

# apache kafka

- O que é o apacha kafka ?
  - é uma plataforma distribuida de streaming.
  - De forma distribuída ele consegue processar uma grande quantidade de dados e entregá-los em tempo real aos seus consumidores.
- CARACTERISTICAS
  - Plataforma
  - trabalha de forma distribuida
  - Banco de dados
  - extremamente rápido é com baixa latência
  - Utiliza discos ao invés de memoria para processar os dados
  - **NÃO É APENAS UM SISTEMA TRADICIONAL DE FILAS COMO O RABBITMQ**
    - TRABALHA TANTO COM FILHAS MAS TAMBÉM TRABALHA COM PUBLISH E SUB

## ▼ Conceitos básicos

- producer(quem produz a mensagem) e consumer(quem consome a mensagem)
  - imagine que um sistema A envia uma mensagem diretamente para o sistema B

O kafka ele ficaria no meio termo ele ficaria entre essa transação de mensagem

onde o producer é o sistema A que produz a mensagem e envia para um tópico do kafka é o sistema B é o consumer que consome essa mensagem do tópico

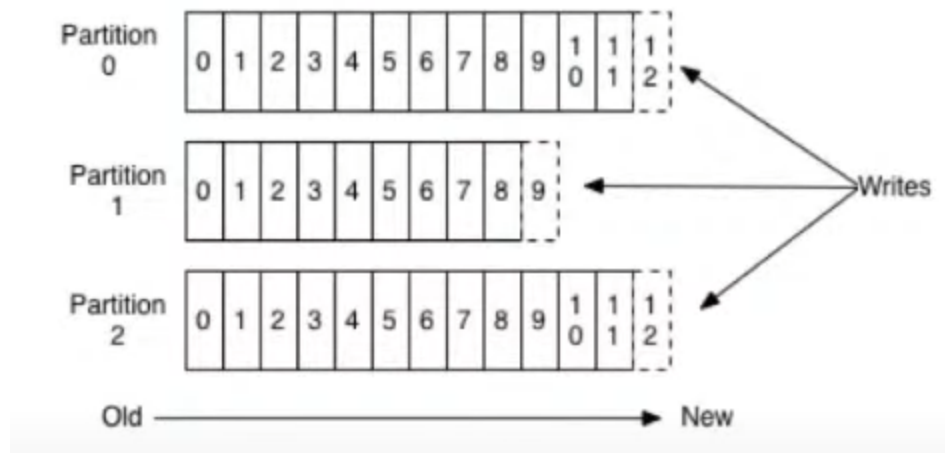
producer(quem produz a mensagem) e consumer(quem consome a mensagem)

tópico "file"

