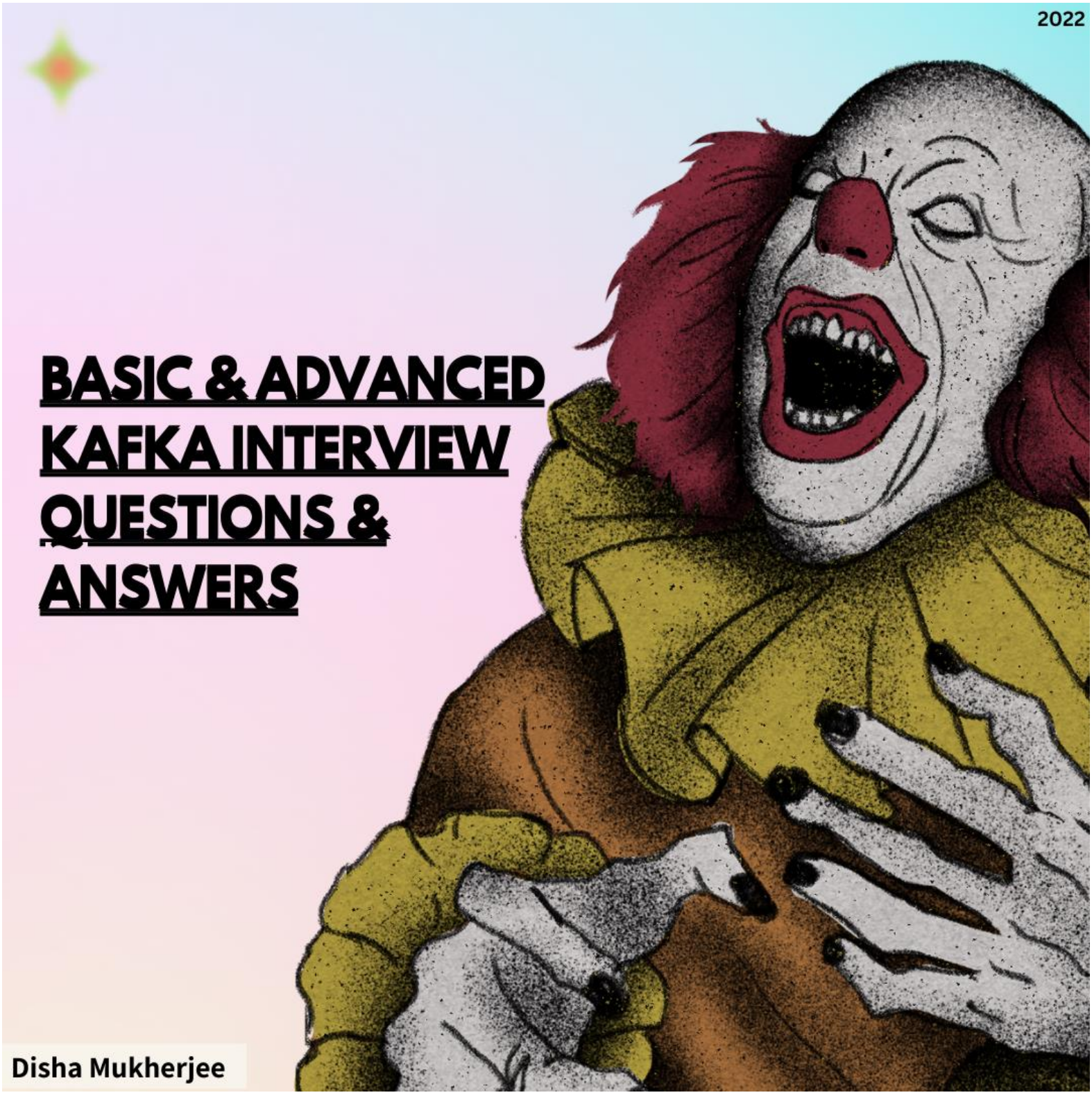

Source—Simplilearn
@Dishamukherjee

The questions have been divided into two parts: Basic and Advanced Kafka Interview Questions.

Basic Kafka Interview Questions

Let us begin with the basic Kafka interview questions!

1. What is the role of the offset?

In partitions, messages are assigned a unique ID number called the offset. The role is to identify each message in the partition uniquely.

2. Can Kafka be used without ZooKeeper?

It is not possible to connect directly to the Kafka Server by bypassing ZooKeeper. Any client request cannot be serviced if ZooKeeper is down.

3. In Kafka, why are replications critical?

Replications are critical as they ensure published messages can be consumed in the event of any program error or machine error and are not lost.

