

kafka

Apache Kafka



Indice

- Cos'è Apache Kafka
- Componenti Principali
- Architettura
- Funzionalità
- Creazione del Messaggio
- Caso d'uso

INTRODUZIONE:

COS'E' APACHE KAFKA

Apache Kafka è una piattaforma di streaming distribuito, progettata per gestire flussi di dati in tempo reale. È utilizzato principalmente per costruire sistemi scalabili, altamente disponibili e con bassa latenza, dove grandi volumi di dati devono essere trasmessi, elaborati e archiviati in modo efficiente. Kafka nasce come sistema di messaggistica che consente di inviare dati tra produttori e consumatori, con una particolare attenzione alla persistente gestione dei dati e alla tolleranza ai guasti. Grazie alla sua architettura distribuita e alla capacità di scalare orizzontalmente, Kafka è diventato uno degli strumenti più popolari per costruire applicazioni di streaming, analisi dei dati in tempo reale e integrazione tra sistemi.

COMPONENTI PRINCIPALI

Producer:

- Invia i messaggi a Kafka.
- Pubblica messaggi su un topic.

Consumer:

- Legge i messaggi da Kafka.
- Si iscrive ai topic per ricevere i messaggi.

Broker:

- Server che gestisce i messaggi e li distribuisce tra i consumatori.

Topic:

- Categoria in cui i messaggi sono organizzati.
- Ogni topic può essere suddiviso in più partizioni.

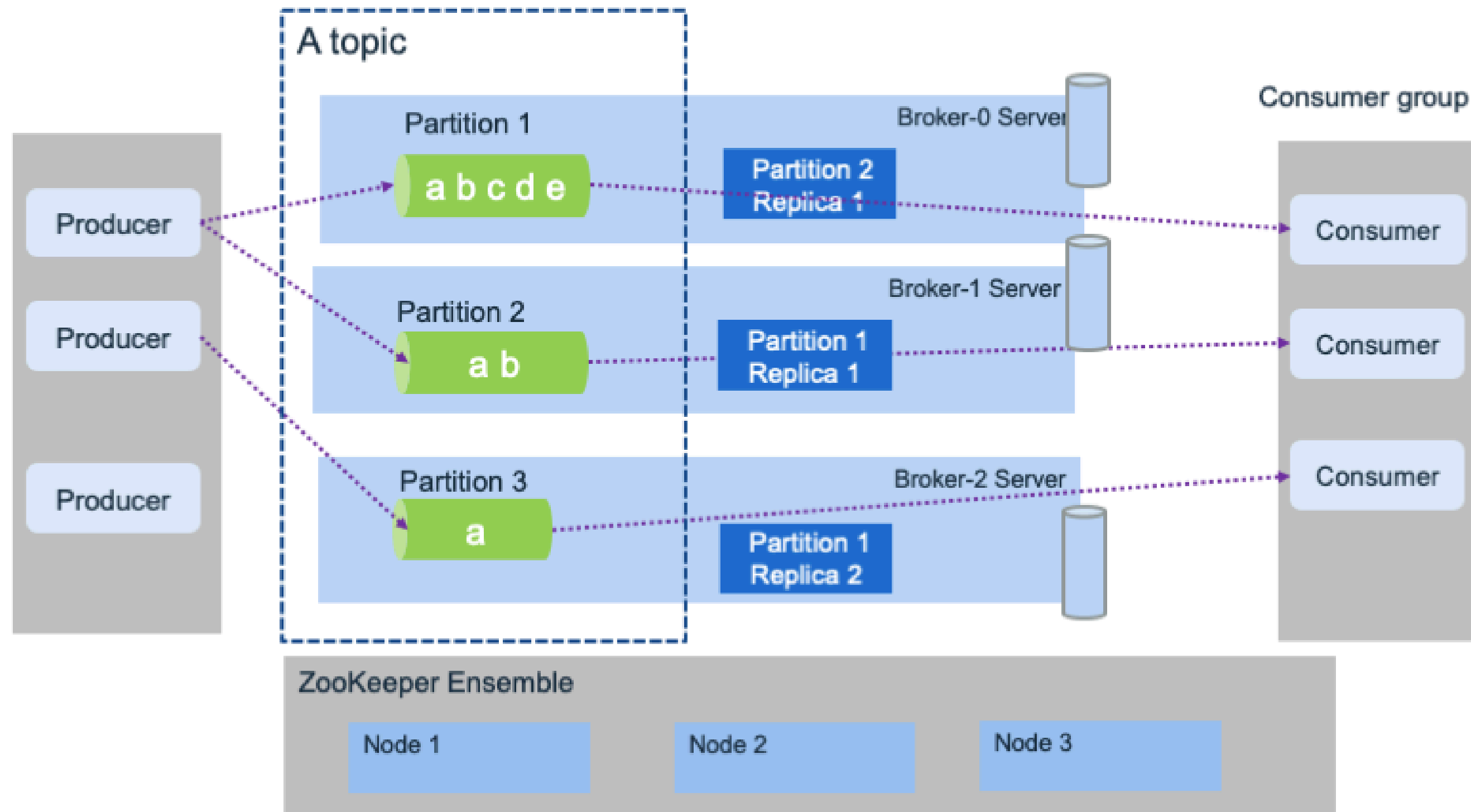
Zookeeper:

