Quiz Elastic Search BigData

1) What is Elasticsearch?

Elasticsearch is a NoSQL database. It is based on the Lucene search engine, and it is built with RESTful APIS. It offers simple deployment, maximum reliability, and easy management. It also provides advanced queries to perform detailed analysis and stores all the data centrally. It helps execute a quick search of the documents.

2) What are the important features of Elasticsearch?

Here are important features of Elasticsearch:

- An open-source search server written using Java.
- Used to index any kind of heterogeneous data
- Has REST API web-interface with JSON output
- Full-Text Search
- Near Real-Time (NRT) search
- Sharded, replicated searchable, JSON document store.
- Schema-free, REST & JSON based distributed document store
- Multi-language & Geolocation support

3) What is a Cluster?

A cluster is a collection of nodes which together holds data and provides joined indexing and search capabilities.

4) Explain Index

A node is an elastic search Instance. It is created when an elasticsearch instance begins.

5) What is a document in Elastic Search?

In an Elastic search, a document is a basic unit of information that can be indexed. It is expressed in `JSON (key: value) pair. '{"user": "nullcon"}'`. Every single Document is associated with a type and a unique id.

6) Define the Term Shard

Every index can be split into several shards to be able to distribute data. The shard is the atomic part of an index, which can be distributed over the cluster if you want to add more nodes.

7) What are the important advantages of Elastic Search?

Here are the important advantages of Elasticsearch:

- Store schema-less data and also creates a schema for your data.
- Manipulate your data record by record with the help of Multi-document APIs
- Perform filtering and querying your data for insights
- Based on Apache Lucene and provides RESTful API
- It provides horizontal scalability, reliability, and multitenant capability for real-time use of indexing.
- Helps you to scale vertically and horizontally

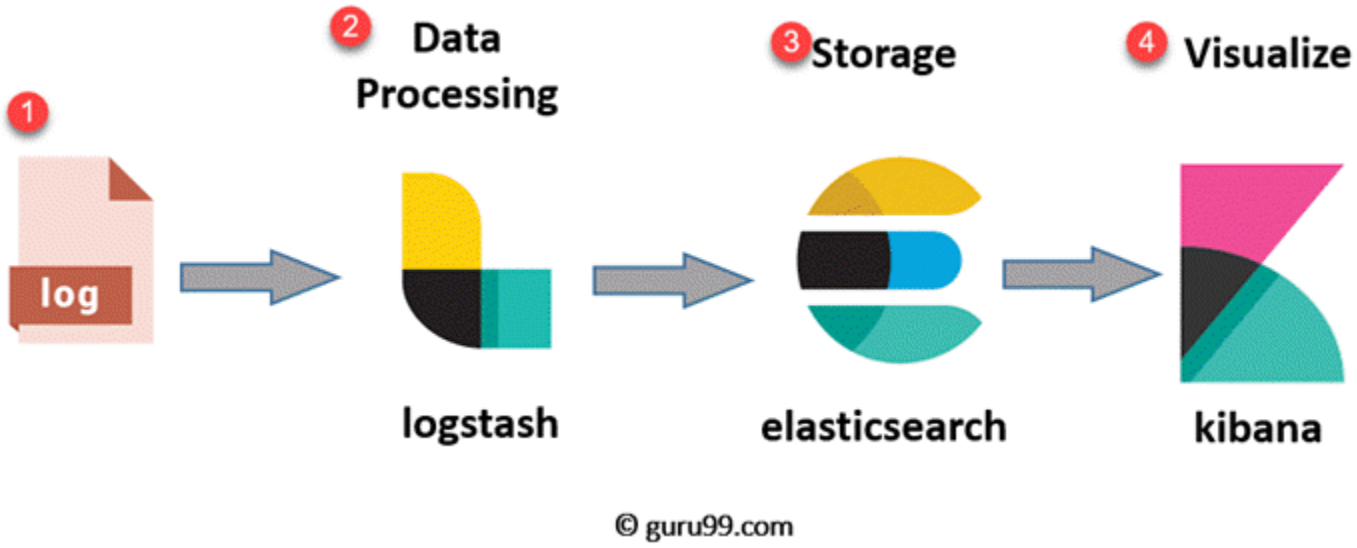
8) What is the ELK stack?

The ELK Stack is a collection of three open-source products — Elasticsearch, Logstash, and Kibana. They are all developed, managed, and maintained by the company Elastic.

- **E stands for ElasticSearch**: It is used for storing logs.
- **L stands for LogStash:** It is used for both shipping as well as the processing and storing logs.
- **K stands for Kibana:** It is a visualization tool (a web interface) that is hosted through Nginx or Apache.

9) Explain ELK stack architecture

ELK Stack is designed to allow users to take to data from any source, in any format, and to search, analyze, and visualize that data in real-time.



- **Logs:** Server logs that need to be analyzed are identified
- **Logstash:** Collect logs and events data. It even parses and transforms data.
- **ElasticSearch:** The transformed data from Logstash is Store, Search, and indexed.
- **Kibana:** Kibana uses Elasticsearch DB to Explore, Visualize, and Share

10) What are the reason for using ELK stack?

