

Elasticsearch

Scott Stirling & Hernán Ibarra

What is Elasticsearch

- It is a horizontally distributed NoSQL (non-relational) database/ search engine.
- Elasticsearch was created by Shay Bannon and was released in February 2010.
- To make search and update the database elastic search uses a RESTful API. To query we use HTTP VERBS - GET, POST, PUT, DELETE, etc.
- Used in text-search, business analytics and geo-spatial analysis.

Overview

- Documents == JSON files
- Inverse index (because querying faster)
- Shards (to horizontally scale faster)
- Nodes (to use different machines)
- Cluster
- Indexes (templates/namespaces for different types of data, address within the cluster)
- replicas (for redundancy)

The Structure of Elasticsearch

In Elasticsearch data is organised in the following structure:

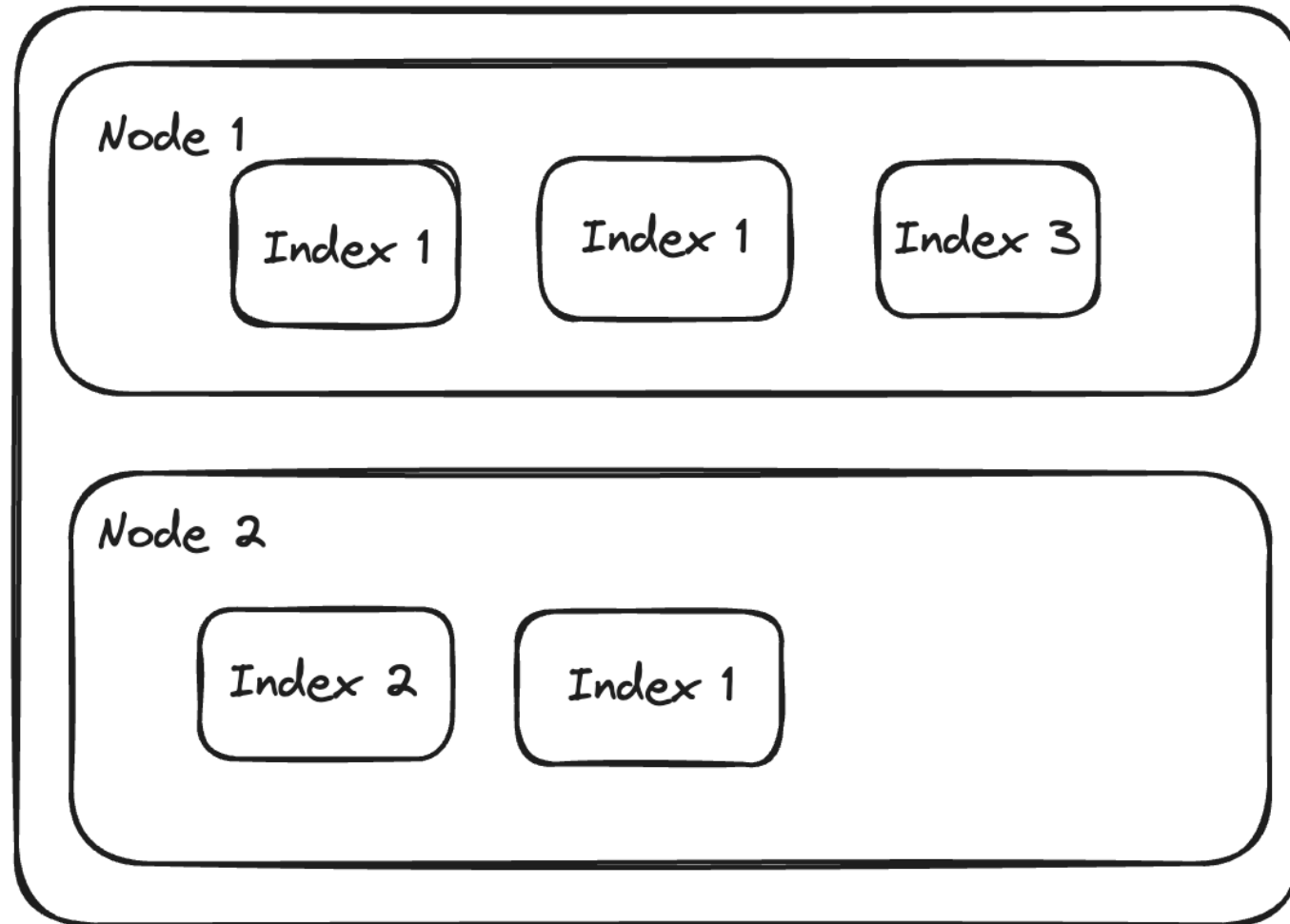
- First we have indices which are analogous to tables in a relational database.
- Each of these are comprised of shards which are parts of an index (more on the next slide).
- Then we have documents which are individual pieces of JSON.

Shards & Nodes

So how does this help us?

- We can then split the index shards and put them on nodes.
- Meaning we can search indices in parallel.

Structure



The CAP Theorem

- Back in 2010, Shay Bannon said that Elasticsearch gives up on partition tolerance.
- We found an Elastic team member in 2019, saying that the CAP theorem technically doesn't apply and so there are times where it behaves like CP and others where it behaves like AP.

Licensing

7.10 and Earlier

BINARY DISTRIBUTION
Elastic License

SOURCE CODE

Platinum+
(opt in)

Source Code:
Elastic License

Free / Basic

Source Code:
Elastic License

Free

Source Code:
Apache 2.0

7.11 – 8.15

BINARY DISTRIBUTION
Elastic License 2.0

SOURCE CODE
Elastic License

Platinum+
(opt in)

Source Code:
Elastic License 2.0

Free / Basic

Source Code:
Elastic License 2.0

Free

Source Code:
SSPL 1.0 or
Elastic License 2.0

8.16+

BINARY DISTRIBUTION
Elastic License 2.0

SOURCE CODE
Elastic License

Platinum+
(opt in)

Source Code:
Elastic License 2.0

Free / Basic

Source Code:
Elastic License 2.0

Free / OSS

Source Code:
AGPL v3, SSPL 1.0 or
Elastic License 2.0

What is OpenSearch

- A fork of Elasticsearch created in 2021.

Who uses it and why?

- Uber uses it for searching and storing geospatial data (2017).
- Netflix uses ElasticSearch for its recommendation system (2021).

Thanks for Listening

Any Questions?