46 Kafka interview questions to hire top developers

Questions

1. Can you explain what Apache Kafka is and its primary use cases?
2. What are the main components of a Kafka ecosystem?
3. How does Kafka achieve high throughput and low latency?
4. What is the role of ZooKeeper in a Kafka setup?
5. Can you describe the difference between a Kafka topic and a partition?
6. What is the purpose of consumer groups in Kafka?
7. How does Kafka handle message retention?
8. What is the significance of the offset in Kafka?
9. Can you explain the concept of replication in Kafka?
10. What are some common Kafka monitoring metrics you're familiar with?
11. How would you explain the role of a Kafka broker?
12. Can you describe what a Kafka producer does?
13. What is a Kafka consumer, and what is its primary function?
14. How would you explain the concept of a Kafka stream?
15. What are the key benefits of using Kafka in a data pipeline?
16. How do Kafka partitions contribute to scalability?
17. Can you explain what exactly an offset is in Kafka?
18. What is the difference between Kafka and traditional messaging systems?
19. How does Kafka ensure message order within a partition?
20. Can you describe how you would handle schema evolution in Kafka?
21. What is the role of a Kafka Connect and how does it differ from a Kafka producer?
22. How would you approach error handling in a Kafka consumer application?
23. What strategies would you implement for ensuring message processing reliability?
24. Can you explain the difference between at-least-once and exactly-once delivery semantics?
25. What are the implications of using compacted topics in Kafka?
26. How would you implement a message processing pipeline using Kafka Streams?
27. What techniques can be used to optimize Kafka consumer performance?
28. How do you monitor and troubleshoot lag in Kafka consumers?
29. Can you explain the concept of idempotence in Kafka producers?
30. What is the significance of the consumer lag metric in Kafka?
31. How would you implement security measures in a Kafka setup?
32. Can you describe a situation where you had to scale Kafka, and what steps you took?
33. How would you integrate Kafka with other big data tools or platforms?
34. How does Kafka handle network partitions in a distributed environment?
35. Can you explain the concept of eventual consistency in the context of Kafka?
36. How does Kafka ensure fault tolerance in a distributed setup?
37. What are the challenges of implementing exactly-once semantics in a distributed streaming system like Kafka?
38. How does Kafka handle data consistency across multiple data centers?
39. Explain the concept of log compaction in Kafka and its importance in distributed systems.
40. How would you handle a situation where a Kafka cluster begins to experience high latency issues?
41. Imagine you have a Kafka topic with a high rate of message drops. How would you approach solving this issue?
42. How would you manage a scenario where you need to migrate data from one Kafka cluster to another with minimal downtime?
43. Suppose you encounter a Kafka topic with unevenly distributed partitions. How would you address this issue?
44. How would you tackle a situation where a Kafka consumer group is experiencing frequent rebalances?
45. What steps would you take if you discover a critical security vulnerability in your Kafka setup?
46. How would you approach optimizing the performance of a Kafka-based data pipeline that processes large volumes of data?