Top 14 Kafka Interview Questions & Answers

1) Mention what is Apache Kafka?

Apache Kafka is a publish-subscribe messaging system developed by Apache written in Scala. It is a distributed, partitioned and replicated log service.

2) Mention what is the traditional method of message transfer?

The traditional method of message transfer includes two methods

- **Queuing:** In a queuing, a pool of consumers may read message from the server and each message goes to one of them
- **Publish-Subscribe:** In this model, messages are broadcasted to all consumers
Kafka caters single consumer abstraction that generalized both of the above- the consumer group.

3) Mention what is the benefits of Apache Kafka over the traditional technique?

Apache Kafka has following benefits above traditional messaging technique

- **Fast:** A single Kafka broker can serve thousands of clients by handling megabytes of reads and writes per second
- **Scalable**: Data are partitioned and streamlined over a cluster of machines to enable larger data
- **Durable:** Messages are persistent and is replicated within the cluster to prevent data loss
- **Distributed by Design:** It provides fault tolerance guarantees and durability

4) Mention what is the meaning of broker in Kafka?

In Kafka cluster, broker term is used to refer Server.

5) Mention what is the maximum size of the message does Kafka server can receive?

The maximum size of the message that Kafka server can receive is 1000000 bytes.

6) Explain what is Zookeeper in Kafka? Can we use Kafka without Zookeeper?

Zookeeper is an open source, high-performance co-ordination service used for distributed applications adapted by Kafka.

No, it is not possible to bye-pass Zookeeper and connect straight to the Kafka broker. Once the Zookeeper is down, it cannot serve client request.

- Zookeeper is basically used to communicate between different nodes in a cluster

- In Kafka, it is used to commit offset, so if node fails in any case it can be retrieved from the previously committed offset
- Apart from this it also does other activities like leader detection, distributed synchronization, configuration management, identifies when a new node leaves or joins, the cluster, node status in real time, etc.

7) Explain how message is consumed by consumer in Kafka?

Transfer of messages in Kafka is done by using sendfile API. It enables the transfer of bytes from the socket to disk via kernel space saving copies and call between kernel user back to the kernel.

8) Explain how you can improve the throughput of a remote consumer?

If the consumer is located in a different data center from the broker, you may require to tune the socket buffer size to amortize the long network latency.

9) Explain how you can get exactly once messaging from Kafka during data production?

During data, production to get exactly once messaging from Kafka you have to follow two things **avoiding duplicates during data consumption** and **avoiding duplication during data production.**

Here are the two ways to get exactly one semantics while data production:

- Avail a single writer per partition, every time you get a network error checks the last message in that partition to see if your last write succeeded
- In the message include a primary key (UUID or something) and de-duplicate on the consumer

10) Explain how you can reduce churn in ISR? When does broker leave the ISR?

ISR is a set of message replicas that are completely synced up with the leaders, in other word ISR has all messages that are committed. ISR should always include all replicas until there is a real failure. A replica will be dropped out of ISR if it deviates from the leader.

11) Why replication is required in Kafka?

Replication of message in Kafka ensures that any published message does not lose and can be consumed in case of machine error, program error or more common software upgrades.

12) What does it indicate if replica stays out of ISR for a long time?

If a replica remains out of ISR for an extended time, it indicates that the follower is unable to fetch data as fast as data accumulated at the leader.

13) Mention what happens if the preferred replica is not in the ISR?

If the preferred replica is not in the ISR, the controller will fail to move leadership to the preferred replica.

14) Is it possible to get the message offset after producing?

You cannot do that from a class that behaves as a producer like in most queue systems, its role is to fire and forget the messages. The broker will do the rest of the work like appropriate metadata handling with id's, offsets, etc.

As a consumer of the message, you can get the offset from a Kafka broker. If you gaze in the **SimpleConsumer** class, you will notice it fetches **MultiFetchResponse** objects that include offsets as a list. In addition to that, when you iterate the Kafka Message, you will have **MessageAndOffset** objects that include both, the offset and the message sent.

Guru99 Provides FREE ONLINE TUTORIAL   on Various courses like

Java	MIS	MongoDB	BigData	Cassandra

Web Services	SQLite	JSP	Informatica	Accounting

SAP Training	Python	Excel	ASP Net	HBase

Project Management	Test Management	Business Analyst	Ethical Hacking	PMP

Live Project	SoapUI	Photoshop	Manual Testing	Mobile Testing

Data Warehouse	R Tutorial	Tableau	DevOps	AWS

Jenkins	Agile Testing	RPA	JUnit	Software Engineering

Selenium	CCNA	AngularJS	NodeJS	PLSQL