**ELK Stack Interview Questions and Answers**

**🔹 1. What is the ELK Stack?**

**Answer:** The ELK Stack is a collection of open-source tools:

* **Elasticsearch** – A distributed search and analytics engine.
* **Logstash** – A data processing pipeline that ingests data from multiple sources, transforms it, and sends it to a storage backend (like Elasticsearch).
* **Kibana** – A visualization layer for Elasticsearch data.

**🔹 2. What are the key features of Elasticsearch?**

**Answer:**

* Distributed and scalable
* Full-text search capabilities
* Near real-time search and analytics
* RESTful API interface
* Schema-less JSON documents
* Built-in clustering and replication

**🔹 3. How does Logstash process data?**

**Answer:** Logstash uses a pipeline with three stages:

* **Input**: Collects data (e.g., from beats, files, syslog)
* **Filter**: Transforms data (e.g., with Grok, mutate, date filters)
* **Output**: Sends processed data (e.g., to Elasticsearch)

**🔹 4. What is Kibana used for?**

**Answer:** Kibana is used to:

* Visualize Elasticsearch data
* Create dashboards and charts
* Explore logs and metrics
* Use Dev Tools for query debugging
* Set up alerts and manage index patterns

**🔹 5. What is a Grok filter in Logstash?**

**Answer:** Grok is a Logstash filter plugin used to parse unstructured log data into structured JSON by using regex-like patterns.  
**Example:**

ruby

CopyEdit

grok {

match => { "message" => "%{IP:client\_ip} - - \[%{HTTPDATE:timestamp}\] \"%{WORD:method} %{URIPATH:uri}" }

}

**🔹 6. How do Beats fit into the ELK Stack?**

**Answer:** Beats are lightweight data shippers:

* **Filebeat** – For forwarding and centralizing log files
* **Metricbeat** – For system and service metrics
* **Packetbeat, Winlogbeat, Auditbeat** – For network, Windows events, audit data respectively

**🔹 7. What is an Elasticsearch Index?**

**Answer:** An index in Elasticsearch is similar to a database in traditional RDBMS. It stores a collection of documents and has mappings to define field types.

**🔹 8. How does Elasticsearch achieve high availability?**

**Answer:**

* **Shards**: Split the index into smaller parts for horizontal scaling
* **Replicas**: Copies of primary shards for fault tolerance
* **Cluster Nodes**: Master, data, ingest, coordinating nodes for distributed workload

**🔹 9. Explain an Ingest Pipeline in Elasticsearch.**

**Answer:** An ingest pipeline consists of processors (like grok, set, rename, date) used to preprocess documents before indexing them.

**Example:**

json

CopyEdit

PUT \_ingest/pipeline/log\_pipeline

{

"processors": [

{

"grok": {

"field": "message",

"patterns": ["%{COMMONAPACHELOG}"]

}

}

]

}

**🔹 10. What is the difference between Kibana Discover and Dashboard?**

**Answer:**

* **Discover**: Explore raw log data interactively
* **Dashboard**: Visual layout containing multiple visualizations and saved searches for monitoring trends

**🔹 11. How do you secure an ELK Stack?**

**Answer:**

* Enable **TLS encryption** between components
* Use **X-Pack Security** for user authentication and role-based access
* Set up **API keys and token-based auth**
* Use **IP filtering and firewalls**

**🔹 12. What is the purpose of the \_bulk API in Elasticsearch?**

**Answer:** The \_bulk API allows batching multiple indexing or update operations into a single HTTP request to improve performance.

**🔹 13. What is a Kibana Index Pattern?**

**Answer:** An index pattern tells Kibana which Elasticsearch indices to explore (e.g., logstash-\*). It enables field discovery and visualizations.

**🔹 14. How can you monitor ELK Stack performance?**

**Answer:**

* Use **Monitoring UI** in Kibana (Stack Monitoring)
* Integrate with **Metricbeat** for node and cluster stats
* Monitor **Heap usage, disk I/O, CPU, index size**

**🔹 15. How do you handle different log formats in Logstash?**

**Answer:**

* Use **conditional filters**
* Combine multiple **Grok patterns** with fallbacks
* Use **mutate**, **json**, and **csv** filters to normalize data

**🔹 16. Can you write a simple Logstash config to read a file and send to Elasticsearch?**

**Answer:**

ruby

CopyEdit

input {

file {

path => "/var/log/syslog"

start\_position => "beginning"

}

}

filter {

grok {

match => { "message" => "%{SYSLOGTIMESTAMP:timestamp} %{WORD:program}" }

}

}

output {

elasticsearch {

hosts => ["http://localhost:9200"]

index => "syslog-index"

}

}

**🔹 17. What is the difference between a Master node and a Data node in Elasticsearch?**

**Answer:**

* **Master Node**: Manages cluster state and metadata (index creation, shard allocation)
* **Data Node**: Stores actual data and handles indexing/search operations

**🔹 18. How do you troubleshoot issues in the ELK stack?**

**Answer:**

* **Check logs** of each component
* Use Kibana **Dev Tools > Console** to query cluster health
* Use \_cat/indices, \_cluster/health, and \_nodes/stats
* Look for pipeline failures in Logstash logs

**🔹 19. What are index lifecycle policies in Elasticsearch?**

**Answer:** ILM (Index Lifecycle Management) automates index rollover, shrink, delete operations based on data age or size.

**🔹 20. What’s the difference between Query DSL’s match and term?**

**Answer:**

* match: Full-text search (analyzed fields)
* term: Exact value match (non-analyzed fields)