

Confluent Kafka — Basic Interview Q&A

1) What is Apache Kafka?

Kafka is a **distributed commit log** for high-throughput, fault-tolerant, real-time event streaming. It stores records in **topics** partitioned across brokers for parallelism and scalability.

2) Topic, Partition, Offset

- **Topic**: Named stream of records.
- **Partition**: Ordered, append-only log; unit of parallelism.
- **Offset**: Monotonically increasing position within a partition.

3) Ordering Guarantees

Kafka guarantees ordering **within a partition**. Use a **key** to ensure the same key routes to the same partition.

4) Consumer Group

A group of consumers sharing a group id; partitions are assigned across consumers for horizontal scaling while each partition is processed by **one** consumer in the group.

5) Durability & Fault Tolerance

Replication across brokers; producers can require `acks=all`. Consumers commit offsets to track progress.

6) Log Retention vs Compaction

- **Retention**: Delete records older than a configured time/size.
- **Compaction**: Retain only the latest record per key (and tombstones) to build **KTable**-like history.