Elasticsearch: A Game Changer

Guillaume Delporte, Aly Abdelaleem, Jordi Hoorelbeke

24/05/2023

Objective

Gain a **comprehensive understanding of Elasticsearch**, including its **advanced capabilities**, and discover **how leading companies leverage its powerful features** for various applications.

24/05/2023

What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Advanced Search Techniques** with Elasticsearch.
5. **Kibana** and Elasticsearch: The Perfect Combination.
6. **Conclusion**: Key Takeaways on Elasticsearch.

What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Advanced Search Techniques** with Elasticsearch.
5. **Kibana** and Elasticsearch: The Perfect Combination.
6. **Conclusion**: Key Takeaways on Elasticsearch.

An **Introduction** to Elasticsearch.

What is elasticsearch?

- Elasticsearch: Distributed, Open-Source search, and analytics engine.
- Handles various types of data: textual, numerical, structured, unstructured.
- Designed for scalability and can search and index diverse document formats.
- Created by Shay Banon between 2004-2009

Shay Banon, Elasticsearch's creator

elasticsearch and APACHE LUCENE™

- Based on Apache Lucene.
- Lucene: Free and open-source information retrieval software library.
- Lucene developed by Doug Cutting.
- Elasticsearch leverages Lucene's capabilities for full-text search.

Doug Cutting,
Open-Source software developer

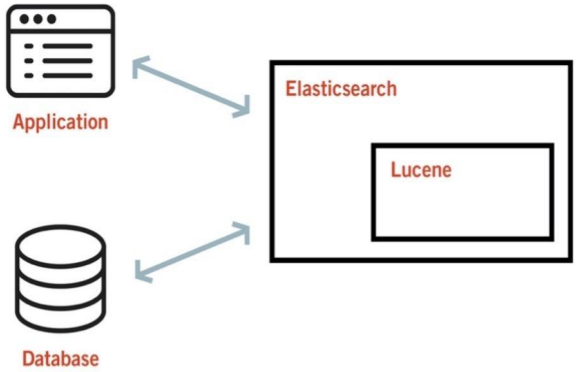


Figure from : https://www.endava.com/en/blog/Engineering/2021/Elasticsearch-and-apache-lucene-fundamentals-behind-the-relevance-score

How Does elasticsearch Work?

- Data stored as JSON (JavaScript Object Notation).
- Document organized in indices, similar as in relational databases.
- Indices further divided into shards, distributed across cluster nodes.

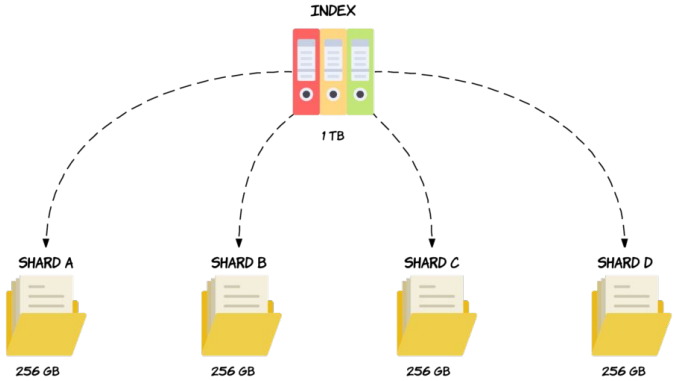
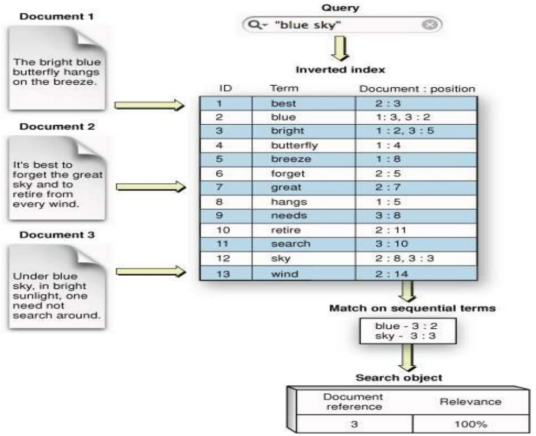


