

Kafka

Apache Kafka es una plataforma de streaming distribuida que permite la creación y gestión de flujos de datos en tiempo real. Fue desarrollada inicialmente por LinkedIn y posteriormente donada a la fundación Apache. Kafka permite que sistemas y aplicaciones intercambien información en tiempo real mediante la publicación y suscripción a eventos (mensajes). Su diseño es altamente escalable, tolerante a fallos y garantiza una baja latencia en el procesamiento de datos, lo que lo hace ideal para entornos de datos de alto volumen.

Características Principales de Apache Kafka

- **Ítem de lista desordenada**Publicación/Suscripción de Mensajes en Tiempo Real: Kafka funciona en un modelo publish-subscribe, donde los productores (producers) publican mensajes a "tópicos" (topics) y los consumidores (consumers) se suscriben a esos tópicos para recibir los mensajes en tiempo real.
- **Ítem de lista desordenada**Persistencia y Durabilidad: Kafka guarda los mensajes en disco de manera duradera, permitiendo que los datos persistan incluso después de ser consumidos. A diferencia de otros sistemas de mensajería, donde los mensajes se eliminan una vez consumidos, Kafka permite múltiples consumos de los mismos datos.
- **Ítem de lista desordenada**Distribución y Escalabilidad Horizontal: Kafka está diseñado para ser un sistema distribuido. Cada componente de Kafka (brokers, productores, consumidores) puede escalar horizontalmente agregando más instancias, permitiendo que el sistema maneje grandes volúmenes de datos de manera efectiva.
- **Ítem de lista desordenada**Tolerancia a Fallos: Kafka distribuye los datos entre varios nodos (brokers) y réplicas, lo que permite que el sistema continúe funcionando correctamente aunque uno de sus nodos falle.
- **Ítem de lista desordenada**Baja Latencia y Alto Rendimiento: Kafka está optimizado para la transferencia de grandes volúmenes de datos con baja latencia, lo que lo hace ideal para aplicaciones que requieren respuestas rápidas en entornos de transmisión de datos.

Componentes Técnicos de Kafka

- **Ítem de lista desordenada**Producers

Los producers son las aplicaciones o sistemas que envían datos a Kafka. Envían mensajes a un tópico específico y pueden definir una clave para determinar en qué partición de ese tópico se almacenará el mensaje.

- **Ítem de lista desordenada**Topics y Partitions

Un topic es una categoría o canal donde se publican los mensajes. Cada tópico se subdivide en varias particiones, que permiten distribuir los datos entre diferentes nodos para mejorar la escalabilidad y paralelismo. Los mensajes en una partición son ordenados y numerados (mediante un offset), lo que permite a los consumidores leer de manera secuencial y en el orden en que fueron producidos.

- **Ítem de lista desordenada**Consumers y Consumer Groups

Los consumers son las aplicaciones o sistemas que leen datos de Kafka. Se suscriben a uno o más

tópicos para recibir mensajes. Los consumer groups permiten que varios consumidores trabajen juntos para leer datos de un tópico de manera paralela. Cada partición se asigna a un único consumidor dentro de un grupo, asegurando que cada mensaje sea procesado una sola vez por el grupo.

- Ítem de lista desordenada Brokers

Los brokers son los servidores que forman el clúster de Kafka y almacenan los datos de los tópicos en sus particiones. Cada broker es responsable de una parte de las particiones y sus réplicas. Kafka distribuye automáticamente las particiones y réplicas entre los brokers para equilibrar la carga.

- Ítem de lista desordenada Zookeeper

Apache Kafka utiliza Zookeeper para la coordinación del clúster. Zookeeper ayuda a gestionar la configuración y el estado de cada broker en el clúster, mantiene la información de particionamiento y facilita la elección de líderes para cada partición. A partir de Kafka 2.8, se está moviendo hacia un nuevo sistema de metadatos que no depende de Zookeeper.