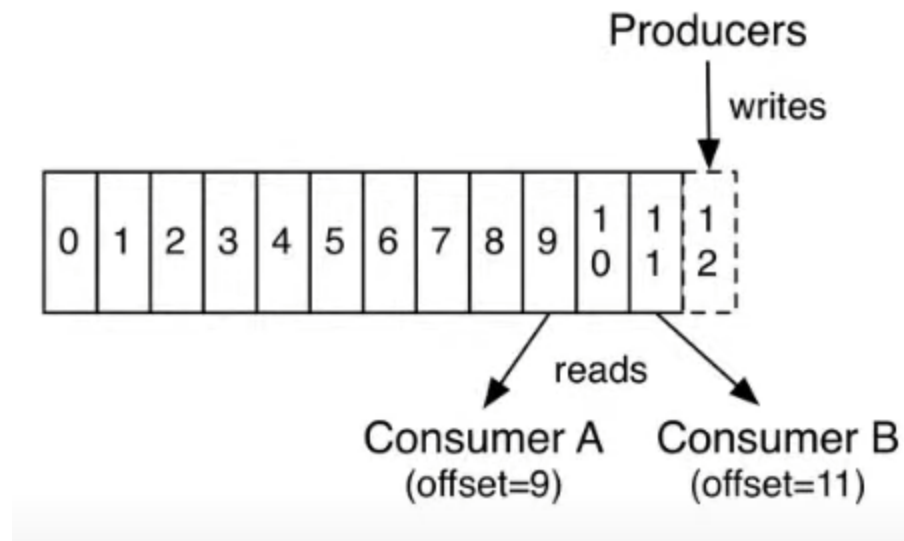
## Conceito básico



- o que é um topic
  - Stream de dados que atua como um banco de dados
    - ele salva todas as mensagens enviadas. podem ser lidas ou processadas posteriormente
  - Todos os dados ficam armazenados, ou seja, cada Topic tem seu "local" para armazenar seus dados.
  - Tópico possui diversas partições
    - Cada partição é definido por número.Ex:0,1,2
    - Você é obrigado a definir a quantidade de partições quando for criar um Topic
  - Anatomia de um topic
    - Cada topic é formado de partições



Cada partição tem offsets onde as mensagens ficam armazenadas onde consumidores diferentes podem ler as mensagens desses OBS não é uma leitura na ordem cada de envio

