## ElasticSearch interview questions

### [1. How to check elastic search server is running?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled1)

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

### [2. Can you list some companies that use Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled2)

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

### [3. What is Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled3)

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](https://www.onlineinterviewquestions.com/json-interview-questions/) (JavaScript Object Notation) documents.It is developed in Java and is an open source released under Apache License.

### [4. How to start elastic search server?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled4)

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.

### [5. What is a type of Elastic search?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled5)

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

### [6. How to add a Mapping in an Index?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled6)

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

### [7. How relevancy and scoring is done in Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled7)

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”.

### [8. What is Document?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled8)

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.

### [9. What are SHARDS?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled9)

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.

### [10. What are different ways of searching in Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled10)

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.

### [11. What is the difference between Term-based and Full-text queries?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled11)

* 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.

### [12. How to add or create an index in Elastic Search Cluster?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled12)

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

### [13. How to list all indexes of a Cluster in ES.?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled13)

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

### [14. List different types of queries supported by Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled14)

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.

### [15. How can you retrieve a document by ID in ES.?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/#collapseUnfiled15)

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

## ElasticSearch interview questions

### [1. How does aggregation work in Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled1)

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.

### [2. What is an Index?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled2)

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.

### [3. Where is Elasticsearch data stored?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled3)

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.

### [4. Please Explain Mapping?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled4)

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”.

### [5. What is a Cluster in Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled5)

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.

### [6. How to delete an index in Elastic search?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled6)

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

Ex: DELETE /website

### [7. What is a Node?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled7)

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

### [8. Can Elasticsearch replace database?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled8)

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.

### [9. What is a current stable version of Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled9)

As on March 2018, the version 6.2.2 is the latest and stable version of Elasticsearch.

### [10. What is REPLICAS?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled10)

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.

### [11. List the software requirements to install Elasticsearch?](https://www.onlineinterviewquestions.com/elasticsearch-interview-questions/page/2/#collapseUnfiled11)

Since Elasticsearch is built using [Java](https://www.onlineinterviewquestions.com/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.

**Q. What is an index in ElasticSearch ?**   
**An index is similar to a table in relational databases. The difference is that relational databases would store actual values, which is optional in ElasticSearch. An index can store actual and/or analyzed values in an index.**

**Q. What is a document in ElasticSearch ?**   
**A document is similar to a row in relational databases. The difference is that each** document in an index can have a different structure (fields), but should have same data type for common fields.  
Each field can occur multiple times in a document with different data types. Fields can contain other documents too.

**Q. Does ElasticSearch have a schema ?**  
**Yes, ElasticSeach can have mappings which can be used to enforce schema on documents.**

**Q. What is a document type in ElasticSearch ?**  
**A document type can be seen as the document schema /** [dynamic mapping](https://mindmajix.com/elasticsearch/dynamic-mapping-overview) definition, which has the mapping of all the fields in the document along with its data types.

**Q. What is indexing in ElasticSearch ?**  
**The process of storing data in an index is called indexing in ElasticSearch. Data in ElasticSearch can be dividend into write-once and read-many segments. Whenever an update is attempted, a new version of the document is written to the index.**

**Q. What is a node in ElasticSearch ?**  
**Each instance of ElasticSearch is called a node. Multiple nodes can work in harmony to form an ElasticSearch Cluster.**

**Q.What is a shard in ElasticSearch ?**  
**Due to resource limitations like RAM, vCPU etc, for scale-out, applications need to employ multiple instances of ElasticSearch on separate machines. Data in an index can be divided into multiple partitions, each handled by a separate node (instance) of ElasticSearch. Each such partition is called a shard. By default an ElasticSearch index has 5 shards.**

**Q.What is a replica in ElasticSearch ?**  
**Each shard in ElasticSearch has 2 copy of the shard. These copies are called replicas. They serve the purpose of high-availability and fault-tolerance.**

[Check Out Elasticsearch Tutorials](https://mindmajix.com/elasticsearch)

**Q.What is an Analyzer in ElasticSearch ?**  
**While indexing data in ElasticSearch, data is transformed internally by the Analyzer defined for the index, and then indexed. An analyzer is built of tokenizer and filters. Following types of Analyzers are available in ElasticSearch 1.10.**  
**1. STANDARD ANALYZER**  
**2. SIMPLE ANALYZER**  
**3. WHITESPACE ANALYZER**  
**4. STOP ANALYZER**  
**5. KEYWORD ANALYZER**  
**6. PATTERN ANALYZER**  
**7. LANGUAGE ANALYZERS**  
**8. SNOWBALL ANALYZER**  
**9. CUSTOM ANALYZER**

**Q.What is a Tokenizer in ElasticSearch ?**  
**A Tokenizer breakdown fields values of a document into a stream, and inverted indexes are created and updates using these values, and these stream of values are stored in the document.**

**Q.What is a Filter in ElasticSearch ?**  
**After data is processed by Tokenizer, the same is processed by Filter, before indexing. Following types of Filters are available in ElasticSearch 1.10.**  
**1. AND FILTER**  
**2. BOOL FILTER**  
**3. EXISTS FILTER**  
**4. GEO BOUNDING BOX FILTER**  
**5. GEO DISTANCE FILTER**  
**6. GEO DISTANCE RANGE FILTER**  
**7. GEO POLYGON FILTER**  
**8. GEOSHAPE** FILTER  
9. GEOHASH CELL FILTER  
10. HAS CHILD FILTER  
11. HAS PARENT FILTER  
12. IDS FILTER  
13. INDICES FILTER  
14. LIMIT FILTER  
15. MATCH ALL FILTER  
16. MISSING FILTER  
17. NESTED FILTER  
18. NOT FILTER  
19. OR FILTER  
20. PREFIX FILTER  
21. QUERY FILTER  
22. RANGE FILTER  
23. REGEXP FILTER  
24. SCRIPT FILTER  
25. TERM FILTER  
26. TERMS FILTER  
27. TYPE FILTER

**Q.What is the query language of ElasticSearch ?**  
**ElasticSearch uses the Apache Lucene query language, which is called Query DSL.**

What is ElasticSearch?