- Strumento per la gestione e il coordinamento dei cluster Kafka.
- Gestisce la configurazione e la gestione dei broker.

Cluster Kafka:

- Insieme di broker che lavorano insieme per garantire la distribuzione dei dati e l'alta disponibilità.

ARCHITETTURA:



FUNZIONALITÀ:

Persistenza:

I dati vengono memorizzati su disco.

Scalabilità Orizzontale:

I broker possono essere aggiunti facilmente per gestire più carico.

Fault Tolerance:

La replica dei messaggi tra i broker garantisce la tolleranza ai guasti.

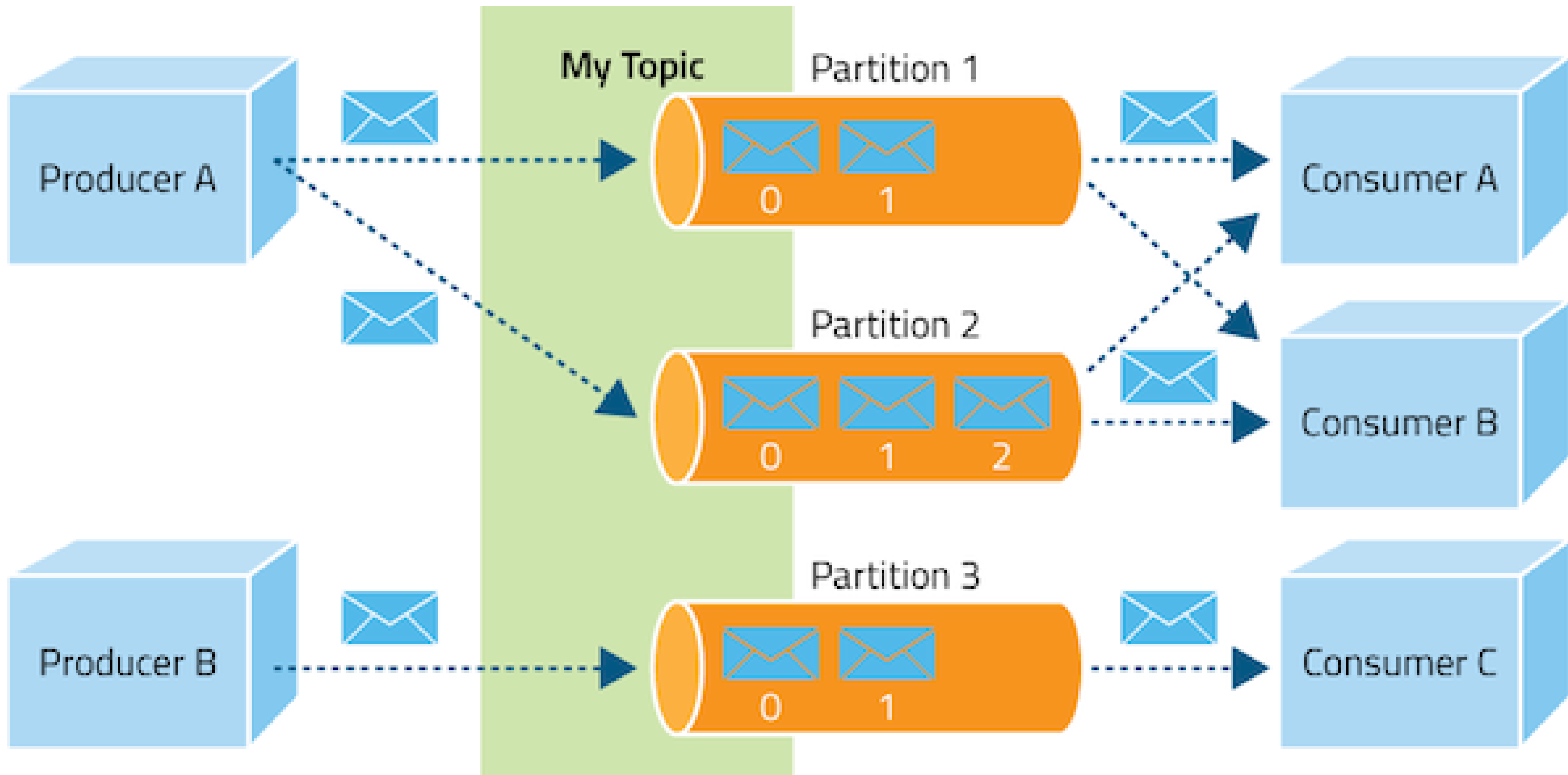
Elaborazione in tempo reale:

