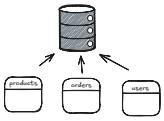
Topics & Partitions



Why Topics?

- One log is not enough
- We need to be able to separate events that belongs to different domains
- Similar concept as database tables
- Each "category" of data becomes a topic

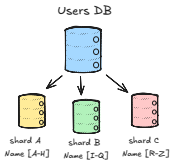
What is a Topic?

- A topic is a logical stream of data of the same type
- Think of it as a named log

Note:

Kafka clusters often hosts hundreds or even thousands of topics.
Each microservice or data domain can publish and consume its own topic.
This separation keeps data organized, secure, and easy to scale.

Quick view on sharding



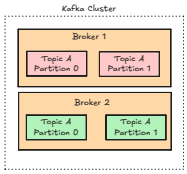
What is sharding?

- The process of dividing a large dataset into smaller pieces called "shards". Each shard is stored on a separate database server to improve performance, scalability and availability.

Note:

In Kafka, sharding refers to dividing a topic into smaller ordered logs called partitions

Partitions

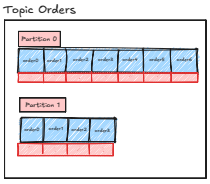


What are partitions?

- A partition is an independent log, stored and replicated across brokers (servers in a cluster)

- It allows the topic to scale horizontally

- A topic can have just 1 partition, but it's not common. It generally have dozens of it.



Partitions characteristics

- Each partition stores its own ordered logs of records
- Has offsets starting from 0
- Can be read and written independently
- More partitions = more scalability
- Ordering is only guaranteed inside each partition, not across the entire topic

Note:

Topic = collection of partitions
Topics organize your data
Partitions make your data scale