Elasticsearch is a search engine based on Lucene. It has a distributed, multitenant-able full-text search engine. Elasticsearch is developed in [Java](https://tekslate.com/tutorials/java-tutorials/) and is released as open source under the terms of the Apache License.

What is the use of attributes- enabled, index and store?

The enabled attribute is applicable to several ElasticSearch created fields like \_index and \_size.

Store implies the data stored by Lucene, which will again return when necessary. Stored fields are not searchable.

The index is employed for searching. Indexed fields are transformed during analysis, and cannot retrieve the original data when necessary.

What is an Analyzer in ElasticSearch?

While indexing data, it is transformed internally via the defined Analyzer for the index.

Analyzers are made of one Tokenizer, preceded by CharFilters and zero or many TokenFilters. On the other hand, analysis module refers Analyzers under the name of mapping definitions or any APIs.

Elasticsearch is prebuilt with analyzers that are ready to use. However, you can integrate the built in character, token filters, along with tokenizers to create custom analyzers.

What is Character Filter in Elasticsearch Analyzer?

A character filter obtains the ideal text as stream of characters, later on modifies it by adding, deleting, or altering characters. For example, any character filter in usage has the ability to convert Hindu-Arabic numerals (٠‎١٢٣٤٥٦٧٨‎٩‎) into Arabic-Latin numerals (0123456789), and even sometimes strip [HTML elements](https://tekslate.com/html-formatting-elements-tags/) via the stream.

What is Token filters in Elasticsearch Analyzer?

A token filter obtains the token stream, later on add, delete, or alter the tokens. For instance, a lowercase token filter modifies all tokens into lowercase, a stop token filter deletes stop words, and a synonym token filter includes synonyms into the token stream.

Token filters will be unable to change the position or character offsets of any certain token.

What is a Tokenizer?

Tokenizers break down a string into stream of tokens. A single tokenizer split the string into terms when working with punctuation and whitespace. Elasticsearch has a number of built in tokenizers which can be used to build custom analyzers.

What is a Filter?

After Tokenizer ends the process of data, the same is carried by Filter.

Certain types of Filters available in ElasticSearch 1.10, are.

* AND FILTER
* EXISTS FILTER
* GEO DISTANCE FILTER
* GEO POLYGON FILTER
* GEOHASH CELL FILTER
* HAS PARENT FILTER
* INDICES FILTER
* MATCH ALL FILTER
* NESTED FILTER
* OR FILTER
* QUERY FILTER
* REGEXP FILTER
* TERM FILTER
* TYPE FILTER

What are the advantages of Elasticsearch?

* Elasticsearch is compatible on any platform.
* Elasticsearch is Near Real Time (NRT), making it searchable on engine.
* Elasticsearch cluster is distributed, scalable and easy to integrate.
* Elasticsearch REST uses JSON objects, making it to invoke the Elasticsearch server along with different programming languages.
* Elasticsearch supports every document type except text rendering.

What is Elasticsearch REST API and use of it?

Elasticsearch provides a very comprehensive and powerful REST API that you can use to interact with your cluster. Among the few things that can be done with the API are as follows:

* Check your cluster, node, and index health, status, and statistics
* Administer your cluster, node, and index data and metadata
* Perform CRUD (Create, Read, Update, and Delete) and search operations against your indexes
* Execute advanced search operations viz. aggregations, filtering, paging, scripting, sorting, among many others

What are the Disadvantages of Elasticsearch?

* Elasticsearch does not support multiple languages while handling request and response data in JSON.
* In rare cases, it has a problem of Split Brain situations.

Does ElasticSearch have a schema?

Yes, [Elasticsearch](https://tekslate.com/elasticsearch-training) can have a schema. A schema is a description of one or more fields that describes the document type and how to handle the different fields of a document. The schema in Elasticsearch is a mapping that emphasizes the JSON document fields and other data type, as well as Lucene indexes under the hood. Because of this, in Elasticsearch terms, we usually call this schema a “mapping”.

What is a cluster in ElasticSearch?

Cluster is a collection of nodes that holds data together and enables indexing and search abilities across each. Each cluster is recognized by a unique default name i.e. "elasticsearch". This name is important because a node can only be part of a cluster if the node is set up to join the cluster by its name.

What is a node in ElasticSearch?

Node is a minute server and forms a part of the cluster. It stores the data and enjoys the clusters indexing and search functionalities.

What is Ingest Node in Elasticsearch?

Ingest nodes can execute pre-processing an ingest pipeline. It effectively transform and works on the document prior to indexing. Dedicated ingest nodes mark the master and data nodes either as false or true.

What is Elasticsearch Data Node?

Data nodes hold shards that handle indexed documents. They execute data related CRUD and search aggregation operations etc. Set node.data=true to make node as Data Node.

Data Node operations are I/O-, memory-, and CPU-intensive. Data nodes benefit the separation of the master and data roles.

What is Master Node and Master Eligible Node in Elasticsearch?

Master Node control cluster-wide operations like to create or remove an index, track nodes of the [cluster](https://en.wikipedia.org/wiki/Cluster), and decide to allocate shards on nodes. It is important for cluster health to have a stable master node. Master Node elected based on configuration properties node.master=true (Default).

Master Eligible Node decide based on below configuration

discovery.zen.minimum\_master\_node : number (default 1)

and above number decide based (master\_eligible\_nodes / 2) + 1

What is Tribe Node and Coordinating Node in Elasticsearch?

Tribe node connect variant clusters and execute search operations across each connected clusters. This node is configured by settings *tribe*.

Coordinating Node is just like a Smart Load balancer that handle master duties, to hold data, and pre-process documents, then you are left with a coordinating node that can only route requests, handle the search reduce phase, and distribute bulk indexing.

Every node can be termed as a coordinating node which has all three node.data, node.ingest and node.master, set to false. This node is impossible to disable as it possess enough memory and CPU to deal with the gather phase.

What is an index in ElasticSearch?

