elastic

Elasticsearch

Search Engine on your server

Aravind Putrevu
Developer | Evangelist
@aravindputrevu | aravindputrevu.in
elastic.co/community

Agenda

1. Terms
2. Talking to Elasticsearch
3. Mappings
4. Analyzers and Aggregations
5. Capacity Planning

elactir

Agenda

1. Terms
2. Talking to Elasticsearch
3. Mappings
4. Analyzers and Aggregations
5. Capacity Planning

elactir

Agenda

1. Terms
2. Talking to Elasticsearch
3. Mappings
4. Analyzers and Aggregations
5. Capacity Planning

elactir

Agenda

1. Terms
2. Talking to Elasticsearch
3. Mappings
4. Analyzers and Aggregations
5. Capacity Planning

elactic

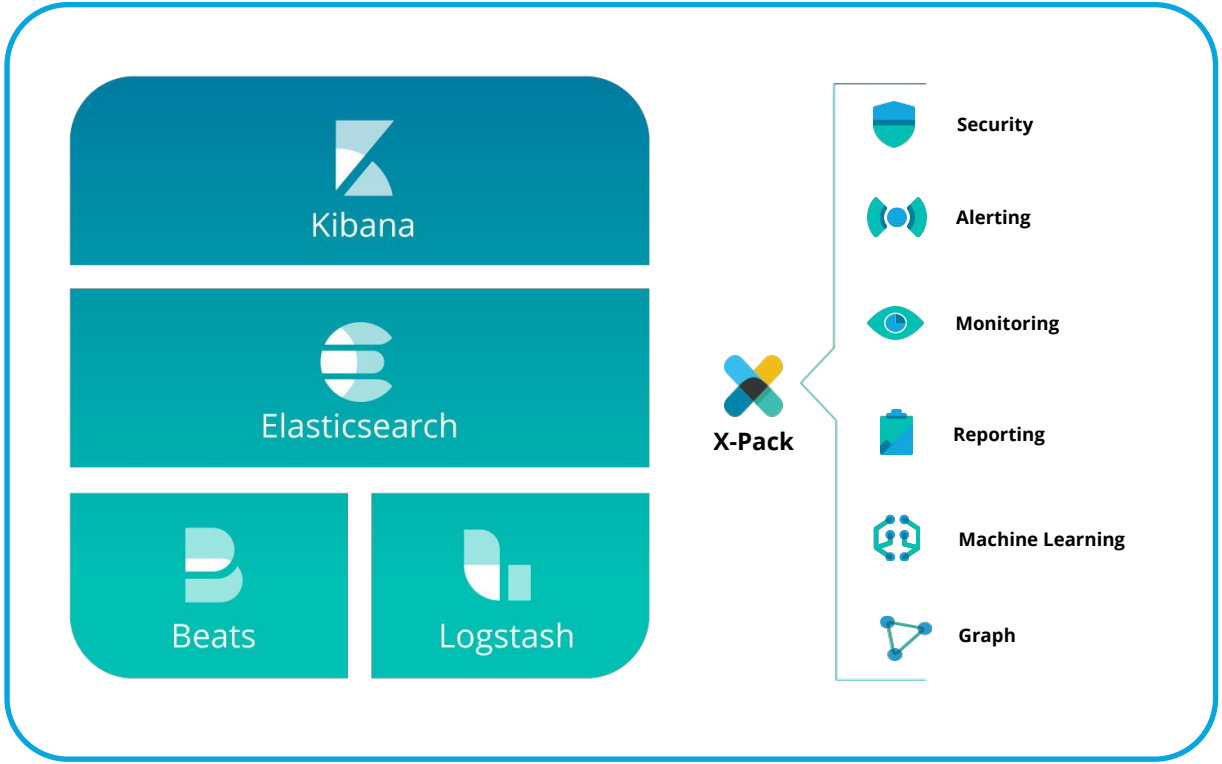
Agenda

1. Terms
2. Talking to Elasticsearch
3. Mappings
4. Analyzers and Aggregations
5. Capacity Planning

elactir

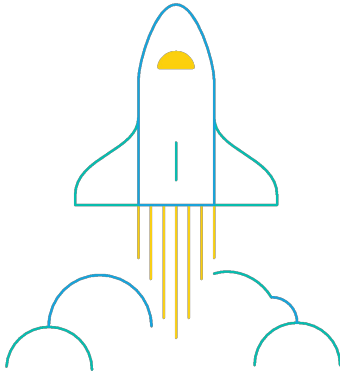
Elastic Stack

No enterprise edition
All new versions with 6.2

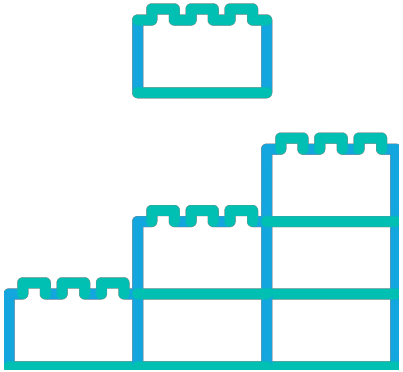


elastic

Why it is Popular?



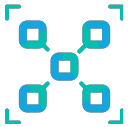
Speed



Scale

Relevance

Terms



Cluster

A cluster is a collection of one or more n

Node

A node is a single server that is part of y
participates in the cluster's indexi

Index

An index is a collection of documents that have somewhat similar characteristics

Type

Deprecated in 6.0.0