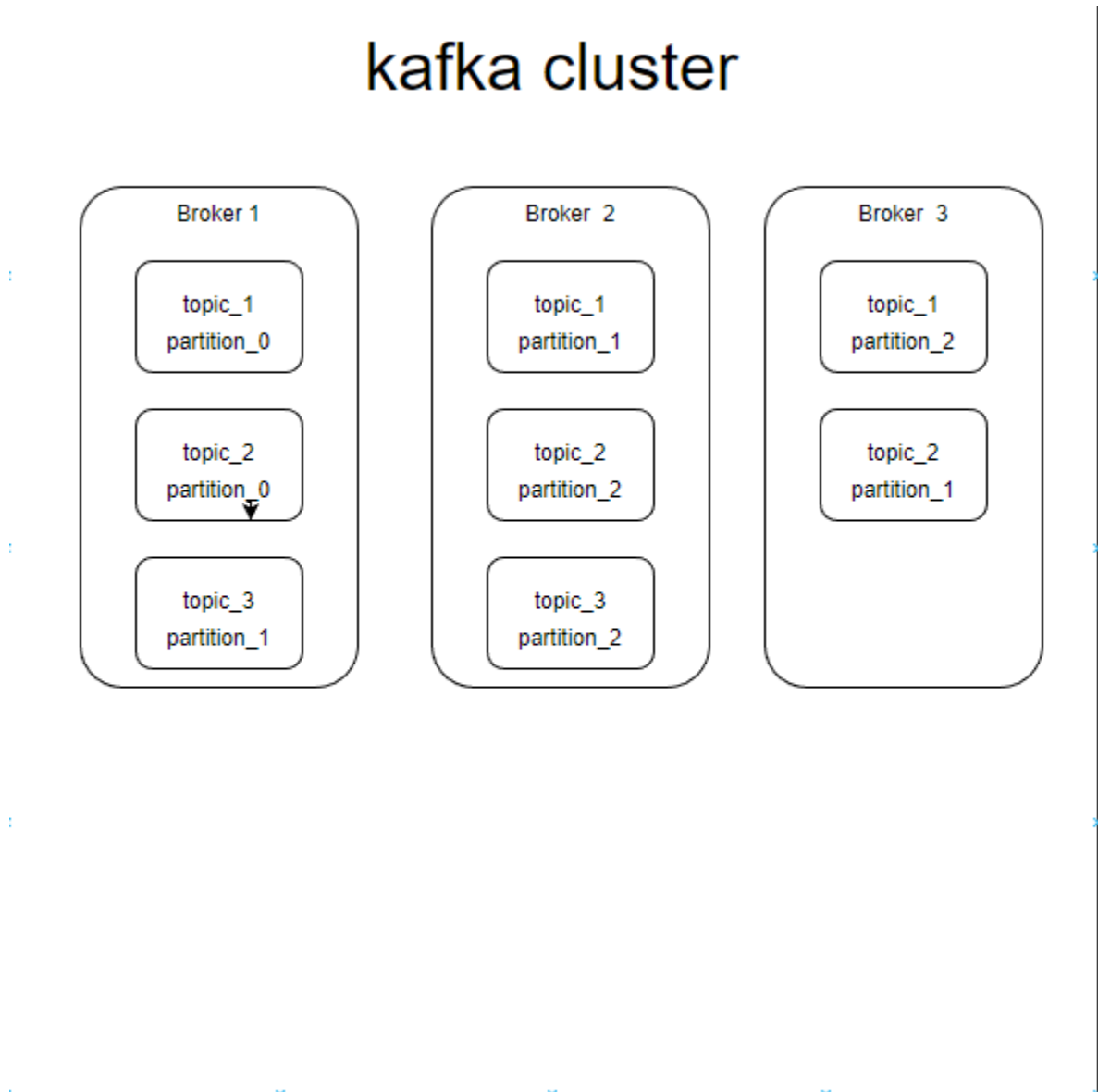


#### ▼ cluster kafka

- Conjunto de Brokers
- Cada Broker é um server por isso ele funciona de forma distribuida
- Cada Broker é responsável por armazenar os dados de uma partição
- Cada partição de Topic está distribuído em diferente brokers

exemplos:

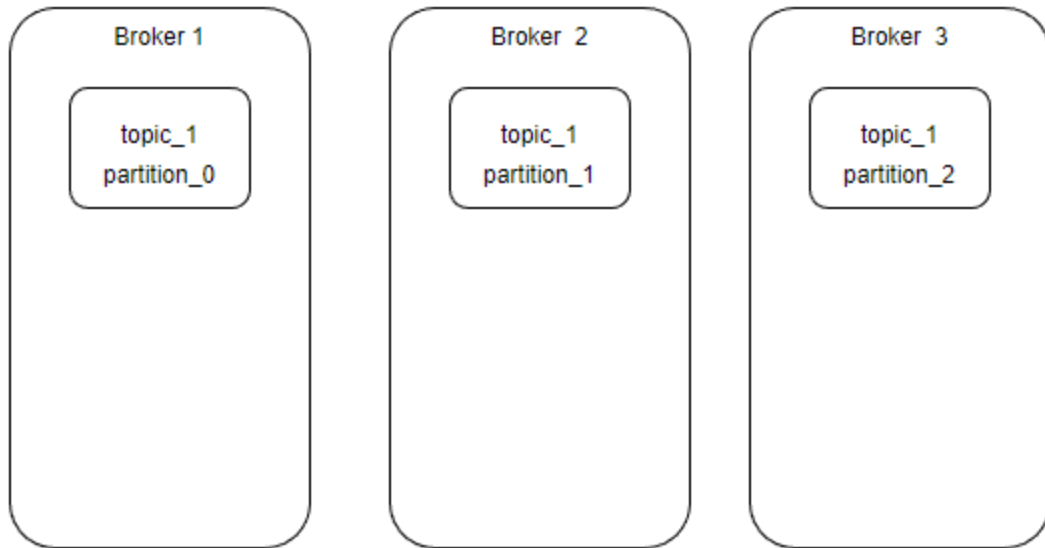
Perceba que os tópicos estão e partições estão distribuídos em diferentes brokers



No momento em que um topico é criado é obrigatório escolher o numero de partições que esse topico tera e ao mesmo tempo sera preciso escolher o replication factor

# kafka cluster

## replication factor



o que é o replication factor: é basicamente quantas replicas vc quer garantir que uma partição tenha ( as replicas ficam em brokers diferentes )

perceba que no exemplo abaixo a caso o broker 1 caia o broker 2 tem uma copia da partition\_0.

# kafka cluster replication factor