- Ítem de lista desordenada Kafka Connect

Es una herramienta para la integración de datos con Kafka, que permite la conexión con bases de datos, sistemas de archivos, sistemas de mensajería y otras fuentes de datos externas mediante connectors predefinidos. Esto facilita la importación y exportación de datos entre Kafka y otros sistemas.

- Ítem de lista desordenada Kafka Streams

Kafka Streams es una biblioteca para el procesamiento de datos en tiempo real directamente sobre Kafka. Permite realizar transformaciones y análisis en los flujos de datos sin necesidad de una infraestructura adicional. **Funcionamiento Básico Producción de Mensajes:** Los productores envían mensajes a un tópico específico. Kafka almacena estos mensajes en la partición designada, donde los consumidores pueden leerlos en el orden en que fueron publicados.

Almacenamiento de Mensajes: Kafka almacena los mensajes en disco y permite que estos persistan durante un tiempo configurado o hasta que el clúster alcance un tamaño límite definido. Los mensajes permanecen en las particiones hasta que se eliminan de acuerdo a estas configuraciones de retención.

Consumo de Mensajes: Los consumidores se suscriben a uno o más tópicos y leen los mensajes en el orden en que fueron almacenados. El consumo es asíncrono, y cada consumidor dentro de un grupo de consumidores puede leer de una partición específica.

Offset: Cada mensaje en una partición tiene un identificador único, llamado offset. Este offset permite a los consumidores realizar un seguimiento de los mensajes que ya han procesado y retomar la lectura desde el último mensaje consumido en caso de fallos o interrupciones.

Ejemplo Práctico Supongamos que tienes una aplicación de e-commerce que necesita procesar eventos de compra en tiempo real. Kafka podría usarse para manejar el flujo de estos eventos de la siguiente manera:

Cada vez que se realiza una compra, se genera un evento que el productor envía al tópico compras. Cada mensaje en el tópico compras contiene detalles de la transacción: ID de cliente, productos

comprados, importe, etc. Los consumers, como los sistemas de inventario, notificaciones y análisis de datos, se suscriben al tópico compras para procesar los eventos en tiempo real. El sistema de inventario actualiza las existencias, el sistema de notificaciones envía un mensaje de confirmación de compra al cliente, y el sistema de análisis registra la compra para futuras evaluaciones de negocio.

Ventajas de Apache Kafka

- Alto Rendimiento y Baja Latencia:** Kafka es capaz de procesar millones de mensajes por segundo, lo que lo hace ideal para aplicaciones de tiempo real.
- Escalabilidad Horizontal:** Kafka permite añadir más brokers, productores y consumidores para aumentar su capacidad de procesamiento.
- Persistencia de Datos:** Kafka almacena los datos en disco y permite que se mantengan durante un período configurado, facilitando su uso en aplicaciones de análisis histórico.
- Tolerancia a Fallos:** La replicación de particiones asegura que los datos estén protegidos ante fallos de nodo.

Casos de Uso Comunes

- Procesamiento de Datos en Tiempo Real:** Se usa en aplicaciones de streaming, como monitoreo de sistemas, detección de fraudes y análisis de redes sociales.
- Integración de Sistemas:** Kafka es ideal para comunicar diferentes servicios en arquitecturas de microservicios.
- ETL (Extracción, Transformación y Carga):** Kafka permite recolectar y transformar datos de diferentes fuentes antes de cargarlos en un data warehouse o un sistema de análisis.
- Pipeline de Datos:** Kafka puede actuar como una columna vertebral para un pipeline de datos, conectando sistemas y aplicaciones en toda una organización.

En cada servidor (desarrollo(10.1.120.41) ,QA (10.1.40.34) y producción (10.1.40.35)) hay un servidor kafka instalado en la maquina (no esta contenerizado por el momento).

[←Regresar](#)

From:
<http://wiki.adacsc.co/> - Wiki

Permanent link:
<http://wiki.adacsc.co/doku.php?id=ada:howto:sicoferp:factory:new-migracion-sicoferp:kafka>

Last update: **2024/11/06 20:34**