4. What is a partitioning key?

Ans. The partitioning key indicates the destination partition of the message within the producer. A hashing based partitioner determines the partition ID when the key is given.

5. What is the critical difference between Flume and Kafka?

Kafka ensures more durability and is scalable even though both are used for real-time processing.

6. When does QueueFullException occur in the producer?

QueueFullException occurs when the producer attempts to send messages at a pace not handleable by the broker.

7. What is a partition of a topic in Kafka Cluster?

Partition is a single piece of Kafka topic. More partitions allow excellent parallelism when reading from the topics. The number of partitions is configured based on per topic.

8. Explain Geo-replication in Kafka.

The Kafka MirrorMaker provides Geo-replication support for clusters. The messages are replicated across multiple cloud regions or datacenters. This can be used in passive/active scenarios for recovery and backup.

9. What do you mean by ISR in Kafka environment?

ISR is the abbreviation of In sync replicas. They are a set of message replicas that are synced to be leaders.

10. How can you get precisely one messaging during data production?

To get precisely one messaging from data production, you have to follow two things avoiding duplicates during data production and avoiding duplicates during data consumption. For this, include a primary key in the message and de-duplicate on the consumer.

11. How do consumers consumes messages in Kafka?

The transfer of messages is done in Kafka by making use of send file API. The transfer of bytes occurs using this file through the kernel-space and the calls between back to the kernel and kernel user.

12. What is Zookeeper in Kafka?

Source—Simplilearn
@Dishamukherjee

One of the basic Kafka interview questions is about Zookeeper. It is a high performance and open source complete coordination service used for distributed applications adapted by Kafka. It lets Kafka manage sources properly.

13. What is a replica in the Kafka environment?

The replica is a list of essential nodes needed for logging for any particular partition. It can play the role of a follower or leader.

14. What does follower and leader in Kafka mean?

Partitions are created in Kafka based on consumer groups and offset. One server in the partition serves as the leader, and one or more servers act as a follower. The leader assigns itself tasks that read and write partition requests. Followers follow the leader and replicate what is being told.

15. Name various components of Kafka.

The main components are:

1. Producer – produces messages and can communicate to a specific topic
2. Topic: a bunch of messages that come under the same topic
3. Consumer: One who consumes the published data and subscribes to different topics
4. Brokers: act as a channel between consumers and producers.

16. Why is Kafka so popular?

Kafka acts as the central nervous system that makes streaming data available to applications. It builds real-time data pipelines responsible for data processing and transferring between different systems that need to use it.

17. What are consumers in Kafka?

Kafka tags itself with a user group, and every communication on the topic is distributed to one use case. Kafka provides a single-customer abstraction that discovers both publish-subscribe consumer group and queuing.

Source—Simplilearn
@Dishamukherjee

18. What is a consumer group?

When more than one consumer consumes a bunch of subscribed topics jointly, it forms a consumer group.

19. How is a Kafka Server started?

To start a Kafka Server, the Zookeeper has to be powered up by using the following steps:

> bin/zookeeper-server-start.sh config/zookeeper.properties

> bin/kafka-server-start.sh config/server.properties

20. How does Kafka work?

Kafka combines two messaging models, queues them, publishes, and subscribes to be made accessible to several consumer instances.

21. What are replications dangerous in Kafka?

This is because duplication assures that issued messages are absorbed in plan fault, appliance mistake or recurrent software promotions.

22. What is the role of Kafka Producer API play?

It covers two producers: kafka.producer.async.AsyncProducer and kafka.producer.SyncProducer. The API provides all producer performance through a single API to its clients.

23. Discuss the architecture of Kafka.

A cluster in Kafka contains multiple brokers as the system is distributed. The topic in the system is divided into multiple partitions. Each broker stores one or multiple partitions so that consumers and producers can retrieve and publish messages simultaneously.

24. What advantages does Kafka have over Flume?

Source—Simplilearn
@Dishamukherjee

Kafka is not explicitly developed for Hadoop. Using it for writing and reading data is trickier than it is with Flume. However, Kafka is a highly reliable and scalable system used to connect multiple systems like Hadoop.

25. Why are the benefits of using Kafka?

Kafka has the following advantages:

1. Scalable- Data is streamlined over a cluster of machines and partitioned to enable large information.
2. Fast- Kafka has brokers which can serve thousands of clients
3. Durable- message is replicated in the cluster to prevent record loss.
4. Distributed- provides robustness and fault tolerance.

Advanced Kafka Interview Questions

In the next section let us have a look at the advanced Kafka interview questions.