A type used to be a logical category/partition of your index to allow you to store different types of documents in the same index

Document

A document is a basic unit of information that can be indexed. This document is expressed in JSON (JavaScript Object Notation) which is a ubiquitous internet data interchange format.

Shard

Elasticsearch provides the ability to subdivide your index into multiple pieces called shards

Replica

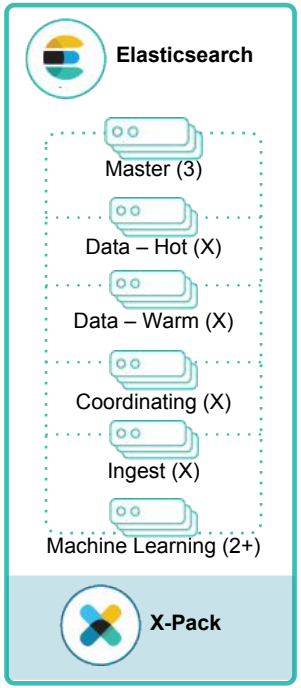
To this end, Elasticsearch allows you to make one or more copies of your index's shards into what are called replica shards, or replicas for short

elastic

https://www.elastic.co/guide/en/elasticsearch/reference/current/glossary.html

Elasticsearch Node Types

Nodes can play one or more roles, for workload isolation and scaling



- Master Nodes
  - Control the cluster, requires a minimum of 3, one is active at any given time
- Data Nodes
  - Hold indexed data and perform data related operations
  - Differentiated Hot and Warm Data nodes can be used
- Coordinating Nodes
  - Route requests, handle search reduce phase, distribute bulk indexing
  - All nodes function as coordinating nodes
- Ingest Nodes
  - Use ingest pipelines to transform and enrich before indexing
- Machine Learning Nodes
  - Run machine learning jobs

What powers Elasticsearch?