Here, are reasons for using ELK stack:

- ELK works best when logs from various Apps of an enterprise converge into a single ELK instance
- It provides amazing insights for this single instance and also eliminates the need to log into a hundred different log data sources.
- Rapid on-premise installation
- Easy to deploy Scales vertically and horizontally
- Elastic offers a host of language clients, which includes Ruby. Python. PHP, Perl, .NET, Java, and JavaScript, and more
- Availability of libraries for different programming and scripting language

11) Explain Tokenizer in ElasticSearch

A Tokenizer breakdown fields which values of a document into a stream. Inverted indexes are created and updated by using these values. After that, these stream of values are stored in the document.

12) What is a replica in ElasticSearch?

Each shard in ElasticSearch has 2 copy, which is called replicas. They help you for high-availability and fault-tolerance.

13) What Are The Main Operations You Can Perform On A Document?

Here, are important operation performed on documents:

- Indexing a document
- Fetching documents
- Updating documents
- Deleting documents

14) What is a Cluster in Elasticsearch?

Cluster is a collection of single or multiple nodes that holds your entire data and offers federated indexing and search abilities across all nodes.

15) How you can delete an index in Elastic search?

To delete an index in Elasticsearch, You need to write command:

```
DELETE /index name.
```

For example, `DELETE /website`.

16) Explain the method to add a mapping in an Index

Elasticsearch allows you to create the mapping according to the data provided by the user in the request body. Its bulk feature can be used to add more than one JSON object in the index.

For example, `POST website /_bulk`.

17) What are the various ways of searching in Elasticsearch?

Following are the way of search in Elasticsearch:

Multi-index, Multitype search: You can search APIs that can be applied across all multiple indices by using the multi-index support system.

In Elastic search, we can create certain tags across all indices across all indices and all types.

- **URI search:** A search request is executed using a URI by providing requested parameters.
- **Request body search:** A search request need to be executed by a search DSL. It includes the query DSL within the body.

18) What is the latest version of Elasticsearch?

Latest version of Elastic Search on Jan 2020, which is the latest and stable version of Elasticsearch.

19) What is Mapping?

Mapping is a process that helps you define how a document is mapped to the search engine. Its searchable characteristics are included fields are tokenized as well as searchable.

20) Where is Elastic Search stored?

You can store Elasticsearch is a distributed document, which is a store with various types of directories. You can also retrieve the complex data structures that can be serialized as JSON documents.

21) What is Apache Lucene?

Apache Lucene is an open-source information retrieval software library. It is originally written in Java language.

22) Here, are important configuration management tool supported by Elasticsearch:

- Puppet – puppet-elasticsearch
- Chef – cookbook-elasticsearch
- Ansible – ansible-elastic search

23) What is NRT in Elasticsearch?

NRT is a full form of (Near Real-Time Search) platform. It is a near real-time search platform. It means there is a slight latency (mostly one second) from when you index a document until it becomes very searchable.

24) Where do you configure settings for X-Pack?

You can configure settings for X-Pack. It has features in the elasticsearch, logstash, and kibana.yml (ELK stack) configuration files.

25) What is cat API in Elasticsearch?

These commands accept a query string parameter. This helps to see all info and headers and info they provide and the /_cat command, which allows you to lists all the available commands.

26) What are the various commands available in Elasticsearch cat API?

Command using with cat API are:

- Cat aliases, cat allocation, cat count, cat field data
- Cat health, cat indices, cat master, pending tasks, cat plugins, cat recovery
- cat repositories, cat snapshots, cat templates

27) What is Ingest node?

Ingest node is use for pre-process documents before the actual document indexing happens. It helps you to intercepts bulk and index requests. It also applies transformations, and then it passes the documents back to the bulk API and index.

28) What are the various ways of using X-Pack Commands?

Here, are X-Pack commands that help you configure security:

- Certgen
- migrate
- syskeygen
- certutil
- saml-metadata
- setup-passwords
- users

29) What is Single document APIs in Elasticsearch?

- Get API
- Index API
- Delete API
- Update API

30) Explain Explore API in Elasticsearch

The Graph explore API allows you to extract and summarize information regarding the documents.

31) How can you create an Index in Elasticsearch?

For example:

```
PUT /client?pretty
```

```
GET /_cat/indices?v
```

32) What are Aggregations?

The aggregations framework helps you to provide aggregated data based on a search query. It is based on simple building blocks known as aggregations. It can be composed to build complex summaries of the data.

33) Does Elasticsearch Have A Schema?

ElasticSearch mappings that can be used to enforce a schema on documents.

34) What is Query DSL in Elasticsearch?

Elasticsearch offers full Query DSL (Domain Specific Language) based on JSON to define queries.

35) What is Elasticsearch Data Node?

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

36) What is a document in ElasticSearch?

The document is very similar to a row in relational databases. Every document in the index possess different structure but has the same data type for respective fields.

- MySQL => Databases => Tables => Columns/Rows
- ElasticSearch => Indices => Types => Documents with Properties

37) Explain type in ElasticSearch

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

38) What is the query language of Elasticsearch?

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

39) What is dynamic mapping in Elasticsearch?

Dynamic mapping helps the user to index documents without unwanted configurations for the field name. Instead, it will be added automatically through the Elasticsearch with some custom rules.

40) What is fuzzy search Elasticsearch?

Fuzzy search is a process in which web page document locations should be identified. That is resembling with the search argument. It also works when the argument is not relevant to the search correspondent for particular information.