By OnlineInterviewQuestions.com

ElasticSearch Interview Questions

Elasticsearch is a real-time distributed, RESTful search and analytics engine built on the top of Apache Lucene which is a full-text search engine. you can see Elasticsearch as a distributed storage that features Real-time Analytics. It is document-oriented and stores object as document and make them indexable so the content of documents is searchable.

Few Known Facts about ElasticSearch

- **Built on Top of Lucene** (A full-text search engine by Apache)
- **Document-Oriented** (Stores data structured JSON documents)
- **Full-Text Search** (Supports Full-text search indexing which gives faster result retrieval)
- **Schema-Free** (Uses NoSQL)
- **Restful API** (Support Restful APIs for storage and retrieval of records)
- **Supports Autocompletion & Instant Search**

Following is the list of Top 25 ElasticSearch Interview questions with their answers

Q1. What is Elasticsearch?

Elasticsearch is a search engine that is based on Lucene.It offers a distributed, multitenant – capable full-text search engine with as HTTP (Hyper Text Transfer Protocol) web interface and Schema-free JSON (JavaScript Object Notation) documents.It is developed in Java and is an open source released under Apache License.

Q2. What is a current stable version of Elasticsearch?

As on Jan 2020, version **7.5** is the latest and stable version of Elasticsearch.

Q3. List the software requirements to install Elasticsearch?

Since Elasticsearch is built using Java, we require any of the following software to run Elasticsearch on our device.

- The latest version of Java 8 series
- Java version 1.8.0_131 is recommended.

Q4. How to start elastic search server?

Run Following command on your terminal to start Elasticsearch server:

```
cd elasticsearch
./bin/elasticsearch
```

curl ‘http://localhost:9200/?pretty’ command is used to check ElasticSearch server is running or not.

Q5. Can you list some companies that use Elasticsearch?

Some of the companies that use Elasticsearch along with Logstash and Kibana are:

- Wikipedia
- Netflix
- Accenture
- Stack Overflow
- Fujitsu
- Tripwire
- Medium
- Swat.io
- Hip chat
- IFTTT

Q6. What is a Cluster in Elasticsearch?

It is a set or a collection of one or more than one nodes or servers that hold your complete data and offers federated indexing and search capabilities across all the nodes.It is identified by a different and unique name that is “Elasticsearch” by default.
This name is considered to be important because a node can be a part of a cluster only if it is set up to join the cluster by its name.

Q7. What is a Node?

Each and every instance of Elasticsearch is a node.And, a collection of multiple nodes which can work in harmony form an Elasticsearch cluster.

Q8. What is an Index?

An index in Elasticsearch is similar to a table in relational databases.The only difference lies in storing the actual values in the relational database, whereas that is optional in Elasticsearch.
An index is capable of storing actual or analyzed values in an index.

Q9. What is a type of Elastic search?

A type in Elasticsearch is a logical category of the index whose semantics are completely up to the user.

Q10. Please Explain Mapping?

Mapping is a process which defines how a document is mapped to the search engine, searchable characteristics are included such as which fields are tokenized as well as searchable.
In Elasticsearch an index created may contain documents of all “mapping types”.

Q11. What is Document?

A document in Elasticsearch is similar to a row in relational databases.The only difference is that every document in an index can have a different structure or fields but having the same data type for common fields is mandatory.Each field with different data types can occur multiple times in a document.
The fields can also contain other documents.

Q12. What are SHARDS?

There are resource limitations like RAM, vCPU etc., for scale out, due to which applications employ multiple instances of Elasticsearch on separate machines.
Data in an index can be partitioned into multiple portions which are managed by a separate node or instance of Elasticsearch.Each such portion is called a Shard.And an Elasticsearch index has 5 shards by default.

Q13. What is REPLICAS?

Each shard in elastic search has again two copies of the shard that are called the replicas.
They serve the purpose of fault tolerance and high availability.

Q14. How to add or create an index in Elastic Search Cluster?