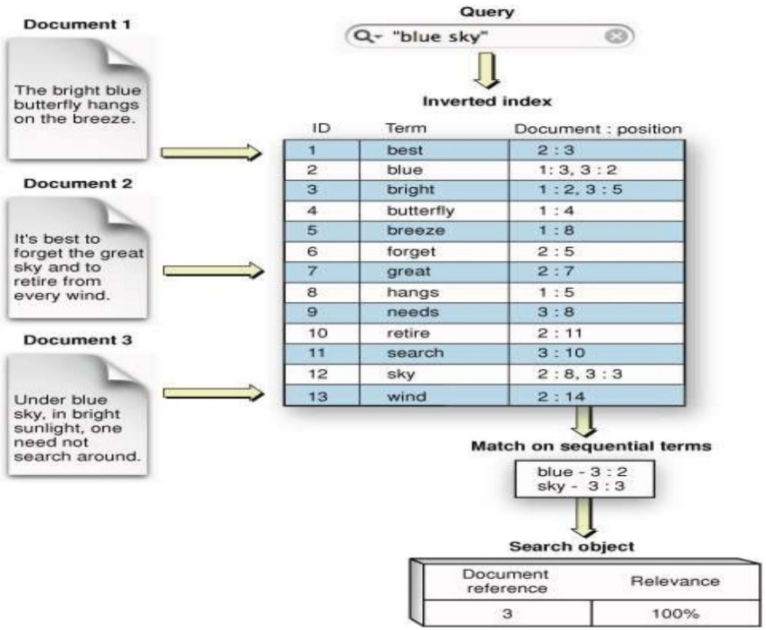
Figure from : https://codingexplained.com/coding/elasticsearch/understanding-sharding-in-elasticsearch



Credit: https://developer.apple.com/library/mac/documentation/userexperience/conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html

Inverted Index for Speed and Flexibility

- elasticsearch uses an inverted index for fast search.
- Inverted index stores unique words and documents they appear in.
- Enables quick lookup and retrieval of search results.



Credit: https://developer.apple.com/library/mac/documentation/userexperience/conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html

elasticsearch and the elastic stack

- Elasticsearch is part of the Elastic Stack.
- Elastic Stack components: Elasticsearch, Logstash, Kibana, Beats.
- Logstash: Data processing pipeline for ingesting and transforming data.
- Kibana: Visualization and analysis layer for Elasticsearch data.

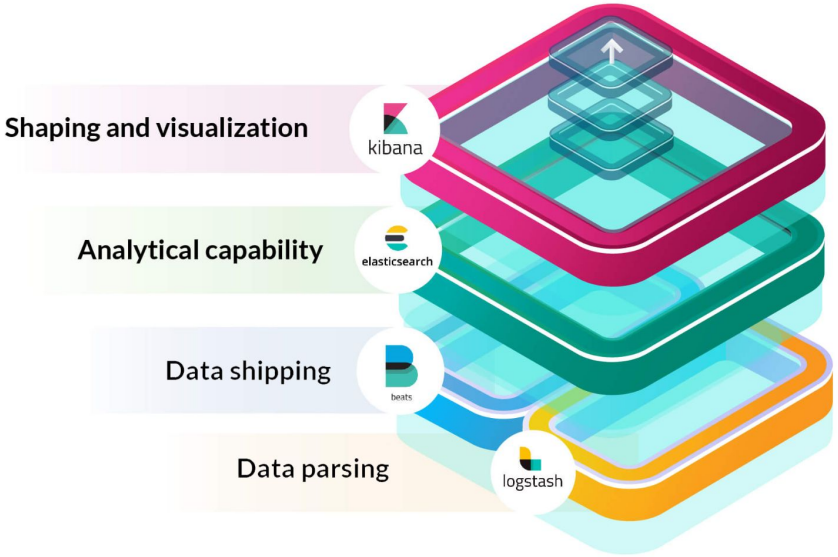
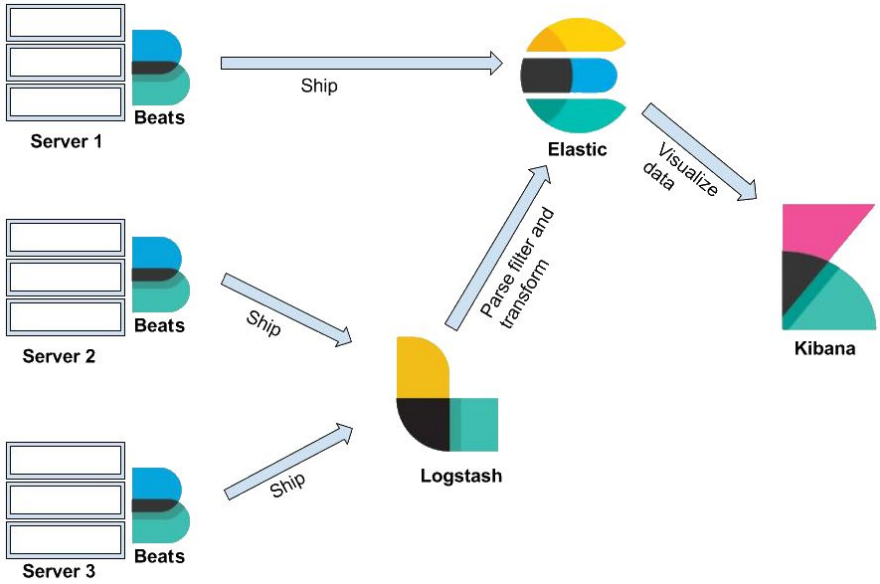