1. Is getting message offset possible after producing?

This is not possible from a class behaving as a producer because, like in most queue systems, its role is to forget and fire the messages. As a message consumer, you get the offset from a Kaka broker.

2. How can the Kafka cluster be rebalanced?

When a customer adds new disks or nodes to existing nodes, partitions are not automatically balanced. If several nodes in a topic are already equal to the replication factor, adding disks will not help in rebalancing. Instead, the Kafka-reassign-partitions command is recommended after adding new hosts.

3. How does Kafka communicate with servers and clients?

The communication between the clients and servers is done with a high-performance, simple, language-agnostic TCP protocol. This protocol maintains backwards compatibility with the earlier version.

4. How is the log cleaner configured?

It is enabled by default and starts the pool of cleaner threads. For enabling log cleaning on particular topic, add: log.cleanup.policy=compact. This can be done either by using alter topic command or at topic creation time.

5. What are the three broker configuration files?

The essential configuration files are broker.id, log.dirs, zookeeper.connect.

6. What are the traditional methods of message transfer?

The traditional method includes:

1. Queuing- a pool of consumers read a message from the server, and each message goes to one of the consumers.
2. Publish-subscribe: Messages are broadcasted to all consumers.

7. What is a broker in Kafka?

The broker term is used to refer to Server in Kafka cluster.

8. What maximum message size can the Kafka server receive?

The maximum message size that Kafka server can receive is 10 lakh bytes.

9. How can the throughput of a remote consumer be improved?

If the consumer is not located in the same data center as the broker, it requires tuning the socket buffer size to amortize the long network latency.

10. How can churn be reduced in ISR, and when does the broker leave it?

Source—Simplilearn
@Dishamukherjee

ISR has all the committed messages. It should have all replicas till there is a real failure. A replica is dropped out of ISR if it deviates from the leader.

11. If replica stays out of ISR for a long time, what is indicated?

If a replica is staying out of ISR for a long time, it indicates the follower cannot fetch data as fast as data is accumulated at the leader.

12. What happens if the preferred replica is not in the ISR?

The controller will fail to move leadership to the preferred replica if it is not in the ISR.

13. What is meant by SerDes?

SerDes (Serializer and Deserializer) materializes the data whenever necessary for any Kafka stream when SerDes is provided for all record and record values.

14. What do you understand by multi-tenancy?

This is one of the most asked advanced Kafka interview questions. Kafka can be deployed as a multi-tenant solution. The configuration for different topics on which data is to be consumed or produced is enabled.

15. How is Kafka tuned for optimal performance?

To tune Kafka, it is essential to tune different components first. This includes tuning Kafka producers, brokers and consumers.

16. What are the benefits of creating Kafka Cluster?

When we expand the cluster, the Kafka cluster has zero downtime. The cluster manages the replication and persistence of message data. The cluster also offers strong durability because of cluster centric design.

17. Who is the producer in Kafka?

Source—Simplilearn
@Dishamukherjee

The producer is a client who publishes and sends the record. The producer sends data to the broker service. The producer applications write data to topics that are ready by consumer applications.

18. Tell us the cases where Kafka does not fit.

Kafka ecosystem is a bit difficult to configure, and one needs implementation knowledge. It does not fit in situations where there is a lack of monitoring tool, and a wildcard option is not available to select topics.

19. What is the consumer lag?

Ans Reads in Kafka lag behind Writes as there is always some delay between writing and consuming the message. This delta between the consuming offset and the latest offset is called consumer lag.

20. What do you know about Kafka Mirror Maker?

Kafka Mirror Maker is a utility that helps in replicating data between two Kafka clusters within the different or identical data centres.

21. What is fault tolerance?

In Kafka, data is stored across multiple nodes in the cluster. There is a high probability of one of the nodes failing. Fault tolerance means that the system is protected and available even when nodes in the cluster fail.

22. What is Kafka producer Acknowledgement?

An acknowledgement or ack is sent to the producer by a broker to acknowledge receipt of the message. Ack level defines the number of acknowledgements that the producer requires before considering a request complete.

23. What is load balancing?

The load balancer distributes loads across multiple systems in caseload gets increased by replicating messages on different systems.

Source—Simplilearn
@Dishamukherjee

24. What is a Smart producer/ dumb broker?

A smart producer/dumb broker is a broker that does not attempt to track which messages have been read by consumers. It only retains unread messages.

25. What is meant by partition offset?

The offset uniquely identifies a record within a partition. Topics can have multiple partition logs that allow consumers to read in parallel. Consumers can read messages from a specific as well as an offset print of their choice.

Source—Simplilearn
@Dishamukherjee