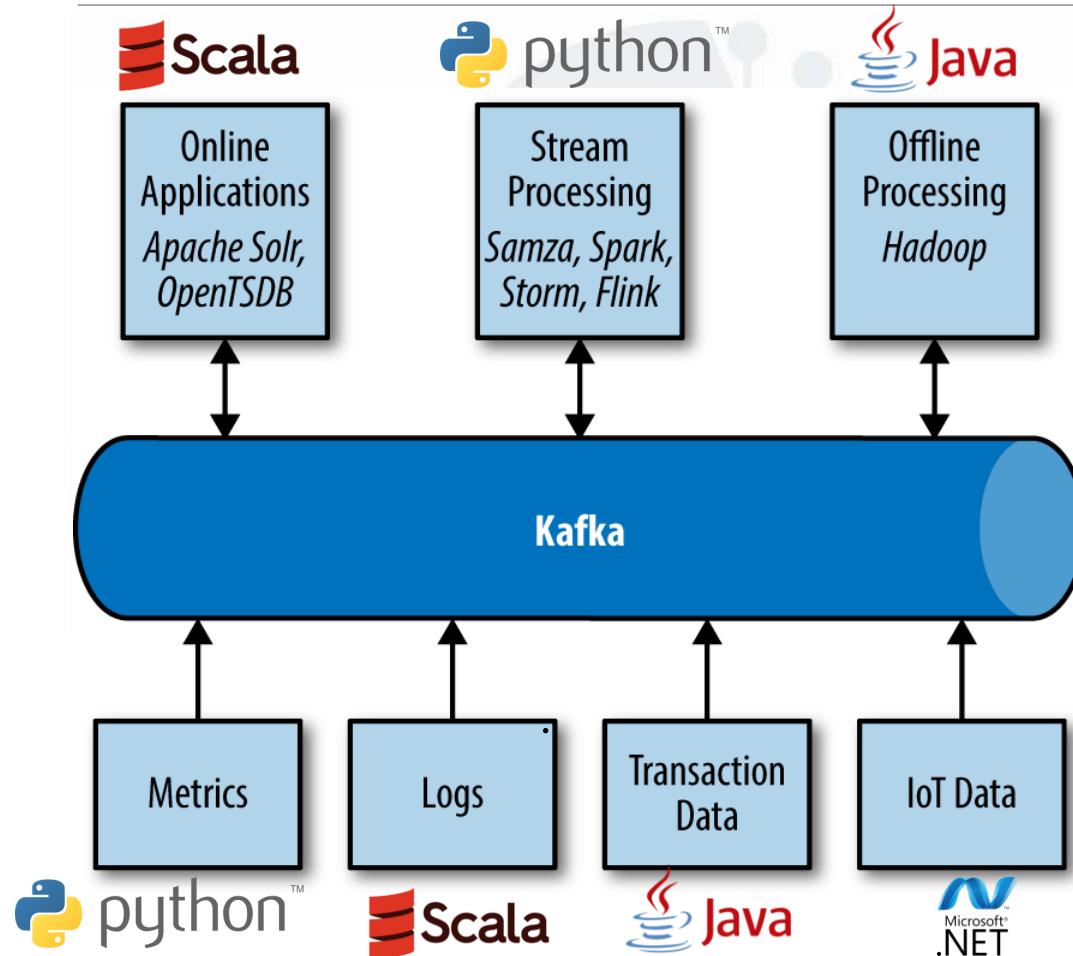


El patrón publicador/suscriptor en Kafka



El “producer” y el “consumer” no los gestiona Kafka, están implementados en algún lenguaje de programación. El tópico es como una “tabla” que gestiona Kafka

Integrando sistemas real time con Kafka



3. **Crear el consumer** en algún lenguaje de programación que extraiga los datos del sistema y los envíe a otros sistemas