- A Java library
- Great for full-text search

But

- Challenging to use
- Not designed for scale

elastic

https://www.elastic.co/blog/found-elasticsearch-top-down

Talking to Elasticsearch

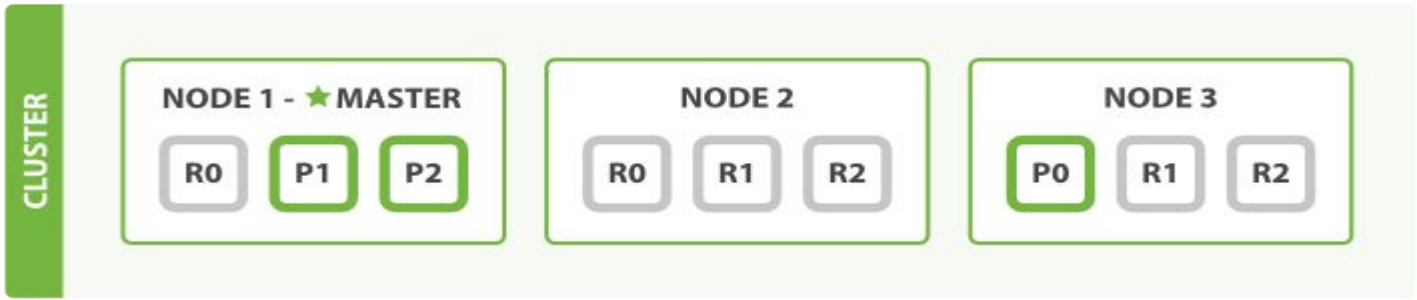
elastic

https://www.elastic.co/guide/en/elasticsearch/client/index.html

Indexing a document

```
PUT /customer/_doc/1?pretty
{
  "name": "John Doe"
}
```

```
{
  "_index" : "customer",
  "_type" : "_doc",
  "_id" : "1",
  "_version" : 1,
  "result" : "created",
  "_shards" : {
    "total" : 2,
    "successful" : 1,
    "failed" : 0
  },
  "_seq_no" : 0,
  "_primary_term" : 1
}
```



elastic

https://www.elastic.co/blog/how-many-shards-should-i-have-in-my-elasticsearch-cluster

Inserting data _bulk

```
curl -X POST "localhost:9200/_bulk" -H 'Content-Type: application/json' -d'
{ "delete": { "_index": "website", "_type": "blog", "_id": "123" }}
{ "create": { "_index": "website", "_type": "blog", "_id": "123" }}
{ "title":    "My first blog post" }
{ "index":  { "_index": "website", "_type": "blog" }}
{ "title":    "My second blog post" }
{ "update": { "_index": "website", "_type": "blog", "_id": "123", "_retry_on_conflict" : 3} }
{ "doc" : {"title" : "My updated blog post"} }
'
```

elastic

Where will my data go?

```
shard = hash(routing) % number_of_primary_shards
```

The default value used for _routing is the document's _id.

0 < shard < number_of_primary_shards - 1

elastic

https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-routing-field.html

Mappings

```
PUT data/_doc/1
{
  "count": 5
}
```

```
GET data/_mapping
{
  "data": {
    "mappings": {
      "_doc": {
        "properties": {
          "count": {
            "type": "long"
          }
        }
      }
    }
  }
}
```

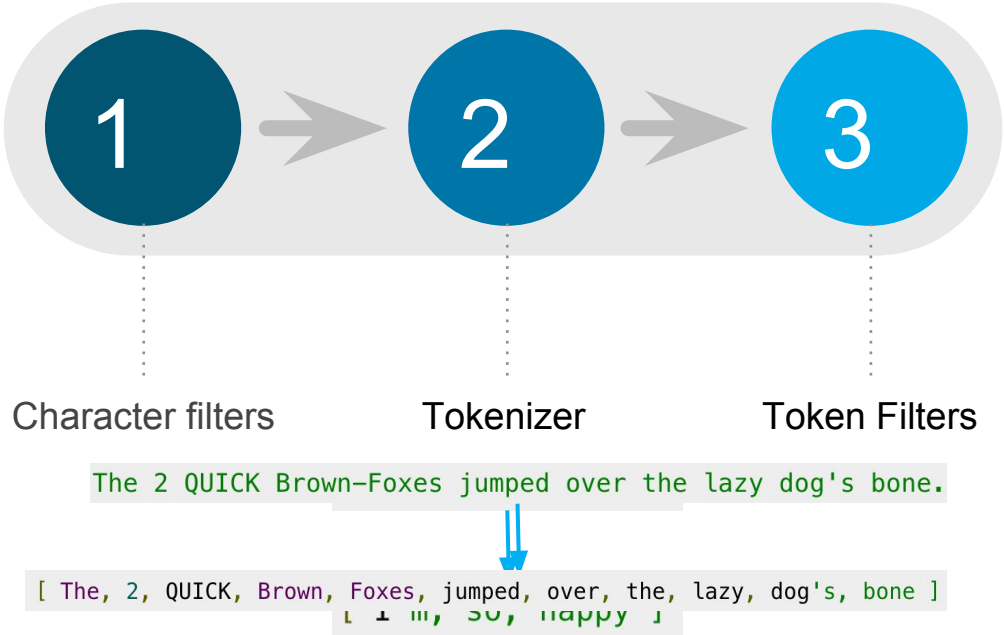
elastic

Full Text Analysis

Inverted Index

Analyzer

Helps in converting text into tokens for better search capability



elastic

Aggregations

- Metrics
- Bucket
- Pipeline
- and so on...

```
{
    ...
    "aggreg
        "t_
    }
}
```

```
SELECT COUNT(color) ❶
FROM table
GROUP BY color ❷
```

```
{
    }
}
```