Index is a ‘database’ within relational database. Its mapping defines multiple types and maps to one or many primary shards and can have zero or many replica shards.

MySQL => Databases

ElasticSearch => Indices

What is inverted index in Elasticsearch?

Inverted Index is backbone of Elasticsearch which make full-text search fast. Inverted index consists of a list of all unique words that occurs in documents and for each word, maintain a list of documents number and positions in which it appears.

For Example: There are two documents and having content as:

1: FacingIssuesOnIT is for ELK.

2: If ELK check FacingIssuesOnIT.

To make inverted index each document will split in words (also called as terms or token) and create below sorted index .

|  |  |  |
| --- | --- | --- |
| Term | Doc\_1 | Doc\_2 |
| FacingIssuesOnIT | X | X |
| Is | X |  |
| For | X |  |
| ELK | X | X |
| If |  | X |
| Check |  | X |

Now when we do some full-text search for String will sort documents based on existence and occurrence of matching counts.

Usually in Books we have inverted indexes on last pages. Based on the word we can thus find the page on which the word exists.

What is a shard?

Different applications need to employ multiple ElasticSearch instances on separate machines. Data in every index is divided into multiple partitions, each controlled by a separate ElasticSearch instance. Each such partition is termed as shard. By default, each ElasticSearch index possess 5 shards.

ElasticSearch Online Training

What is a replica?

Each shard has 2 copies called replicas. They are highly-available and fault-tolerant.

What is a document in ElasticSearch?

Document is similar to a row in relational databases. Each document in the index possess different structure, but has same data type for mutual fields.

MySQL => Databases => Tables => Columns/Rows

ElasticSearch => Indices => Types => Documents with Properties

What are the basic operations you can perform on a document?

The following operations can be performed on documents

* INDEXING A DOCUMENT USING ELASTICSEARCH.
* FETCHING DOCUMENTS USING ELASTICSEARCH.
* UPDATING DOCUMENTS USING ELASTICSEARCH.
* DELETING DOCUMENTS USING ELASTICSEARCH.

What is a type in ElasticSearch?

Type is a logical index partition whose semantics are entirely upon the user.

What are common area of use Elasticsearch?

It’s useful in application where need to do analysis, statics and need to find out anomalies on data based on pattern.

It’s useful where need to send alerts when particular condition matched like stock market, exception from logs etc.

It’s useful with application where log analysis and issue solution provide because of full search in billions of records in milliseconds.

It’s compatible with application like Filebeat, Logstash and Kibana for storage of high volume data for analysis and visualize in form of chart and dashboards.

Define Analyzer in ElasticSearch?

In ElasticSearch, Data is transformed while indexing internally by the analyzer specifically defined for the index and then indexed. Analyzers are built of filters and tokenizes. The major types of analyzers available in ElasticSearch 1.10 are as follows:

* simple analyzer
* standard analyzer
* keyword analyzer
* language analyzers
* snowball analyzer
* custom analyzer
* pattern analyzer
* whitespace analyzer
* stop analyzer

What is the query language of Elasticsearch?

Apache Lucene query language which is also called as Query DSL is used by Elasticsearch.

What is Elasticsearch?

Answers:

Elasticsearch is a search engine based on Apache Lucene that supports full-text search engine with scheme free JSON objects and HTTP web interface. This is s free and open source project developed in Java and licensed under Apache License terms. The key components of Elasticsearch are Node, Cluster, Index, Type, Document, Shard and Replicas. Elastic search has the capacity to perform a fast incisive search over large chunks of data.

Elasticsearch can be used to search different kinds of documents that provide scalable search, multi-tenancy, and real-time search. Elasticsearch is also available in Amazon Cloud as Amazon Web Services Elasticsearch Cloud. Elasticsearch is a distributed, RESTful search analytics engine that is capable of helping in solving numerous use cases for the business requirement in big data or data science environment.

2. What is an index and inverted index in Elasticsearch?

Answer:

Elasticsearch has a concept called index which is similar to that of a table in a relational database structure. An index has mappings that define multiple types. An index maps one or more multiple shards and can have zero or many replica shards. Here Shard is an index which is split into multiple elements. Elasticsearch will have one replica for each index. The main reason for fast searching in elastic search is that index will be searched instead of content making it so faster.

The inverted index is an index which is used to make very fast full-text searches which is a key component. This is used to search and make a list of all unique words searched in all documents. To create the inverted index, first, the field of each document should be split into separate elements. In an inverted index, to store a mapping from content, words or numbers can be used in a database to its file location. The inverted index is a key component and structure of the elastic search to provide very fast full-text searches.

3. What is a document in Elasticsearch?

Answer:

A document in elastic search is a top level or a root component object that is serialized into JSON object and will be stored in Elastic search under a unique id. The entities or objects in most of the applications can be serialized to JSON with keys and values where the key is the property or name of the field and value is the data present for that key such as String or Number or Boolean etc.,

Documents in elastic search are indexed and stored and will be available to search using the index. In the key-value pairs, the indexes can be generated using auto-generated id values. Inelastic search, document, and object are often interchangeable words. Mapping is the process to define a document, and the fields it contains which are stored and indexed. In a document, each index will have one mapping type that defines how mapping can be done and a document will be indexed. Each mapping type will have meta fields and fields, where meta fields can be used to customize the document’s metadata. Each field will have data types like Boolean, double, long, date or text etc.

4. What is a Node in Elasticsearch?

Answer:

A node is an important component in an Elasticsearch which is needed before starting an instance of Elasticsearch. A group of nodes is called a cluster. If a single node of Elasticsearch is running, then it is called a cluster of one node. In networking, the transport layer is used to establish communication between nodes of a cluster. Each and every node existing in a cluster can send client requests to each other and can establish communication with each other.

There are several types of nodes such as master node, data node, ingest node and tribe node. A Master node is a node that controls the entire cluster. A data node is a node that holds data in it and performs logical operations on the data. An ingest node is a node that can be used to ingest pipeline which means a series of processors to a document to perform some transformations before indexing the document. A tribe node is a node that performs some coordination to connect to multiple clusters across all the connected clusters and perform some logical operations or searches. By default, a node will always be a master node and a data node, but depending on the large requirements, node configurations should be carried out.

