Title: ELK, Splunk & Kafka Interview Questions and Answers

Introduction: This document contains 50+ interview questions and answers covering ELK (Elasticsearch, Logstash, Kibana), Splunk, and Kafka integration scenarios. Useful for software engineers, DevOps, and observability professionals preparing for interviews.

ELK Stack Questions

Basic

1. **What is ELK Stack?**
2. ELK Stack consists of Elasticsearch, Logstash, and Kibana for log ingestion, storage, and visualization.
3. **Explain Elasticsearch.**
4. Elasticsearch is a distributed search and analytics engine that stores logs as JSON documents.
5. **What is Logstash?**
6. Logstash is a data collection and processing pipeline that ingests logs from multiple sources, transforms them, and sends them to Elasticsearch.
7. **What is Kibana?**
8. Kibana is a visualization tool to create dashboards and perform analytics on Elasticsearch data.
9. **What are Beats?**
10. Beats are lightweight agents (Filebeat, Metricbeat, etc.) used to ship logs/metrics to Logstash or Elasticsearch.

Intermediate

1. **What is an Elasticsearch index?**
2. An index is like a database in Elasticsearch storing documents.
3. **Difference between Elasticsearch and Logstash?**
4. Elasticsearch stores and searches data, Logstash processes and forwards data.
5. **Explain index lifecycle management (ILM).**
6. ILM automates index rollover, retention, and deletion policies.
7. **How does ELK scale for large data?**

8. Use clusters with master, data, and client nodes, and optionally a buffering system like Kafka.

9. **What is the difference between ELK and EFK?**

10. EFK uses Fluentd instead of Logstash for log shipping.

Advanced

1. **How to handle log transformations in Logstash?**
2. Using filters like grok, mutate, date, kv, and dissect.

3. **How to integrate Kafka with ELK?**

4. Kafka acts as a buffer. Logstash consumes Kafka topics and indexes data into Elasticsearch.

5. **How to secure ELK stack?**

6. Use X-Pack for authentication, encryption, and role-based access control.

7. **Difference between ELK open-source and Elastic Cloud?**

8. Elastic Cloud is managed SaaS, open-source ELK is self-hosted.

9. **How to monitor ELK performance?**

10. Using Kibana monitoring, cluster health APIs, and node statistics.

Splunk Questions

Basic

1. **What is Splunk?**
2. Splunk is a platform for collecting, indexing, and analyzing machine-generated data.

3. **Main components of Splunk?**

4. Forwarder, Indexer, Search Head, Deployment Server.

5. **What is HEC in Splunk?**

6. HTTP Event Collector allows apps to send JSON events directly to Splunk.

7. **What is SPL?**

8. Splunk Processing Language used for querying and analyzing logs.

9. **Types of Splunk licenses?**

10. Free, Enterprise, Cloud.

Intermediate

1. **Explain Splunk indexes.**
2. Indexes store logs with retention policies and enable fast searches.
3. **Difference between Universal Forwarder and Heavy Forwarder.**
4. UF is lightweight, forwards raw logs. HF can parse and filter before sending.
5. **How to monitor Splunk performance?**
6. Use Monitoring Console for indexer, search head, and forwarder metrics.
7. **Difference between Splunk Free and Enterprise.**
8. Free: 500 MB/day, no clustering. Enterprise: unlimited, supports clustering, advanced analytics.
9. **How to enable HEC?**
10. Settings → Data Inputs → HTTP Event Collector → Enable token.

Advanced

1. **How to integrate Kafka with Splunk?**
2. Use **Splunk Connect for Kafka** or HEC: Kafka consumers push logs to Splunk HEC.
3. **Explain Splunk clustering.**
4. Indexer clustering for HA, Search head clustering for scaling queries.
5. **How to handle high volume of logs in Splunk?**
6. Use indexer clustering, load balancing, and tokenized forwarders.
7. **Difference between Splunk Cloud and Enterprise.**
8. Cloud is SaaS-managed, Enterprise is on-premises.
9. **What is Splunk Machine Learning Toolkit?**
10. Built-in ML module for anomaly detection and predictive analytics.

Combined ELK & Splunk + Kafka Questions

1. **Why use Kafka between apps and log platforms?**

2. Kafka acts as a buffer for high-throughput log streams and decouples producers from consumers.

3. **How to send Spring Boot logs to both ELK and Splunk?**

4. Configure logback/log4j appenders for HEC (Splunk) and Logstash (ELK).

5. **What is the difference between Kafka → ELK and Kafka → Splunk setups?**

6. Kafka → ELK: Logstash consumes topics and indexes.
7. Kafka → Splunk: Splunk Kafka Connectors or custom consumers push to HEC.

8. **How to ensure data consistency between ELK and Splunk?**

9. Use the same Kafka topic as source, and idempotent message IDs where possible.

10. **How to monitor Kafka lag for log pipelines?**

11. Use Kafka consumer group offsets and tools like Burrow or Kafka Manager.

12. **What are the advantages of using both ELK and Splunk?**

13. ELK: open-source, cost-effective, flexible dashboards.
14. Splunk: enterprise-ready, SPL analytics, alerts, and ML.

15. **Example Kafka → ELK → Splunk flow:**

16. Spring Boot logs → Kafka topics → Logstash → Elasticsearch/Kibana dashboards → Splunk HEC for alerting.

17. **How to parse JSON logs in ELK and Splunk?**

18. ELK: Logstash `json` filter or `mutate`.
19. Splunk: HEC accepts JSON natively, configure `sourcetype=json`.

20. **How to handle schema evolution in logs?**

21. Kafka schemas via Schema Registry, ELK dynamic mappings, Splunk JSON events.

22. **How to aggregate metrics across ELK and Splunk?**

23. Use dashboards to combine logs from Elasticsearch and Splunk via APIs.

24. **Explain high availability in a Kafka + ELK + Splunk setup.**

25. Kafka cluster with replication, Elasticsearch cluster with master/data nodes, Splunk indexer cluster.

26. **How to troubleshoot missing logs in Splunk and ELK?**

27. Check forwarders, HEC tokens, Logstash pipelines, Kafka consumer offsets, and firewall rules.

28. **Explain backpressure handling.**

29. Kafka handles spikes; Logstash uses persistent queues; Splunk HEC throttles requests.

30. **Example of structured logging for Kafka + ELK + Splunk:**

```
{
  "timestamp":"2025-09-21T09:00:00Z",
  "level":"INFO",
  "service":"user-service",
  "message":"User created",
  "userId":1234
}
```

31. **How to secure Kafka → ELK → Splunk pipeline?**

32. SSL/TLS for Kafka and HEC, authentication for Elasticsearch, RBAC in Splunk, network policies.

33. **How to visualize combined logs in dashboards?**

34. ELK: Kibana dashboards.

35. Splunk: native Splunk dashboards or Webhooks from ELK.

36. **How to test Kafka → Splunk integration?**

37. Produce test JSON logs to Kafka, verify they appear in Splunk via HEC.

38. **Difference between Logstash and Splunk forwarder in Kafka pipelines.**

39. Logstash: can transform, enrich, buffer logs.

40. Splunk UF/HEC: lightweight forwarding, direct ingestion.

41. **How to perform log retention?**

42. ELK: Index lifecycle policies.

43. Splunk: retention by index or archive policies.

44. **Monitoring and alerting in combined setup**

45. Use Splunk alerts for critical events.
46. Kibana + Elasticsearch watches or OpenSearch alerts for custom dashboards.

Conclusion: This document provides a comprehensive set of questions and answers covering ELK, Splunk, and Kafka integration scenarios, suitable for both beginner and advanced interview preparation.