elastic

Querying Data

- Full Text Queries
- Term Level Queries
- Compound Queries
- Geo Queries

Query DSL

Match Query

```
GET /_search
{
  "query": {
    "multi_match" : {
      "query":    "this is a test", (1)
      "fields": [ "subject", "message" ] (2)
    }
  }
}
```


elastic

Query DSL

Term Queries

```
POST _search
{
  "query": {
    "bool" : {
      "must" : {
        "term" : { "user" : "kimchy" }
      },
      "filter": {
        "term" : { "tag" : "tech" }
      },
      "must_not" : {
        "range" : {
          "age" : { "gte" : 10, "lte" : 20 }
        }
      },
      "should" : [
        { "term" : { "tag" : "wow" } },
        { "term" : { "tag" : "elasticsearch" } }
      ],
      "minimum_should_match" : 1,
      "boost" : 1.0
    }
  }
}
```


elastic

Query DSL

Nested queries

```
GET /_search
{
    "query": {
        "nested" : {
            "path" : "obj1",
            "score_mode" : "avg",
            "query" : {
                "bool" : {
                    "must" : [
                    { "match" : {"obj1.name" : "blue"} },
                    { "range" : {"obj1.count" : {"gt" : 5}} }
                    ]
                }
            }
        }
    }
}
```