By using the command PUT before the index name, creates the index and if you want to add another index then use the command POST before the index name.
Ex: PUT website

An index named computer is created

Q15. How to delete an index in Elastic search?

To delete an index in Elasticsearch use the command DELETE /index name.
Ex: DELETE /website

Q16. How to list all indexes of a Cluster in ES.?

By using GET / _index name/ indices we can get the list of indices present in the cluster.

Q17. How to add a Mapping in an Index?

Basically, Elasticsearch will automatically create the mapping according to the data provided by the user in the request body. Its bulk functionality can be used to add more than one JSON object in the index.

Ex: POST website /_bulk

Q18. How can you retrieve a document by ID in ES.?

To retrieve a document in Elasticsearch, we use the GET verb followed by the _index, _type, _id.
Ex: GET / computer / blog / 123?=pretty

Q19. How relevancy and scoring is done in Elasticsearch?

The Boolean model is used by the Lucene to find the similar documents, and a formula called practical scoring function is used to calculate the relevance.
This formula copies concepts from the inverse document/term-document frequency and the vector space model and adds the modern features like coordination factor, field length normalization as well.
Score (q, d) is the relevance score of document "d" for query "q".

Q20. What are different ways of searching in Elasticsearch?

We can perform the following searches in Elasticsearch:

- **Multi-index, Multitype search:** All search APIs can be applied across all multiple indices with the support for the multi-index system.
  We can search certain tags across all indices as well as all across all indices and all types.
- **URI search:** A search request is executed purely using a URI by providing request parameters.
- **Request body search:**A search request can be executed by a search DSL, that includes the query DSL within the body.

Q21. List different types of queries supported by Elasticsearch?

The Queries are divided into two types with multiple queries categorized under them.

- **Full-text queries:** Match Query, Match phrase Query, Multi match Query, Match phrase prefix Query, common terms Query, Query string Query, simple Query String Query.
- **Term level queries:** term Query, term set Query, terms Query, Range Query, Prefix Query, wildcard Query, regexp Query, fuzzy Query, exists Query, type Query, ids Query.

Q22. What is the difference between Term-based and Full-text queries?

- **Term-based Queries :** Queries like the term query or fuzzy query are the low-level queries that do not have analysis phase.A term Query for the term Foo searches for the exact term in the inverted index and calculates the IDF/TF relevance score for every document that has a term.
- **Full-text Queries :** Queries like match query or query string queries are the high-level queries that understand that mapping of a field.As soon as the query assembles the complete list of items it executes

the appropriate low-level query for every term, and finally combines their results to produce the relevance score of every document.

Q23. How does aggregation work in Elasticsearch?

The aggregation framework provides aggregated data based on search query.It can be seen as a unit of work that builds analytic information over the set of documents.There are different types of aggregations with different purpose and outputs.

Q24. Where is Elasticsearch data stored?

Elasticsearch is a distributed documented store with several directories.It can store and retrieve the complex data structures that are serialized as JSON documents in real time.

Q25. Can Elasticsearch replace database?

Yes, Elasticsearch can be used as a replacement for a database as the Elasticsearch is very powerful.
It offers features like multitenancy, sharding and Replication, distribution and cloud Realtime get, Refresh, commit, versioning and re-indexing and many more, which make it an apt replacement of a database.

Q26. How to check elastic search server is running?

Generally, Elasticsearch uses the port range of 9200-9300.
So, to check if it is running on your server just type the URL of the homepage followed by the port number.

Ex: `mysitename.com:9200`

Q27. What are Analyzers in Elasticsearch?

Analyzer in Elasticsearch is a component that is used for search and analytics. An Analyzer component is used to split a text in multiple terms i.e token that is used to build the inverted index which exists inside Lucene for implementing full-text search.

Elasticsearch comes with a bunch of Built-in Analyzers, few of them are Simple Analyzer Standard Analyzer, Whitespace Analyzer, Language Analyzers, Fingerprint Analyzer, etc.