Kafka è progettato per l'elaborazione e la gestione di flussi di dati in tempo reale.

Distribuzione e Consumo parallelo:

Consente la lettura e scrittura simultanea su più partizioni, migliorando le prestazioni

CREAZIONE DEL MESSAGGIO:



CREAZIONE DEL MESSAGGIO

Componente Coinvolto: Producer

Il Producer è il componente che crea e invia il messaggio al broker Kafka.

Struttura del Messaggio:

Key: Opzionale, può essere utilizzata per determinare la partizione del topic in cui il messaggio verrà scritto.

Value: Il dato principale del messaggio (il corpo del messaggio). Questo può essere un qualsiasi tipo di dato serializzabile (es. JSON, Avro, stringa, ecc.).

Timestamp: Data e ora di quando il messaggio è stato creato.

Headers: Opzionale, metadata addizionali che possono essere usati per identificare o descrivere meglio il messaggio.

Serializzazione:

Prima di inviare il messaggio al broker, i dati devono essere serializzati in un formato binario. Kafka supporta diverse modalità di serializzazione (es. StringSerializer, IntegerSerializer, AvroSerializer, ecc.).

Invio al Broker:

Il producer invia il messaggio al topic Kafka. Se il producer è configurato per farlo, Kafka determinerà la partizione in base alla key del messaggio (se presente) oppure in modo casuale.

CASI D'USO

Streaming di dati:

Raccolta di log, eventi in tempo reale e analisi.

Elaborazione in tempo reale:

Analisi dei dati in tempo reale (ad esempio, sistemi di raccomandazione).

Integrazione di sistemi:

Kafka viene utilizzato come middleware tra sistemi differenti per la trasmissione di messaggi

Tracking di attività:

Kafka viene utilizzato per tracciare le attività degli utenti su applicazioni web, mobile e altri sistemi digitali.

Aggregazione di log:

Kafka è spesso utilizzato per raccogliere e centralizzare i log provenienti da diversi sistemi e applicazioni.