Figure from: https://quintagroup.com/services/the-elastic-stack-and-its-components-elasticsearch-kibana-logstash-and-beats

The elastic stack (Cont'd.)

- Beats: Lightweight data shippers for sending operational data.
- Beats can send data directly to Elasticsearch or via Logstash for processing.
- The Elastic Stack provides a comprehensive data management and analysis solution.

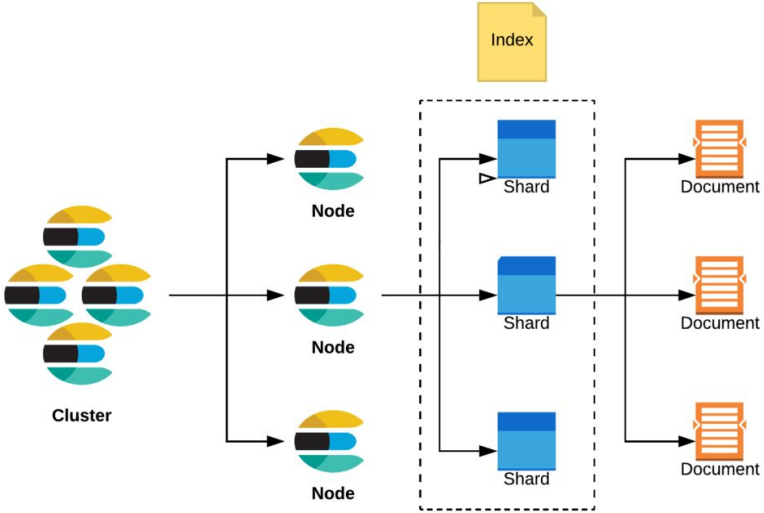


What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Advanced Search Techniques** with Elasticsearch.
5. **Kibana** and Elasticsearch: The Perfect Combination.
6. **Conclusion**: Key Takeaways on Elasticsearch.

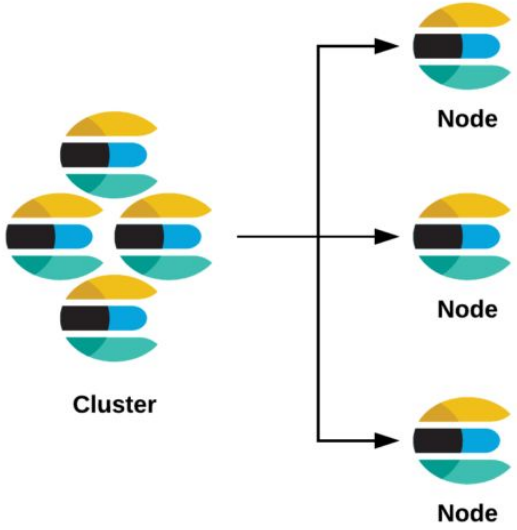
Deep Dive into elasticsearch Architecture

- A cluster contains one or more nodes.
- A node is an instance of Elasticsearch
- Index is used to group shards that contain related data.
- Shard is where data is stored.
- Data is stored as documents.



