**What is Elasticsearch?**

Elasticsearch is a powerful & fast search and analytics engine that works in real time. It is well known for its ability to suggest intelligent results based on prior search queries and returns accurate results for terms entered with misspellings or near matches.

**What is Elasticsearch used for?**

Elasticsearch allows users to create indexes that can be quickly searched through as this search engine can intelligently return search results. Its aggregation function can power filters within an eCommerce website, highlight key parts of results, score results for relevancy and use fuzzy match to return results even when the user has not used exact criteria for an item.

**What is Elasticsearch cluster?**

An Elasticsearch cluster is a group of node instances that are connected together, a cluster also consisting of a single node.

**What is an Elasticsearch Index?**

It functions as a database that contains different types of data for easier organisation and retrieval, its indexes map to primary shards.

**What is a document?**

Documents are JSON objects that are distributed across our elasticsearch cluster which are accessible from any node.

**What is ELK?**

It is known as Elasticsearch, logstash and kibana

**What is logstash?**

Logstash is a powerful pipeline that collects, enriches and transforms data. It works as an ETL tool for collecting log messages and forwarding them onto elasticsearch for visualization within kibana.

**What is kibana?**

Kibana is an open source visualization and reporting UI.

Kibana allows users to produce visualization from a variety of data inputs and can create charts and graphs to reflect data in easily digestible format.

**What are the advantages of using Elasticsearch?**

Key benefits:

It is ability to scale

multilingual

Extensive API

Fast performance even when we are working with massive scale datasets.

**What are the advantages of using Logstash?**

Process unstructured data

pre-built and custom filter

build custom data processing pipelines

work as an ETL tool.

Over 200 plugins available.

**What are the advantages of kiban?**

Real time observability

Integration with Elasticsearch

browser based visualization tool

allows to show graphs,charts and dashboards to get data in digestible format.

**How do we check the version of Elasticsearch that we are working on ?**

Type curl command: curl -XGET ‘http://localhost:9200’

or type localhost:9200 in the browser.

**How do we create an index in Elasticsearch?**

Use the following command:

PUT /index-name

**How to stop Elasticsearch?**

To stop the Elasticsearch use the following command:

ps -ef | grep elas

Here, we can see the Process ids and then type kill PID for elasticsearch.

**What is the curl command?**

Curl command in Elasticsearch allows you to take many actions including but not limited to: deleting indexes, list all indexes, list all documents within an index, query using URL parameters, add data and list index mappings.

**How to delete indexes?**

DELETE /index-name

**What is Filebeat?**

It is a shipper for forwarding and centralizing log data.

Filebeat monitors the log files or locations that we specify, collects log events and forward them either to Elasticsearch or Logstash for indexing.

**What is Metricbeat?**

It is a shipper that periodically collects metrics from the OS and from services running on the server.

Metricbeat takes the metrics and statistics that it collects and ships them to the output that we specify. Such as Es or Logstash.

**How can we make balance shards?**

Shards are balanced when we have an equal number of shards across each node.

ES will run an automatic process of rebalancing shards which moves shards between the nodes that make up our cluster in order to improve performance.

**What is Grok?**

Grok is a filter plugin for logstash that is used to parse the unstructured data. It is often used for transforming apache,syslog and other web server log into a structured and IQueryable to facilitate data analysis to be performed.

**What is Fuzzy search?**

It is used to return positive matches for non-exact matches of the term we are searching for.

**What do we understand by cluster health? how to check the health of a cluster?**

Cluster health helps to show the status of the cluster. It defines how many clusters are running in elasticsearch. The health status is shown by three different colors.

RED: It indicates that some of the primary shards or nodes are not available in the cluster.

RED - The cluster health status will be RED when some of the primary shards or nodes are not available in the cluster.

YELLOW - The cluster health status will be RED when some or all shards are not allocated to any of the cluster.

GREEN- The cluster health status will be RED when the shards are allocated to the node.

**We can’t perform the write operation on frozen indices.**

**What is the ingest node in Elasticsearch?**

Ingest node is used to transform the document before indexing it in Elasticsearch. Ingest nodes pre-process the document before the indexing occurs. Such operations like rename a field name, add or remove field from a document are handled by the ingest node.

**Usage of restore API:**

Es provides \_restore API to restore API to restore data, which backed up to a snapshot.

**Types of Node in Elasticsearch:**

**Master:**

Master node is responsible for lightweight cluster-wide actions such as creating or deleting an index, tracking which nodes are part of the cluster and deciding which shard to allocate to which node.

Node.mater: True

**Data node:**

Data nodes hold the shards that contain documents we have indexed.

No shards will be saved.

node.data: true

**Ingest node:**

Enable → to run ingest pipeline

manipulate documents

before adding indexes to them, like removing some useless fields and adding some more information.

In case of filebeat → off

That means we have two times manipulate the data

1 from files

2 from ingest pipeline.

**ML node:**

Used for machine learning algorithms. If we want to run some algorithms on one of the nodes then we have to set the two settings below.

node.ml = true → enable or disable jobs of ML

xpack.ml.enabled = true/false → used for enable or disable API for ML

**Coordination role:**

It routes queries to the other nodes.

work as a load balancer.

Configure:

node.master =false

node.data=false

node.ingest=false

node.ml=false

xpack.ml.enabled=false

**voting only:**

participate in electing a new master node.

**Built in analyzer in elasticsearch**

Standard analyzer:

IT divides texts into terms on word boundaries, as defined by the unicode text segmentation algorithm. It removes the punctuation,lowercase terms and supports removing stop words.

Simple analyzer:

It divides text into terms whenever it encounters a character which is not a letter. it lowercases all terms.

Whitespace analyzer:

The whitespace is like the simple analyzer but also supports removal of stop words.

Keyword analyzer:

It accepts whatever text it is given and outputs the exact same text as a single line/term.

Pattern analyzer:

It uses a regular expression to

split the text into terms.

It supports lower casing and stop words.

Fingerprint analyzer:

The fingerprint analyzer acreage a fingerprint which can be used for duplication detection.

**Custom analyzer Demo:**

We can create a custom analyzer which uses the combination of

Zero or more character filters.

a tokenizer

zero or more token filters.

Configuration:

| type | Analyzer type |
| --- | --- |
| tokenizer | A built-in or customized tokenizer. |
| char\_filter | built-in or customized character filter. |
| filter | optional array of built-in or customized token filters. |
| postion\_increment\_gap | When indexing an array of text values, Elasticsearch inserts a fake "gap" between the last term of one value and the first term of the next value to ensure that a phrase query doesn’t match two terms from different array elements. Defaults to 100 |

**What is Kibana?**

Kibana is an open source and free front end tool that sits at the top of the elastic stack, allowing users to search and visualize the data which is indexed in elasticsearch.

**What is kibana used for?**

Data indexed in ES is searched, viewed, and visualized and data is analyzed using tables, maps, pie charts, bar charts and histograms.

Using a web interface we can secure, monitor and manage the Elastic stack instance.

Provide centralized access to built in elastic stack solutions for enterprise search applications, security and visibility.

**What is Term Frequency?**

How many times does the term appear in the field for a given documents.

**What is inverse document frequency?**

How often does the term appear within the indexes.

**Field length norm**

How long are the fields.

**What is Elastic relevance?**

Elasticsearch assigns a relevancy, or score to each document, so you can narrow your search to the document with the most relevant result.

The higher the score, the better it matches your query.