5. What is Schema in Elasticsearch?

Answer:

A schema is a structure that describes multiple fields that provides the detailed overview of the document and its type and the way of handling the fields inside the document. The schema is used for mapping in Elasticsearch which describes the fields in JSON documents with its data types. This process is called schema mapping in Elasticsearch. An Elasticsearch server usually contains zero or more indexes. An index contains multiple types which will have multiple documents in them. The other feature of elastic search is that it can also be schema-less by making the documents to be indexed without providing schema clearly.

If a mapping is not explicitly provided in elastic search, then a default mapping will be generated automatically while detecting fields during the process of indexing. This is the process of dynamic mapping generation. The mapping will be done in the form of JSON in elastic search and this will be the hierarchically structured format. Each level in the hierarchy will be having properties configuration to make it work flexibly as per requirement. This means each and every level and its child levels will be having each property set to the last level.

1) What is Logstash? Explain?

A) Logstash is an open source data collection engine with real-time pipelining capabilities. Logstash can dynamically unify data from disparate sources and normalize the data into destinations of your choice. Cleanse and democratize all your data for diverse advanced downstream analytics and visualization use cases.

2) What is Logstash used for?

A) Logstash is an open source tool for collecting, parsing, and storing logs for future use. Kibana 3 is a web interface that can be used to search and view the logs that Logstash has indexed. Both of these tools are based on Elasticsearch.

3) What does Logstash forwarder do?

A) Filebeat is based on the Logstash Forwarder source code and replaces Logstash Forwarder as the method to use for tailing log files and forwarding them to Logstash. The registry file, which stores the state of the currently read files, was changed.

4) What is ELK Stack (Elastic Stack)?

A) Elasticsearch, Logstash, and Kibana, when used together is known as an ELK stack.

5) What is the Power of Logstash?

A) The ingestion workhorse for Elasticsearch and more – Horizontally scalable data processing pipeline with strong Elasticsearch and Kibana synergy

6) What are Logs and Metrics in Logstash?

A) Logs and Metrics – Logstash handle all types of logging data.

Easily ingest a multitude of web logs like Apache, and application logs like log4j for Java

Capture many other log formats like syslog, networking and firewall logs, and more

Enjoy complimentary secure log forwarding capabilities with Filebeat

Collect metrics from Ganglia, collectd, NetFlow, JMX, and many other infrastructure and application platforms over TCP and UDP

7) How does Logstash work with the web?

A) Transform HTTP requests into events

Consume from web service firehoses like Twitter for social sentiment analysis

Webhook support for GitHub, HipChat, JIRA, and countless other applications

Enables many Watcher alerting use cases

Create events by polling HTTP endpoints on demand

Universally capture health, performance, metrics, and other types of data from web application interfaces

Perfect for scenarios where the control of polling is preferred over receiving

8) Which Java version is required to install Logstash?

A) Logstash requires Java 8. Java 9 is not supported.

9) What are the two required elements in Logstash pipeline?

A) A Logstash pipeline has two required elements, input and output, and one optional element, filter. The input plugins consume data from a source, the filter plugins modify the data as you specify, and the output plugins write the data to a destination.

10) What is Filebeat?

A) The Filebeat client is a lightweight, resource-friendly tool that collects logs from files on the server and forwards these logs to your Logstash instance for processing.

Filebeat is designed for reliability and low latency. Filebeat has a light resource footprint on the host machine, and the Beats input plugin minimizes the resource demands on the Logstash instance.

Elasticsearch Logstash Interview Questions

11) What is grok filter plugin?

A) The grok filter plugin enables you to parse the unstructured log data into something structured and queryable.

Because the grok filter plugin looks for patterns in the incoming log data, configuring the plugin requires you to make decisions about how to identify the patterns that are of interest to your use case.

12) What is geoip plugin?

A) geoip plugin looks up IP addresses, derives geographic location information from the addresses, and adds that location information to the logs

13) How do you read data from a Twitter Feed?

A) To add a Twitter feed, you use the twitter input plugin. To configure the plugin, you need several pieces of information:

A consumer key, which uniquely identifies your Twitter app.

A consumer secret, which serves as the password for your Twitter app.

One or more keywords to search in the incoming feed. The example shows using “cloud” as a keyword, but you can use whatever you want.

An oauth token, which identifies the Twitter account using this app.

An oauth token secret, which serves as the password of the Twitter account.

14) Can you explain how Logstash Works?

A) The Logstash event processing pipeline has three stages: inputs -> filters -> outputs. Inputs generate events, filters modify them, and outputs ship them elsewhere. Inputs and outputs support codecs that enable you to encode or decode the data as it enters or exits the pipeline without having to use a separate filter.

15) What are Inputs in Logstash?

A) You use inputs to get data into Logstash.

Some of the more commonly-used inputs are: file, syslog, redis, and beats.

file: reads from a file on the filesystem, much like the UNIX command tail -0F

syslog: listens on the well-known port 514 for syslog messages and parses according to the RFC3164 format

redis: reads from a redis server, using both redis channels and redis lists. Redis is often used as a “broker” in a centralized Logstash installation, which queues Logstash events from remote Logstash “shippers”.

beats: processes events sent by Filebeat.

16) What are Filters in Logstash?

A) Filters are intermediary processing devices in the Logstash pipeline. You can combine filters with conditionals to perform an action on an event if it meets certain criteria. Some useful filters include:

grok: parse and structure arbitrary text. Grok is currently the best way in Logstash to parse unstructured log data into something structured and queryable. With 120 patterns built-in to Logstash, it’s more than likely you’ll find one that meets your needs!

mutate: perform general transformations on event fields. You can rename, remove, replace, and modify fields in your events.

drop: drop an event completely, for example, debug events.

clone: make a copy of an event, possibly adding or removing fields.

geoip: add information about geographical location of IP addresses (also displays amazing charts in Kibana!)

