

# Top 20 Apache Kafka Interview Questions and Answers

## **1. What is Apache Kafka?**

Apache Kafka is a distributed event streaming platform used to build real-time data pipelines and streaming applications. It provides high-throughput, fault-tolerant, and scalable publish-subscribe messaging.

## **2. What are the main components of Kafka?**

Kafka has four main components: Producers, Consumers, Topics, and Brokers. Optionally, Zookeeper (or KRaft in newer versions) is used for cluster coordination.

## **3. What is a Kafka Topic?**

A Kafka Topic is a logical channel to which records are sent by producers and from which consumers read data.

## **4. What is a Partition in Kafka?**

Each topic is divided into partitions. A partition is an ordered, immutable sequence of records that allows Kafka to scale horizontally.

## **5. What is a Kafka Broker?**

A Kafka Broker is a server that stores data and serves client requests. Each broker in a cluster handles partitions and replicates data for fault tolerance.

## **6. What is the role of Zookeeper in Kafka?**

Zookeeper manages metadata, leader election, and configuration synchronization across Kafka brokers. (In newer versions, KRaft mode replaces Zookeeper).

## **7. What is the difference between Kafka and traditional message queues?**

Unlike traditional message queues, Kafka is distributed, persistent, and supports replay of messages and partition-based parallelism.

## **8. What is Kafka Producer?**

A Kafka Producer is an application that publishes messages to Kafka topics.

## **9. What is Kafka Consumer?**

A Kafka Consumer subscribes to one or more topics and processes the stream of records produced to them.

## **10. What is Kafka Consumer Group?**

A group of consumers that share the work of consuming and processing records from topics collectively.

## **11. What is an Offset in Kafka?**

An offset is a unique identifier assigned to each record within a partition. It tracks the consumer's position in the log.

#### **12. What is Retention Policy in Kafka?**

Retention policy defines how long Kafka retains data (based on time or size). Once the limit is reached, old data is deleted.

#### **13. What are Kafka Streams?**

Kafka Streams is a client library for building real-time stream processing applications using data stored in Kafka.

#### **14. What is Idempotent Producer?**

An Idempotent Producer ensures that messages are not duplicated during retries, maintaining exactly-once delivery semantics.

#### **15. What is the difference between Exactly Once, At Least Once, and At Most Once delivery semantics?**

- At Most Once: Messages may be lost but never redelivered.
- At Least Once: Messages are never lost but can be duplicated.
- Exactly Once: Messages are neither lost nor duplicated.

#### **16. What is Kafka Connect?**

Kafka Connect is a framework for integrating Kafka with external systems like databases, file systems, or APIs.

#### **17. What is the significance of replication in Kafka?**

Replication ensures data availability and fault tolerance. Each partition can have multiple replicas across brokers.

#### **18. What is the Leader and Follower concept in Kafka?**

Each partition has one leader and multiple followers. The leader handles all reads/writes, while followers replicate data from the leader.

#### **19. What is Kafka's log compaction feature?**

Log compaction ensures that only the latest value for each key is retained, useful for maintaining the latest state of a dataset.

#### **20. How does Kafka ensure fault tolerance?**

Kafka ensures fault tolerance via data replication, leader election, and commit log persistence across distributed brokers.