1. **Crear en Kafka el tópico** que almacenará los mensajes

2. **Crear el producer** en algún lenguaje de programación que escriba los datos del sistema y los coloque en el tópico

Kafka funciona sobre un clúster

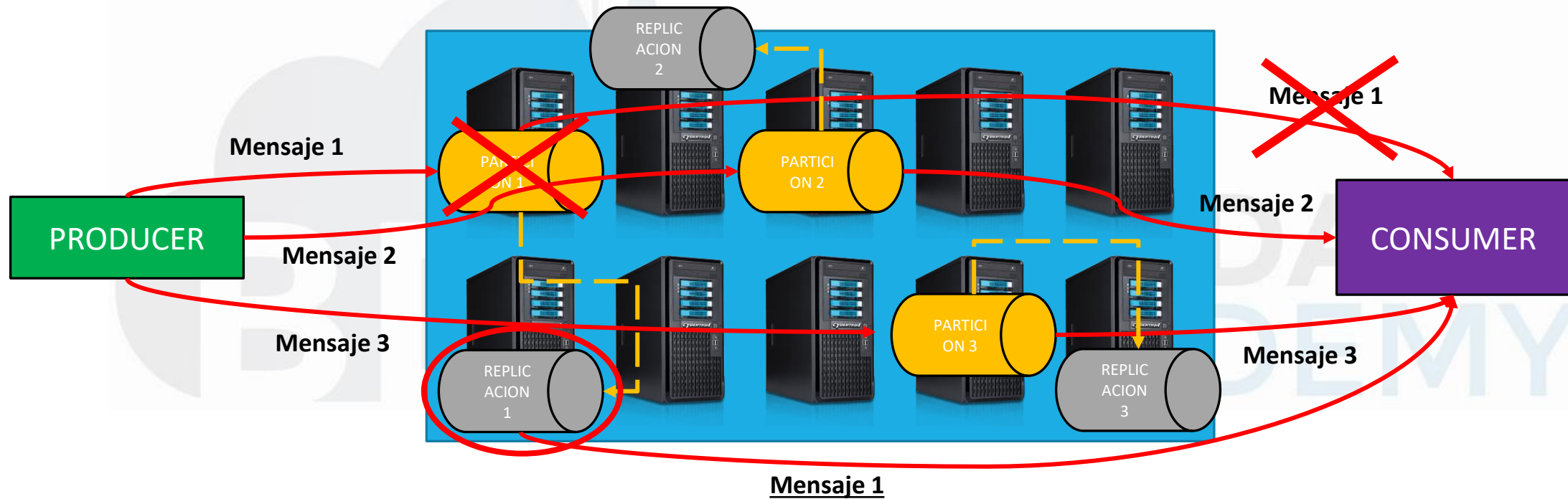


Los mensajes de los tópicos son colocados alrededor de varios nodos.

Anatomía de un tópico

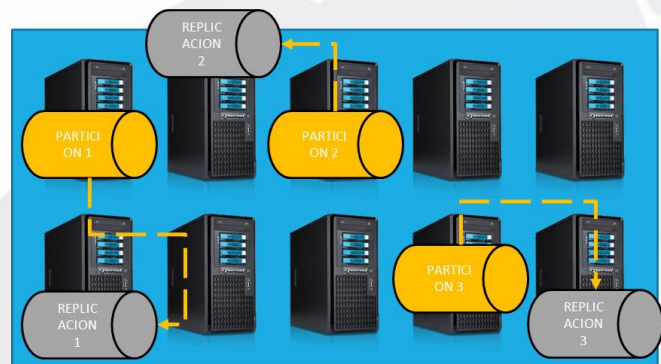


Tolerancia a fallos



Kafka permite configurar un factor de replicación para evitar pérdida de datos cuando un nodo colapse. **La replicación se realiza a nivel de partición.**

Clúster dedicado



Clúster Kafka

Kafka necesita de nodos dedicados para funcionar, no podemos re-usar los nodos de Hadoop