17) What are Outputs in Logstash?

A) Outputs are the final phase of the Logstash pipeline. An event can pass through multiple outputs, but once all output processing is complete, the event has finished its execution. Some commonly used outputs include:

elasticsearch: send event data to Elasticsearch. If you’re planning to save your data in an efficient, convenient, and easily queryable format.

file: write event data to a file on disk.

graphite: send event data to graphite, a popular open source tool for storing and graphing metrics.

statsd: send event data to statsd, a service that “listens for statistics, like counters and timers, sent over UDP and sends aggregates to one or more pluggable backend services”. If you’re already using statsd, this could be useful for you!

18) What are Codecs in Logstash?

A) Codecs are basically streamed filters that can operate as part of an input or output. Codecs enable you to easily separate the transport of your messages from the serialization process. Popular codecs include json, msgpack, and plain (text).

json: encode or decode data in the JSON format.

multiline: merge multiple-line text events such as java exception and stacktrace messages into a single event.

19) Explain the Execution Model of Logstash?

A) The Logstash event processing pipeline coordinates the execution of inputs, filters, and outputs.

Each input stage in the Logstash pipeline runs in its own thread. Inputs write events to a central queue that is either in memory (default) or on disk.

Each pipeline worker thread takes a batch of events off this queue, runs the batch of events through the configured filters, and then runs the filtered events through any outputs.

20) How many types of Logstash Configuration Files are there?

A) Logstash has two types of configuration files: pipeline configuration files, which define the Logstash processing pipeline, and settings files, which specify options that control Logstash startup and execution.

Kibana Interview Questions # 1) What is Kibana?

A) Kibana is an open source data visualization plugin for Elasticsearch. It provides visualization capabilities on top of the content indexed on an Elasticsearch cluster.

Kibana Interview Questions # 2) What is Kibana used for?

A) Logstash is an open source tool for collecting, parsing, and storing logs for future use. Kibana 3 is a web interface that can be used to search and view the logs that Logstash has indexed. Both of these tools are based on Elasticsearch. Elasticsearch, Logstash, and Kibana, when used together is known as an ELK stack.

Kibana Interview Questions # 3) What is Elasticsearch Logstash Kibana?

A) The ELK stack consists of Elasticsearch, Logstash, and Kibana. Although they’ve all been built to work exceptionally well together, each one is a separate project that is driven by the open-source vendor Elastic—which itself began as an enterprise search platform vendor.

Kibana Interview Questions # 4) What is the Filebeat?

A) Filebeat is a log data shipper for local files. Installed as an agent on your servers, Filebeat monitors the log directories or specific log files, tails the files, and forwards them either to Elasticsearch or Logstash for indexing.

Kibana Interview Questions # 5) What is the elastic stack?

A) Elastic Stack is a group of open source products from Elastic designed to help users take data from any type of source and in any format and search, analyze, and visualize that data in real time.

Kibana Interview Questions # 6) What are the main components on Kibana interface?

A) The Kibana interface is divided into four main sections:

Discover

Visualize

Dashboard

Settings

Kibana Interview Questions # 7) What is Kibana Discover interface?

A) When you first connect to Kibana 4, you will be taken to the Discover page. By default, this page will display all of your ELK stack’s most recently received logs.

Kibana Interview Questions # 8) What is Kibana Visualize interface?

A) The Kibana Visualize page is where you can create, modify, and view your own custom visualizations. There are several different types of visualizations, ranging from Vertical bar and Pie charts to Tile maps (for displaying data on a map) and Data tables.

Kibana Interview Questions # 9) What is Kibana Dashboard?

A) The Kibana Dashboard page is where you can create, modify, and view your own custom dashboards. With a dashboard, you can combine multiple visualizations onto a single page, then filter them by providing a search query or by selecting filters by clicking elements in the visualization. Dashboards are useful for when you want to get an overview of your logs, and make correlations among various visualizations and logs.

Kibana Interview Questions # 10) How to create Kibana Dashboard?

A) To create a Kibana dashboard, first, click the Dashboard menu item.

Here is a breakdown of the steps that are being performed:

Clicked Add Visualization icon

Added “Log Counts” pie chart and “Nginx: Top 10 client IP” histogram

Collapsed the Add Visualization menu

Rearranged and resized the visualizations on the dashboard

Clicked Save Dashboard icon

Choose a name for your dashboard before saving it.

Kibana Elasticsearch Interview Questions

Kibana Interview Questions # 11) What are Kibana Settings?

A) The Kibana Settings page lets you change a variety of things like default values or index patterns. In this tutorial, we will keep it simple and focus on the Indices and Objects sections.

Kibana Interview Questions # 12) Is Elasticsearch a Nosql DB?

A) Elasticsearch is a full-text, distributed NoSQL database. In other words, it uses documents rather than schema or tables. It’s a free, open source tool that allows for real-time searching and analyzing of your data.

Kibana Interview Questions # 13) What is Kibana Docker Image?

A) The images are available in two different configurations or “flavors”. The x-pack flavor, which is the default, ships with X-Pack features pre-installed. The oss flavor does not include X-Pack, and contains only open source Kibana.

You can download Kibana docker image at: <https://github.com/elastic/kibana-docker>

Kibana Interview Questions # 14) What is Kibana Port?

A) The default settings configure Kibana to run on localhost:5601 . To change the host or port number, or connect to Elasticsearch running on a different machine, you’ll need to update your kibana.yml file. You can also enable SSL and set a variety of other options.

Kibana Interview Questions # 15) What is kibana.yml?

A) The Kibana server reads properties from the kibana.yml file on startup. To change the host or port number, or connect to Elasticsearch running on a different machine, you’ll need to update your kibana.yml file. You can also enable SSL and set a variety of other options.

**1) What is Elasticsearch and what do you understand about Elasticsearch?**

A) Elasticsearch is a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents.

ElasticSearch is an open source, RESTful search engine built on top of Apache Lucene and released under an Apache license. It is Java-based and can search and index document files in diverse formats. An index can be easily recovered in a case of a server crash.

