Elasticsearch, Solr, and Lucene

Table of Contents

What is Elasticsearch?

Elasticsearch is a **Distributed**, **Free**, **Open search**, **Analytics and Full-text search engine**. Its often used for enabling search functionality for various types of *data*. Amazon AWS has integrated with Elasticsearch to make **Amazon Elasticsearch Service** that makes it easy to **Deploy**, **Operate** and **Scale** Elasticsearch in the AWS Cloud.

- You can imagine this when you shop using amazon. We search for an item in a search box and that's an **Unstructured search**.

- When you click on any item recommended by Amazon on the home page, that is a **Structured search**.

- If you **Sort** the items from *cheap to high* on the products page as I do, that is facilitated using Elasticsearch.
- Now the items have filled the page and we need to click on **Next** to go to the next page. This feature is also facilitated using Elasticsearch.

Elasticsearch is built on **Apache Lucene** and was first released in 2010 and is known for its **Simple REST APIs**, **Distributed nature**, **Speed**, and **Scalability**.

What is it used for?

- Application search
- Website search
- Enterprise search
- Logging and log analytics
- Infrastructure metrics and container monitoring
- Application performance monitoring
- Geospatial data analysis and visualization
- Security analytics
- Business analytics

Why use Elasticsearch?

- Elasticsearch offers REST APIs, A simple HTTP interface and uses schema-free JSON documents, Making it easy to get started and quickly build applications for a variety of use-cases.
- **Elasticsearch is distributed**. The distributed nature of Elasticsearch enables it to process large volumes of data in **parallel**, Quickly finding the best matches for your queries.

- **Complimentary tools and plugins**. Elasticsearch comes integrated with **Kibana**, *A popular visualization and reporting tool*. It also offers integration with **Beats** and **Logstash**, While enabling you to easily transform source data and load it into your Elasticsearch cluster.
- Elasticsearch supports a variety of **programming languages** and Elasticsearch supports **34** majorly text languages. Some of them are:
    - Java
    - Python
    - Go
    - .Net
    - PHP
    - Perl
    - Ruby

- Elasticsearch comes with a wide set of features. In addition to its **Speed**, **Scalability**, and **Resiliency**, Elasticsearch has several powerful built-in features that make storing and searching data even more efficient such as data rollups and index lifecycle management.

Let's understand some queries here. Curl is a command-line tool to transfer data to or from a server. Here, We are requesting a search in a products index for a type of product which is a thinkpadx240 laptop. This is just a simple query. More to it can be found here.

```
curl -X GET 'localhost:9200/products/product/_search?q=thinkpadx240'
```

> Some Technology companies using Solr are Github, Uber, Shopify, Facebook, Udemy, Slack, etc.

What is Solr?

- Solr is a JAVA-based Open source enterprise search client. It's built around Lucene which is a high-performance search engine library.
- Solr is the popular, **Blazing-fast**, **Open source enterprise search platform**, Built on Apache Lucene™. It could is integrated with**Apache Hadoop** and therefore is also capable of using with **Big data**.
- Solr is a **Standalone enterprise search server** with a REST-like API. You put documents in it via JSON, XML, CSV or binary over HTTP.

Features of Solr

- Advanced **Full-text search capabilities**: Searching document in a full-text database by using **Phrases**, **Wildcards**, **Joins**, **Grouping** and much more across any data type.
- Optimized for **High volume traffic**: Solr is proven helpful at extremely large scales of data coming in.
- **Near real-time operations**: Tasks such as **reading** or **writing** data usually take less than a second to complete. This lets you use Elasticsearch for near real-time use cases such as **Application monitoring** and **Anomaly detection**. This makes Solr operations use **less** time.
- Comprehensive **Administration interfaces**: Solr ships with a built-in, Responsive administrative user interface to make it easy to control your Solr instances. Solr also has **Database integrations** like MongoDB, HBase, Cassandra.
- **Highly scalable** and **fault-tolerant**: Solr makes it easy to scale up and down. Solr also features *Replication*, *Distribution*, *Rebalancing* and *Fault tolerance*. Solr can also accommodate rich document handling.
- **Flexible and adaptable** with easy configuration: Solr's is designed to adapt to your needs all while simplifying configuration.

> Some Technology companies using Solr are Flipkart, Instagram, Intuit, Apple, Chegg, eBay, etc.

What is Lucene?

It is a powerful JAVA search library that lets you easily add search or retrieval of information to applications. It is not a server, But an embedded library which is mostly misunderstood.

The structure of Apache Lucene shown in various versions

Why use Lucene?

- **Scalable** and **high-performance indexing**.
- **Powerful**, **Accurate** and efficient search algorithms.
- **Open-source**, Hence scalable.

- Implementations of other programming languages are also available.

> Some companies using Lucene are Nike, Walmart, Peraton, etc.

Conclusion

Considering switching to a different database for a full-text searching to improve performance, We could go for Elasticsearch because of its innumerable benefits.

- Elasticsearch is cloud-ready out of the box which makes it reliable in times of systems failure.
- Elasticsearch runs a search index on multiple servers which makes it reliable in times of systems failure.
- Performance-wise, Both Elasticsearch and Solr are the same.
- Elasticsearch is simpler to work with. Unlike solr, Elasticsearch is a single process with real-time updates making it engaging and helpful in tracking status.

References

- What is ElasticSearch
- Elasticsearch - Beyond Full-text Search Alex Reelsen - GOTO 2013
- Amazon ElasticSearch Service
- Why ElasticSearch
- All about ElasticSearch
- How can I use ElasticSearch
- Solr introduction
- Features of Solr
- System design of Solr
- Difference between Solr and lucene
- Amazon Elastic Service