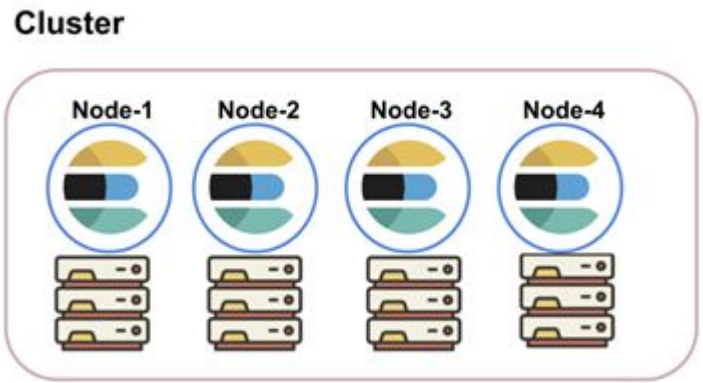
Cluster

- A cluster is a collection of nodes that work together to achieve a common goal.
- Each cluster has a unique name.
- When creating a node, one cluster comes with it by default.
- The cluster is managed by a master node.



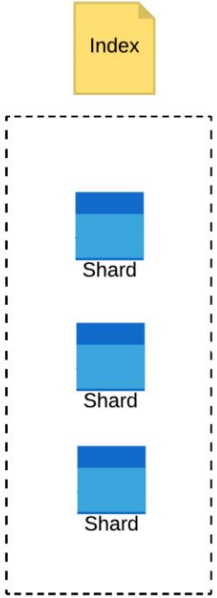
Node

- A node is an instance of Elasticsearch with a unique id and a name.
- A node must belong to a cluster.
- Nodes are recommended to be distributed across several machines, but you can have multiple nodes on one machine.
- A node can have multiple roles.



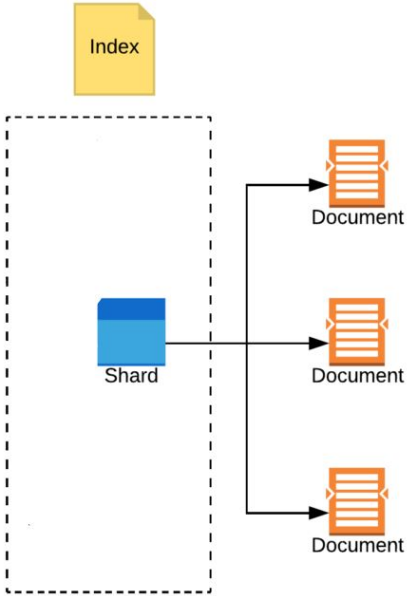
Index

- Index is a logical way to group shards that contain documents that are related to each other.
- Similar to a database in traditional systems.
- When creating an index, one shard comes with it by default.
- Index can be configured to allow multiple shards across multiple nodes (Sharding).



Shard

- Shard is actually where data is stored on disk and where searches are run.
- A shard can be:
  - Primary shard
  - Replica shard
- Elasticsearch allows you to configure the number of replica shards per primary shard.
- Benefits of sharding: horizontal scaling, parallel queries, backup, and load balancing using replica shards.



Document

- Data is stored as documents, which is a JSON object that is stored with a unique id.
- Documents are schema flexible.
- Documents have version number which are incremented with each update.


```
{
  "_index": "blog",
  "_id": "123456",
  "_version": 2,
  "_source": {
    "title": "Introduction to Elasticsearch",
    "author": "John Doe",
    "publish_date": "2022-01-15",
    "content": "Elasticsearch is a scalable search and analytics engine built on Apache Lucene. It offers real-time search capabilities, making it popular for log analytics, full-text search, and more. Its flexible data model and distributed architecture support powerful querying and analysis."
  }
}
```

What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Advanced Search Techniques** with Elasticsearch.
5. **Kibana** and Elasticsearch: The Perfect Combination.
6. **Conclusion**: Key Takeaways on Elasticsearch.