**2) How does Elasticsearch work?**

A) Indexing Documents to the Repository. During an indexing operation, Elasticsearch converts raw data such as log files or message files into internal documents and stores them in a basic data structure similar to a JSON object. Simply do an HTTP POST that transmits your document as a simple JSON object.

**3) What is Amazon Elasticsearch?**

A) Amazon Elasticsearch Service makes it easy to deploy, secure, operate and scale Elasticsearch for log analytics, full-text search, application monitoring, and more. You can set up and configure petabyte-scale Amazon Elasticsearch Service domains in minutes from the AWS Management Console.

**4) What is the functionality of Elasticsearch?**

A) Elasticsearch is developed in Java and is released as open source under the terms of the Apache License. Elasticsearch provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents.

**5) What** is **Kibana and elastic search?**

A) Kibana is an open source data visualization plugin for Elasticsearch. It provides visualization capabilities on top of the content indexed on an Elasticsearch cluster. Users can create bar, line and scatter plots, or pie charts and maps on top of large volumes of data.

**6) What is Apache Lucene?**

A) Apache Lucene is a free and open-source information retrieval software library, originally written completely in Java.

**7) What is NRT in Elasticsearch?**

A) In Elasticsearch NRT stands for Near Real Time Search platform. Elasticsearch is a near real-time search platform. What this means is there is a slight latency (normally one second) from the time you index a document until the time it becomes searchable.

**8) What is a Cluster in Elasticsearch?**

A) A cluster is a collection of one or more nodes (servers) that together holds your entire data and provides federated indexing and search capabilities across all nodes. A cluster is identified by a unique name which by default is “elasticsearch”. This name is important because a node can only be part of a cluster if the node is set up to join the cluster by its name.

**9) What is Node in Elasticsearch?**

A) A node is a single server that is part of your cluster, stores your data, and participates in the cluster’s indexing and search capabilities. Just like a cluster, a node is identified by a name which by default is a random Universally Unique IDentifier (UUID) that is assigned to the node at startup.

### Top 40 Elasticsearch Interview Questions

**Elasticsearch Interview Questions # 10) What is Index in Elasticsearch?**

A) Index – An index is a collection of documents that have somewhat similar characteristics. For example, you can have an index for customer data, another index for a product catalog, and yet another index for order data.

An index is identified by a name (that must be all lowercase) and this name is used to refer to the index when performing indexing, search, update, and delete operations against the documents in it.

**Elasticsearch Interview Questions # 11) What is Document in Elasticsearch?**

A) Document – A document is a basic unit of information that can be indexed. For example, you can have a document for a single customer, another document for a single product, and yet another for a single order. This document is expressed in JSON (JavaScript Object Notation) which is a ubiquitous internet data interchange format.

**Elasticsearch Interview Questions # 12) What are Shards in Elasticsearch and Explain the concept?**

A) An index can potentially store a large amount of data that can exceed the hardware limits of a single node. For example, a single index of a billion documents taking up 1TB of disk space may not fit on the disk of a single node or may be too slow to serve search requests from a single node alone.

To solve this problem, Elasticsearch provides the ability to subdivide your index into multiple pieces called shards. When you create an index, you can simply define the number of shards that you want. Each shard is in itself a fully-functional and independent “index” that can be hosted on any node in the cluster.

**Elasticsearch Interview Questions # 13) What are the benefits of Sharding in Elasicsearch?**

A) Sharding is important for two primary reasons:

* It allows you to horizontally split/scale your content volume
* It allows you to distribute and parallelize operations across shards (potentially on multiple nodes) thus increasing performance/throughput

**Elasticsearch Interview Questions # 14) What are Replicas and Explain what do you understand?**

A) In a network/cloud environment where failures can be expected any time, it is very useful and highly recommended to have a failover mechanism in case a shard/node somehow goes offline or disappears for whatever reason.

To this end, Elasticsearch allows you to make one or more copies of your index’s shards into what are called replica shards, or replicas for short.

**Elasticsearch Interview Questions # 15) What are the benefits of Replicas in Elasticsearch?**

A) Replication is important for two primary reasons:

* It provides high availability in case a shard/node fails. For this reason, it is important to note that a replica shard is never allocated on the same node as the original/primary shard that it was copied from.
* It allows you to scale out your search volume/throughput since searches can be executed on all replicas in parallel.

**Elasticsearch Interview Questions # 16) What is the minimum Java version required to install Elasticsearch?**

A) To install Elasticsearch on a machine, you require having at least Java 8.

**Elasticsearch Interview Questions # 17) How do you interact with Cluster in Elasticsearch?**

A) Elasticsearch provides a very comprehensive and powerful REST API that you can use to interact with your cluster.

**Elasticsearch Interview Questions # 18) What are the benefits of REST** API’s **in Elasticsearch?**

A) There are many benefits of using REST API’s in Elasticsearch, they are:

* Check your cluster, node, and index health, status, and statistics
* Administer your cluster, node, and index data and metadata
* Perform CRUD (Create, Read, Update, and Delete) and search operations against your indexes
* Execute advanced search operations such as paging, sorting, filtering, scripting, aggregations, and many others

**Elasticsearch Interview Questions # 19) How do you create an Index in Elasticsearch?**

A) Now let’s create an index named “customer” and then list all the indexes again:

PUT /customer?pretty  
GET /\_cat/indices?v

he first command creates the index named “customer” using the PUT verb. We simply append pretty to the end of the call to tell it to pretty-print the JSON response (if any).

**Elasticsearch Interview Questions # 20) How do you delete an Index?**

A) Now let’s delete the index that we just created and then list all the indexes again:

DELETE /customer?pretty  
GET /\_cat/indices?v

### Elasticsearch Frequently Asked Interview Questions

**Elasticsearch Interview Questions # 21) What are the different packages available for installing Elasticsearch?**

A) zip/tar.gz – The zip and tar.gz packages are suitable for installation on any system and are the easiest choice for getting started with Elasticsearch on most systems.