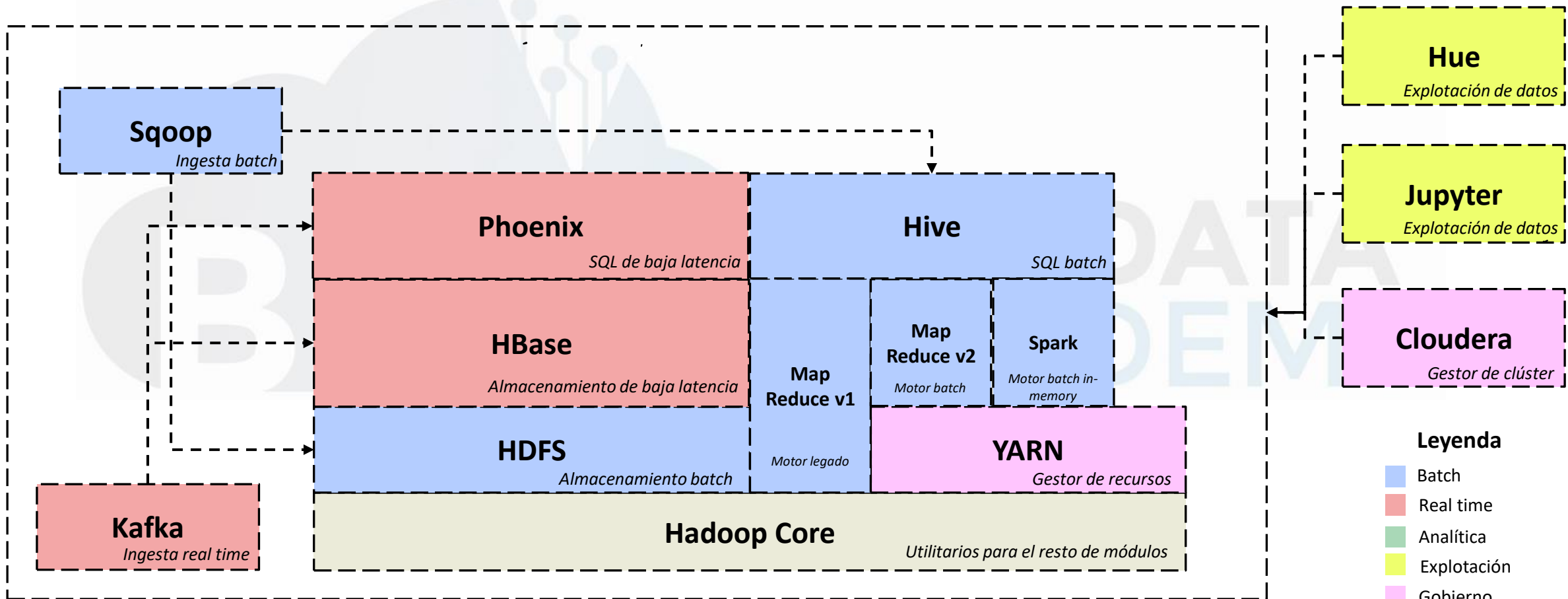


Clúster Hadoop

Arquitectura de componentes



Arquitectura de componentes



Arquitectura de servicios

Arquitectura de servicios



Ejercicios teóricos

Ejercicios teóricos

1. ¿Qué es Kafka?
2. ¿Cuál es el objetivo de Kafka?
3. ¿Cómo implementa Kafka el patrón publicador/suscriptor?
4. ¿Qué es un producer?
5. ¿Qué es un consumer?
6. ¿Qué es un tópico?
7. ¿Por qué Kafka requiere de un clúster dedicado?
8. ¿Cuáles son los servicios de Kafka?
9. Dibuje la arquitectura de componentes del ecosistema Big Data

Resumen

BIG DATA
ACADEMY

Resumen

Hablemos...