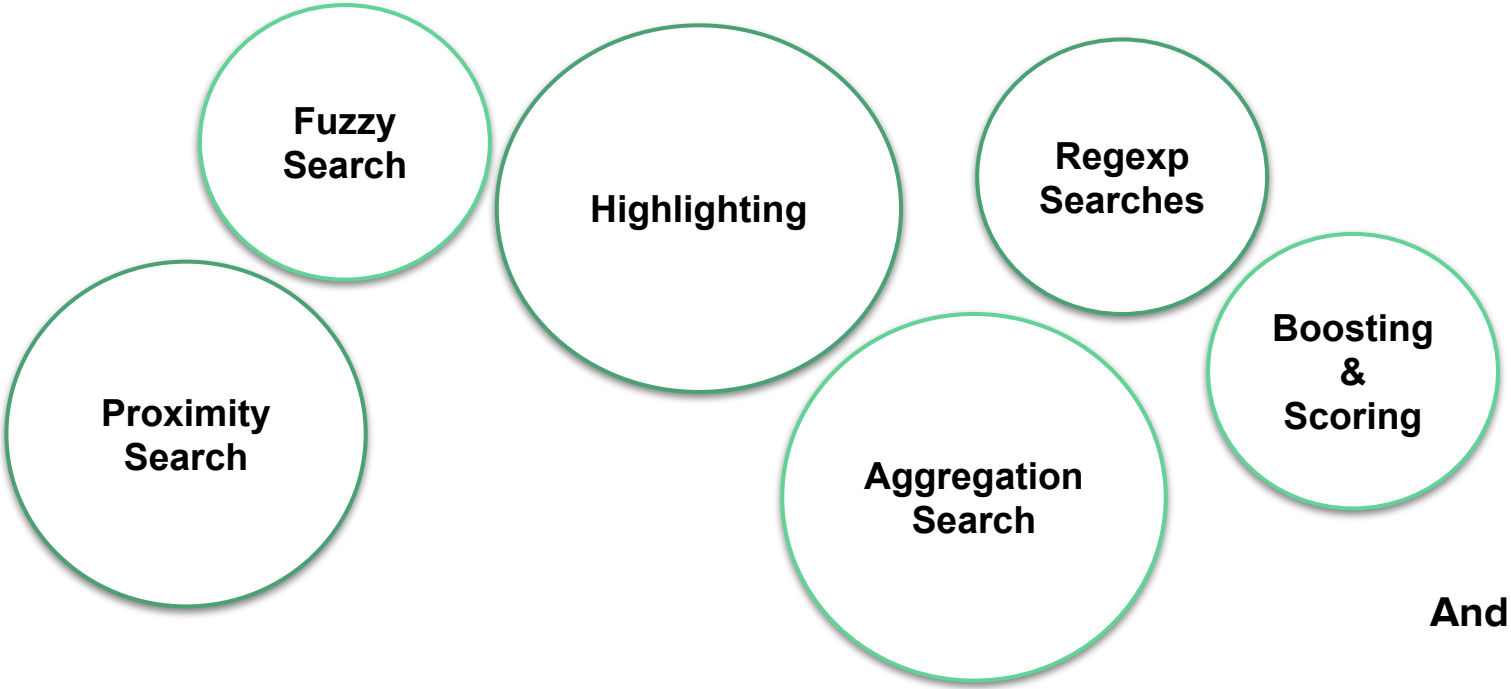
Elasticsearch in Action: **How Companies Utilize It**.

- Vodafone (EG) used it as a log analytics and reporting tool for their Website and App (personal experience).
- eBay's search infrastructure is powered by Elasticsearch help them deliver relevant results.
- NASA uses Elasticsearch to analyze and search through large volumes of telemetry data collected from satellites and space missions.
- Uber uses Elasticsearch to power real-time geospatial search.

What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Advanced Search Techniques** with Elasticsearch.
5. **Kibana** and Elasticsearch: The Perfect Combination.
6. **Conclusion**: Key Takeaways on Elasticsearch.