* deb – The deb package is suitable for Debian, Ubuntu, and other Debian-based systems.
* rpm – The rpm package is suitable for installation on Red Hat, Centos, SLES, OpenSuSE and other RPM-based systems.
* msi – The msi package is suitable for installation on Windows 64-bit systems with at least .NET 4.5 framework installed
* docker – Images are available for running Elasticsearch as Docker containers.

**Elasticsearch Interview Questions # 22) What are the configuration management tools supported by Elasticsearch?**

A) Elasticsearch supports the following configuration management tools to help with large deployments:

* Puppet – puppet-elasticsearch
* Chef – cookbook-elasticsearch
* Ansible – ansible-elasticsearch

**Elasticsearch Interview Questions # 23) How many types of Configuration files are there in Elasticsearch?**

A) Elasticsearch has three configuration files:

1. elasticsearch.yml for configuring Elasticsearch
2. jvm.options for configuring Elasticsearch JVM settings
3. log4j2.properties for configuring Elasticsearch logging

These files are located in the config directory, whose default location depends on whether or not the installation is from an archive distribution (tar.gz or zip) or a package distribution (Debian or RPM packages).

**Elasticsearch Interview Questions # 24) What is X-Pack in Elasticsearch?**

A) X-Pack is an Elastic Stack extension that bundles security, alerting, monitoring, reporting, machine learning, and graph capabilities into one easy-to-install package. To access this functionality, you must install X-Pack in Elasticsearch.

**Elasticsearch Interview Questions # 25) Where do you configure settings for X-Pack?**

A) X-Pack Settings in Elasticsearch – You configure settings for X-Pack features in the elasticsearch.yml, kibana.yml, and logstash.yml configuration files.

**Elasticsearch Interview Questions # 26) What are breaking changes in Elasticsearch?**

A) You need to be aware of when migrating your application from one version of Elasticsearch to another.

As a general rule:

* Migration between minor versions – e.g. 6.x to 6.y – can be performed by upgrading one node at a time.
* Migration between consecutive major versions – e.g. 5.x to 6.x – requires a full cluster restart.
* Migration between non-consecutive major versions – e.g. 2.x to 6.x – is not supported.

**Elasticsearch Interview Questions # 27) What is Single document APIs in Elasticsearch?**

A) Index API, Get API, Delete API, Update API

**Elasticsearch Interview Questions # 28) What are Multi-document APIs?**

A) Multi Get API, Bulk API, Delete By Query API, Update By Query API, Reindex API.

**Elasticsearch Interview Questions # 29) What is Routing in Elasticsearch?**

A) When executing a search, it will be broadcast to all the index/indices shards (round robin between replicas). Which shards will be searched on can be controlled by providing the routing parameter.

**Elasticsearch Interview Questions # 30) What are Aggregations?**

A) The aggregation’s framework helps provide aggregated data based on a search query. It is based on simple building blocks called aggregations, that can be composed in order to build complex summaries of the data.

An aggregation can be seen as a unit-of-work that builds analytic information over a set of documents. The context of the execution defines what this document set is (e.g. a top-level aggregation executes within the context of the executed query/filters of the search request).

### Advanced Elasticsearch Interview Questions

**31) What are the different types of aggregations in Elasticsearch?**

A) There are many different types of aggregations, each with its own purpose and output.

* Metric – Aggregations that keep track and compute metrics over a set of documents.
* Matrix – A family of aggregations that operate on multiple fields and produce a matrix result based on the values extracted from the requested document fields. Unlike metric and bucket aggregations, this aggregation family does not yet support scripting.
* Pipeline – Aggregations that aggregate the output of other aggregations and their associated metrics

**32) What are Indices APIs?**

A) The indices APIs are used to manage individual indices, index settings, aliases, mappings, and index templates.

**33) What is cat API in Elasticsearch?**

A) All the cat commands accept a query string parameter help to see all the headers and info they provide, and the /\_cat command alone lists all the available commands.

**34) What are the different cat commands available in Elasticsearch cat API?**

A) The different commands available in cat APIs are:

* cat aliases, cat allocation, cat count, cat fielddata
* cat health, cat indices, cat master, cat nodeattrs
* cat nodes, cat pending tasks, cat plugins, cat recovery
* cat repositories, cat thread pool, cat shards, cat segments
* cat snapshots, cat templates

**35) What is Query DSL in Elasticsearch?**

A) Elasticsearch provides a full Query DSL (Domain Specific Language) based on JSON to define queries. Think of the Query DSL as an AST (Abstract Syntax Tree) of queries, consisting of two types of clauses:

Leaf query clauses – Leaf query clauses look for a particular value in a particular field, such as the match, term or range queries. These queries can be used by themselves.

Compound query clauses – Compound query clauses wrap other leaf or compound queries and are used to combine multiple queries in a logical fashion (such as the bool or dis\_max query), or to alter their behavior (such as the constant\_score query).

**36) What is Ingest Node?**

A) Use an ingest node to pre-process documents before the actual document indexing happens. The ingest node intercepts bulk and index requests, it applies transformations, and it then passes the documents back to the index or bulk APIs.

**37) What are the different types of X-Pack APIs?**

A) X-Pack APIs – X-Pack exposes REST APIs that are used by the UI components and can be called directly to configure and access X-Pack features.

* Info API
* Graph Explore API
* Machine Learning APIs
* Security APIs
* Watcher APIs
* Migration APIs

**38) What are the different types of X-Pack Commands?**

A) X-Pack includes commands that help you configure security:

* certgen
* certutil
* migrate
* saml-metadata
* setup-passwords
* syskeygen
* users

**39) What is Explore API in Elasticsearch?**

A) The Graph explore API enables you to extract and summarize information about the documents and terms in your Elasticsearch index.

**40) What is Migration APIs in Elasticsearch?**

A) The migration APIs simplify upgrading X-Pack indices from one version to another.