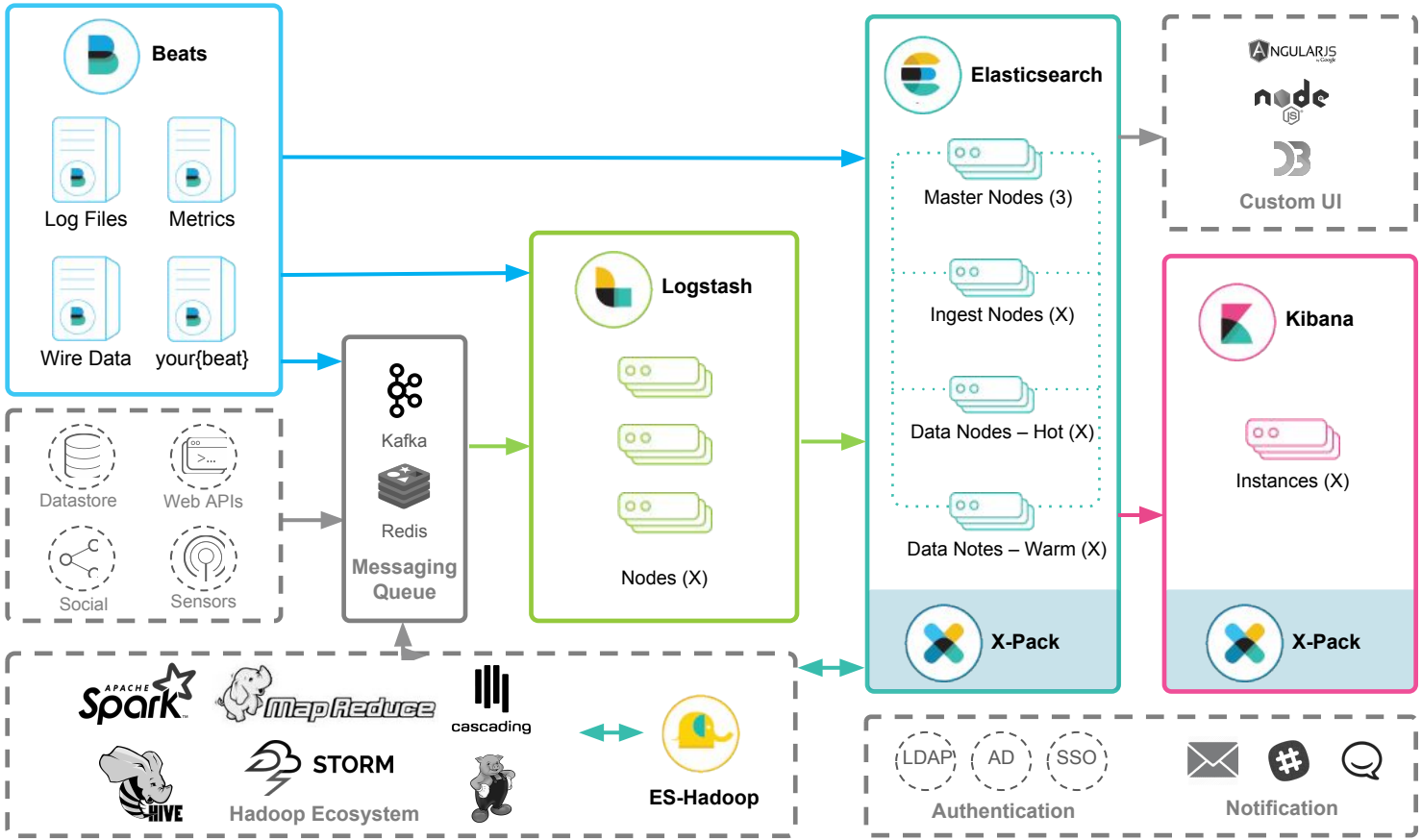
elastic

Query DSL

Geo queries

```
GET /example/_search
{
    "query":{
        "bool": {
            "must": {
                "match_all": {}
            },
            "filter": {
                "geo_shape": {
                    "location": {
                        "shape": {
                            "type": "envelope",
                            "coordinates" : [[13.0, 53.0], [14.0, 52.0]]
                        },
                        "relation": "within"
                    }
                }
            }
        }
    }
}
```

elastic


Beats
Log Files
Metrics
Wire Data
your{beat}
Datastore
Web APIs
Social
Sensors
Kafka
Redis
Messaging Queue
Logstash
Nodes (X)
Elasticsearch
Master Nodes (3)
Ingest Nodes (X)
Data Nodes – Hot (X)
Data Notes – Warm (X)
X-Pack
ANGULARJS
node
Custom UI
Kibana
Instances (X)
X-Pack
APACHE Spark
MapReduce
cascading
HIVE
STORM
Hadoop Ecosystem
ES-Hadoop
LDAP
AD
SSO
Authentication
Notification

elastic

Capacity Planning

It depends...

elastic

https://www.elastic.co/elasticon/conf/2016/sf/quantitative-cluster-sizing

Capacity Planning

What is your use case?

- Full text search
- Logging/Metrics
- Complex Aggregations with lot of users

Each use case needs a different cluster configuration.

elastic

https://www.elastic.co/elasticon/conf/2016/sf/quantitative-cluster-sizing

Capacity Planning

Let us take Logging..

- Inflow of data per day
  - Per day : 10GB
  - Per Month : 300GB
  - Per Year: 3600GB
- Data Retention
  - 15 days
- High Availability (Replication factor)
  - 1 i.e., 7200GB Per Year
- Type of Queries

