Confluent / Kafka Interview Questions & Answers - 2025 Edition

Q: What is Apache Kafka?

A: Kafka is a distributed streaming platform that handles real-time data feeds. It allows producers to send messages to topics and consumers to read them.

Q: What is Confluent?

A: Confluent is the enterprise version of Kafka that adds features like Schema Registry, connectors, ksqlDB, and better management tools.

Q: What are producers, consumers, and topics?

A: Producers send data to topics, topics organize messages, and consumers read messages from those topics.

Q: What are partitions in Kafka?

A: Partitions divide topics into segments for parallel processing. Each message in a partition has an offset for order tracking.

Q: What is an offset in Kafka?

A: An offset is a unique identifier for a message within a partition. Consumers use it to track their read progress.

Q: What are brokers in Kafka?

A: Brokers are servers in a Kafka cluster that store and serve data. Each broker handles partitions and replication.

Q: What is a consumer group?

A: A consumer group is a set of consumers that share the workload of reading messages from topics, ensuring each partition is read by only one consumer in the group.

Q: What is message replication in Kafka?

A: Kafka replicates partitions across multiple brokers for fault tolerance. Replication ensures data durability in case of broker failure.

Q: What is leader and follower in Kafka?

A: Each partition has one leader (handles reads/writes) and followers (replicas that sync data for fault tolerance).

Q: What is exactly-once semantics (EOS)?

A: It ensures that each message is processed exactly once, avoiding duplicates or missed messages in complex pipelines.

Q: What is Kafka Connect?

A: Kafka Connect is a framework for connecting Kafka to external systems using pre-built connectors for data ingestion and delivery.

Q: What is ksqlDB?

A: ksqlDB enables SQL-like queries over Kafka streams for real-time analytics and data transformation.

Q: How does Confluent Schema Registry help?

A: It manages schemas for data serialization formats like Avro or JSON, ensuring producers and consumers are schema-compatible.

Q: What is consumer lag?

A: Consumer lag is the difference between the latest message offset and the consumer's current offset, indicating delayed consumption.

Q: What happens if a Kafka broker fails?

A: If a broker fails, partition leaders are re-elected from replicas automatically, ensuring continuity without data loss.

Q: What is idempotent producer in Kafka?

A: An idempotent producer ensures messages are not duplicated during retries by assigning unique sequence numbers to messages.

Q: What are Kafka Streams?

A: Kafka Streams is a Java library for building real-time processing applications directly on top of Kafka topics.

Q: How does Kafka integrate with Snowflake?

A: Using the Snowflake Kafka Connector, data streams from Kafka topics directly into Snowflake tables for near-real-time analytics.

Q: What are batch vs streaming systems?

A: Batch systems (like ETL) process data periodically, while streaming systems like Kafka handle continuous data in real time.

Q: What are the advantages of using Confluent over open-source Kafka?

A: Confluent offers enterprise-grade connectors, improved monitoring, governance, and security features, reducing setup and maintenance complexity.