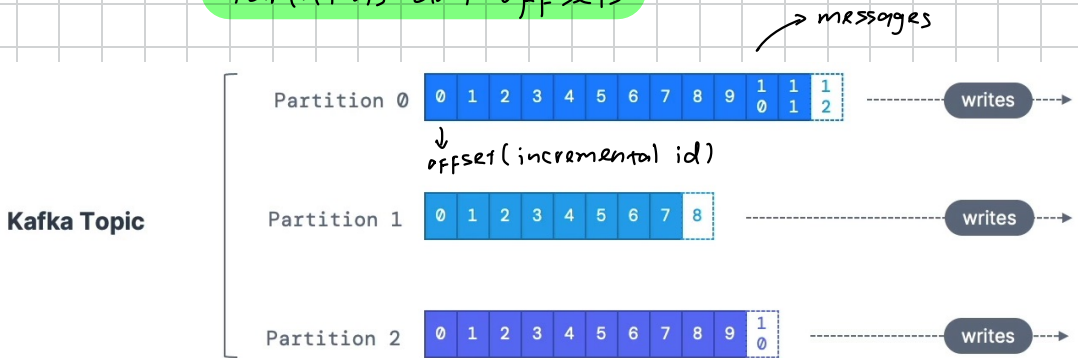


Topics

- A stream of data
- Like a table in database (without constraints)
- A topic is identified by its name
- Support any kind of message format
- Sequence of messages is called data stream
- Cannot query topics

Partitions and Offsets



- Kafka topics are immutable: once data is written into a partition, it cannot be changed

Producers

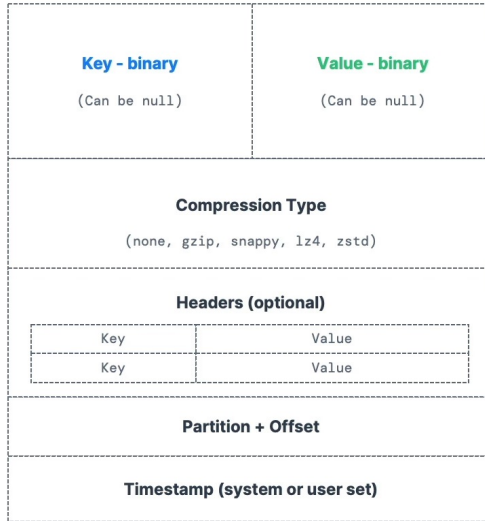


- Write data to Topics
- Know which partition to write to (and which broker has it)
- In case broker failures, producer will automatically recover

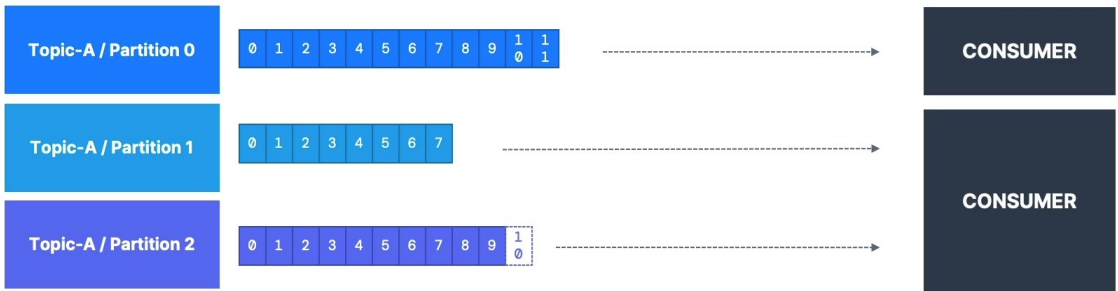
Message keys

- Producer can send a key with the message
 - + key = null → round robin (partition 0, 1, 2, ...)
 - + key != null → same partition (hashing)

Kafka
Message
Created by
the producer



Consumers



- Read data from topics (identified by name)
- Know which broker to read from
- In case of broker failure, know how to recover
- Data is read in order of offset in each partition

Broker

- A Kafka cluster is composed of multiple brokers (servers)
- Identified by ID(integer)
- Contain certain topic partitions