Advanced Search Techniques with Elasticsearch.



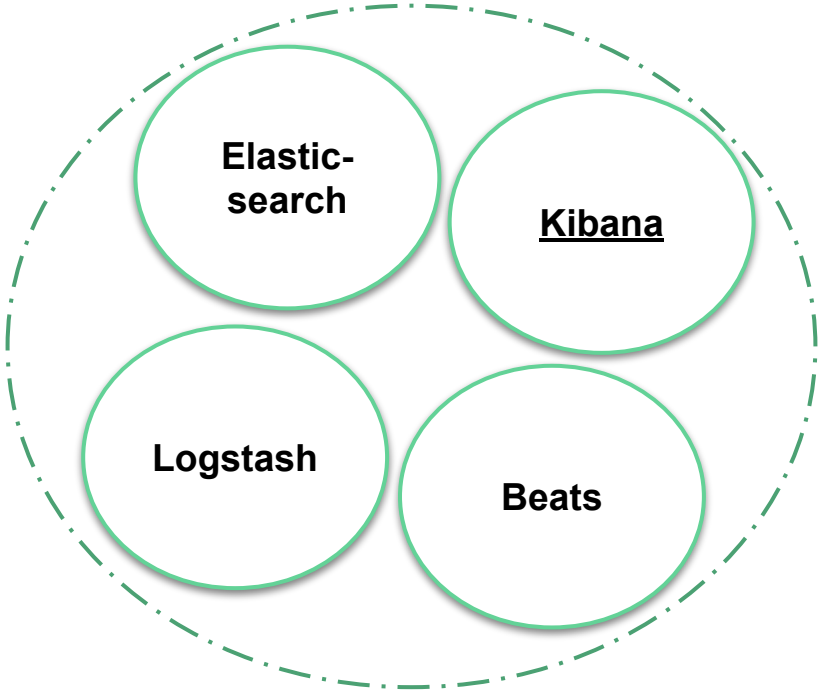
And many more...

What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Advanced Search Techniques** with Elasticsearch.
5. **Kibana** and Elasticsearch: The Perfect Combination.
6. **Conclusion**: Key Takeaways on Elasticsearch.

Kibana and Elasticsearch: The Perfect Combination.

Elastic Stack



What is Kibana?

- **Visualization** capabilities on top of indexed content.
- **Geospatial analysis** tool for creating and layering maps.
- **Development tools** for developers to interact with the Elasticsearch REST API.
- Users can create and manage **alerts** that notify them when real-time data meets certain conditions.

Kibana and Advanced Search: A **Demo**.

What We Cover.

1. An **Introduction** to Elasticsearch.
2. Deep Dive into Elasticsearch's **Architecture**.
3. Elasticsearch in Action: **How Companies Utilize It**.
4. **Kibana** and Elasticsearch: The Perfect Combination.
5. **Advanced Search Techniques** with Elasticsearch.
6. **Conclusion**: Key Takeaways on Elasticsearch.

Conclusion: Key Takeaways on Elasticsearch

- **Elasticsearch**: A potent, distributed search and analytics engine, underpinned by Apache Lucene.
- **Distributed Architecture**: Ensures high availability and resilience, with a structure composed of nodes, indices, shards, and replicas.
- **Elastic Stack Synergy**: Elasticsearch operates in concert with Kibana, Logstash, and Beats to provide a holistic data analysis solution.

Key Takeaways on Elasticsearch (Cont'd)

- **Advanced Search**: Employs sophisticated techniques such as fuzzy search, proximity matching, boosting, and aggregations for precise and intricate searches.
- **Customizable Scoring**: Allows tailoring of relevance scoring to cater to specific needs.
- **Versatility**: Used for a broad range of applications, including full-text search, data analysis, log, and event data management.
- **Scalability and Resilience**: Its ability to scale and remain resilient makes Elasticsearch a preferred choice among various organizations.

Elasticsearch: Going further

Official Elasticsearch Documentation
Tutorialspoint's Elasticsearch tutorial
Tutorial series created by LisaHJung

Thank you for your attention!