Master Node : X

Data Node : X

elastic

https://www.elastic.co/elasticon/conf/2016/sf/quantitative-cluster-sizing

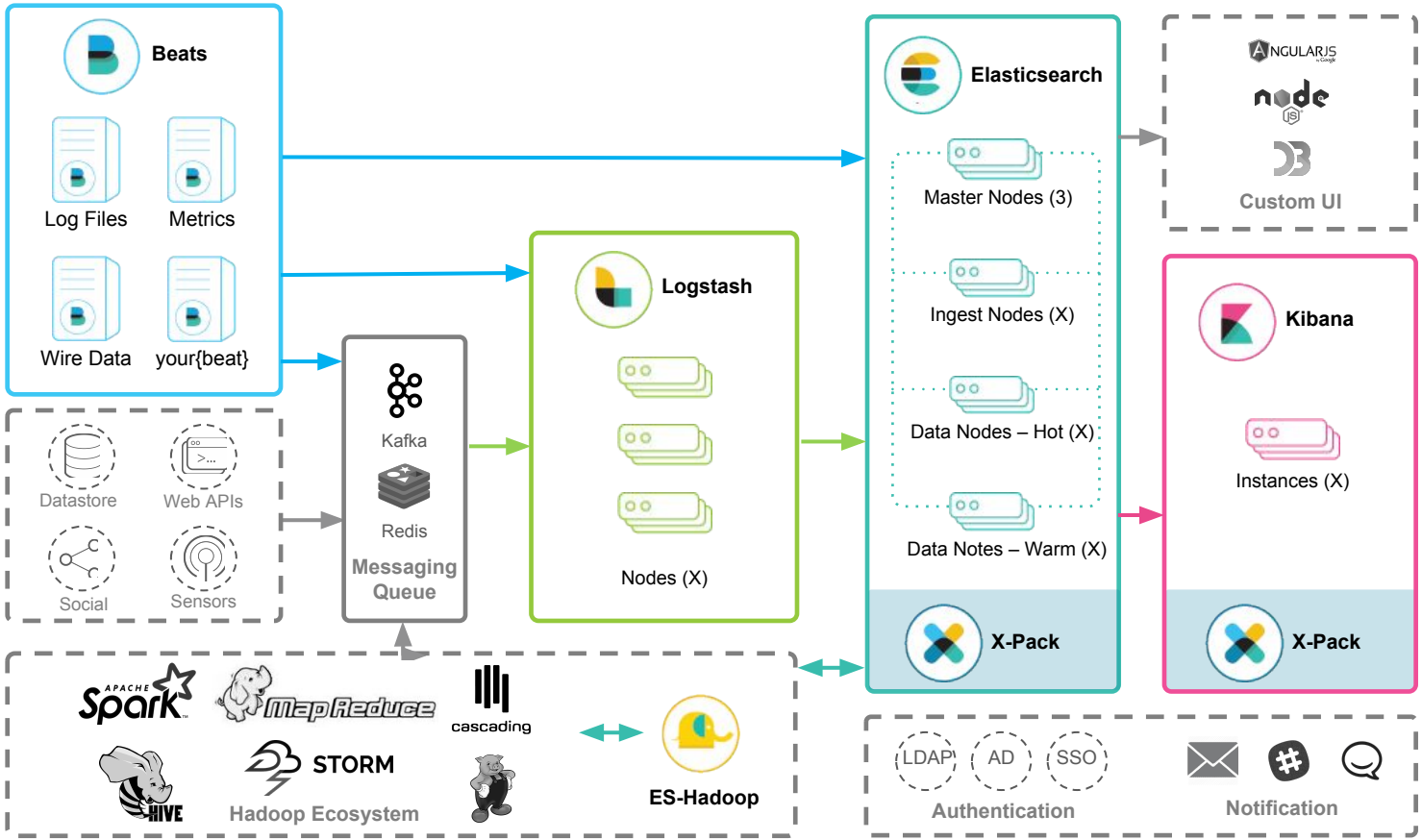
Capacity Planning

Hardware Recommendations

- SSD’s are the best
- Local Disk is king!
- Prefer Medium size machine’s over Large size machine’s
- Only 50% of your RAM to Elasticsearch
- Don’t Cross 32GB Java Heap Space

elastic

https://www.elastic.co/elasticon/conf/2016/sf/quantitative-cluster-sizing



https://www.elastic.co/blog/hot-warm-architecture-in-elasticsearch-5-x

elastic

Certification

Elastic Certified Engineer
Learn More

training.elastic.co

Instructor-Led Training

Elasticsearch Engineer I
Learn More

Elasticsearch Engineer II
Learn More

Elastic Stack Data Administration I
Learn More

Kibana Data Analysis I
Learn More

Elastic Security Analytics
Learn More

On-Demand Training

X-Pack: Security
Learn More

X-Pack: Machine Learning
Learn More

Elastic Cloud Enterprise Administration I
Learn More

elastic

Resources

- https://www.elastic.co/learn
- https://www.elastic.co/blog/category/engineering
- https://discuss.elastic.co/
- https://fb.com/groups/ElasticIndiaUserGroup
- https://elastic.co/community

Fin!

discuss.elastic.co | aravind@elastic.co | @aravindputrevu