Migration Assistance API  
Migration Upgrade API  
Deprecation Info APIs

* + **Question 1. What Is An Index In Elasticsearch ?**
  + **Answer :**
  + An index is similar to a table in relational databases. The difference is that relational databases would store actual values, which is optional in ElasticSearch. An index can store actual and/or analyzed values in an index.
  + **Question 2. What Is A Document In Elasticsearch ?**
  + **Answer :**
  + A document is similar to a row in relational databases. The difference is that each document in an index can have a different structure (fields), but should have same data type for common fields.
  + Each field can occur multiple times in a document with different data types. Fields can contain other documents too.
  + **Question 3. Does Elasticsearch Have A Schema ?**
  + **Answer :**
  + Yes, ElasticSearch can have mappings which can be used to enforce schema on documents.
  + **Question 4. What Is A Document Type In Elasticsearch ?**
  + **Answer :**
  + A document type can be seen as the document schema / mapping definition, which has the mapping of all the fields in the document along with its data types.
  + **Question 5. What Is Indexing In Elasticsearch ?**
  + **Answer :**
  + The process of storing data in an index is called indexing in ElasticSearch. Data in ElasticSearch can be dividend into write-once and read-many segments. Whenever an update is attempted, a new version of the document is written to the index.
  + **Question 6. What Is A Node In Elasticsearch ?**
  + **Answer :**
  + Each instance of ElasticSearch is called a node. Multiple nodes can work in harmony to form an ElasticSearch Cluster.
  + **Question 7. What Is A Shard In Elasticsearch ?**
  + **Answer :**
  + Due to resource limitations like RAM, vCPU etc, for scale-out, applications need to employ multiple instances of ElasticSearch on separate machines. Data in an index can be divided into multiple partitions, each handled by a separate node (instance) of ElasticSearch. Each such partition is called a shard. By default an ElasticSearch index has 5 shards.
  + **Question 8. What Is A Replica In Elasticsearch ?**
  + **Answer :**
  + Each shard in ElasticSearch has 2 copy of the shard. These copies are called replicas. They serve the purpose of high-availability and fault-tolerance.
  + **Question 9. What Is An Analyzer In Elasticsearch ?**
  + **Answer :**
  + While indexing data in ElasticSearch, data is transformed internally by the Analyzer defined for the index, and then indexed. An analyzer is built of tokenizer and filters. Following types of Analyzers are available in ElasticSearch 1.10.
    - STANDARD ANALYZER
    - SIMPLE ANALYZER
    - WHITESPACE ANALYZER
    - STOP ANALYZER
    - KEYWORD ANALYZER
    - PATTERN ANALYZER
    - LANGUAGE ANALYZERS
    - SNOWBALL ANALYZER
    - CUSTOM ANALYZER
  + **Question 10. What Is A Tokenizer In Elasticsearch ?**
  + **Answer :**
  + A Tokenizer breakdown fields values of a document into a stream, and inverted indexes are created and updates using these values, and these stream of values are stored in the document.
  + **Question 11. What Is A Filter In Elasticsearch ?**
  + **Answer :**
  + After data is processed by Tokenizer, the same is processed by Filter, before indexing. Following types of Filters are available in ElasticSearch 1.10.
    - AND FILTER
    - BOOL FILTER
    - EXISTS FILTER
    - GEO BOUNDING BOX FILTER
    - GEO DISTANCE FILTER
    - GEO DISTANCE RANGE FILTER
    - GEO POLYGON FILTER
    - GEOSHAPE FILTER
    - GEOHASH CELL FILTER
    - HAS CHILD FILTER
    - HAS PARENT FILTER
    - IDS FILTER
    - INDICES FILTER
    - LIMIT FILTER
    - MATCH ALL FILTER
    - MISSING FILTER
    - NESTED FILTER
    - NOT FILTER
    - OR FILTER
    - PREFIX FILTER
    - QUERY FILTER
    - RANGE FILTER
    - REGEXP FILTER
    - SCRIPT FILTER
    - TERM FILTER
    - TERMS FILTER
    - TYPE FILTER
  + **Question 12. What Is The Query Language Of Elasticsearch ?**
  + **Answer :**
  + ElasticSearch uses the Apache Lucene query language, which is called Query DSL.
  + **Question 13. What Is Elasticsearch ?**
  + **Answer :**
  + Elasticsearch is a search engine based on Lucene. It provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents. Elasticsearch is developed in Java and is released as open source under the terms of the Apache License.
  + **Question 14. What Are The Basic Operations You Can Perform On A Document ?**
  + **Answer :**
  + The following operations can be performed on documents
    - INDEXING A DOCUMENT USING ELASTICSEARCH.
    - FETCHING DOCUMENTS USING ELASTICSEARCH.
    - UPDATING DOCUMENTS USING ELASTICSEARCH.
    - DELETING DOCUMENTS USING ELASTICSEARCH.
  + Perform basic operations with Elasticsearch.
  + **Question 15. What Is Inverted Index In Elasticsearch ?**
  + **Answer :**
  + Inverted index is the heart of search engines. The primary goal of a search engine is to provide speedy searches while finding the documents in which our search terms occur. Inverted index is a hashmap like data structure that directs users from a word to a document or a web page. It is the heart of search engines. Its main goal is to provide quick searches for finding data from millions of documents.
  + Usually in Books we have inverted indexes as below. Based on the word we can thus find the page on which the word exists.
  + Consider the following statements
    - javainuse is a good website
    - javainuse is one of the good websites.
  + For indexing purpose the above text are tokenized into separate terms and all the unique terms are stored inside the index with information such as in which document this term appears and what is the term position in that document.
  + So the inverted index for the document text will be as follows-
  + When you search for the term website OR websites, the query is executed against the inverted index and the terms are looked out for, and the documents where these terms appear are quickly identified.
  + **Question 16. What Is A Cluster In Elasticsearch ?**
  + **Answer :**
  + Cluster is a collection of one or more nodes (servers) that together holds your entire data and provides federated indexing and search capabilities across all nodes. A cluster is identified by a unique name which by default is "elasticsearch". This name is important because a node can only be part of a cluster if the node is set up to join the cluster by its name.