



Elasticsearch

Duration 3 day(s) (ELASTICSEARCH-03)

Know how to use and configure Elasticsearch, the second generation search engine

Description

Elasticsearch is a new generation Open Source search and indexing engine. Based on the Apache Lucene library, this search engine offers advanced features such as searches by geographic coordinates, analysis and categorization aggregations, filtering results or searching on several different indexes and types of documents. Designed for the Cloud, Elasticsearch has been specifically designed to index very large volumes of data while ensuring high performance scalability and high fault tolerance. This course covers key concepts of search engines before detailing the basic features of Elasticsearch and its architectural principles. Based on concrete examples and actual use cases, the training provides all the necessary knowledge to use and operate Elasticsearch effectively and to solve the problems most often encountered.

Goals

- Identify the issues and keys to success of a search engine
- Learn how to use the elasticsearch API to index and search documents
- Know how to install elasticsearch and perform cluster configuration
- Integrate elasticsearch into an existing application
- Understanding how elasticsearch simplifies the use of Apache Lucene
- Use elasticsearch client and Java API to index and search documents
- Monitor elasticsearch and understand usage statistics

Public

- Architect
- Developer

Prerequisites

- Have knowledge of what are REST / HTTP
- Have knowledge of Json, Yaml format
- Have minimum knowledge of Linux and command lines including Curl, wget ...
- Have minimum knowledge on the network
- **Mandatory for the Java chapter:** Knowledge of the Java language and its ecosystem (maven, IDE)

If you do not know Java, the chapter can be ignored if others do not request it.

Non-compulsory knowledge greatly reduces practical work when it is all to be acquired.

Structure

60% Theory, 40% Practice

Introduction to search engines

- General (Needs, uses, key features of search engines)
- Process of implementing a search engine
- Panorama of open source search solutions

Get started with Elasticsearch

- Principles of cluster operation
- Notions of node, index, partitioning, document ...
- Installation
- Configuration
- REST API overview
- First indexing
- First search

Indexing documents

- Design of the index and its documents
- Index or delete documents with the Rest API
- Mass indexing
- Optimistic version and management of the competition
- Presentation of Lucene storage and refresh
- Other features (routing, consistency, child document, ...)

Mapping

- Definition and role of mapping
- Field types
- Preset fields
- Metadata of an Index

Text analysis and extraction

- The basis of extraction and text analysis
 - Analyzers
 - Char filters
 - Tokenizers
 - Token Filters
- Configuration and use of predefined or customized Analyzers
- Text analysis verification API
- Use case
 - Deleting words
 - Extraction of email addresses and urls
 - Removing HTML tags in text
 - Text analysis in multiple languages
 - Rooting

Search for documents

- Find documents with the Rest API
- Results management
- Search queries
 - Match_all type query

- Request of type query_string / simple_query_string
- Match type query (and derivatives)
- Term and terms query
- Wildcard request
- Range request
- Fuzzy request
- Request of type bool
- Request of type exist / missing
- Request of type and, or and not
- Query of type _type and _id
- Different queries and filters
- Benefits of filters

Advanced search features

- Management of relevance and score
- Suggestions
- Autocompletion
- Highlighting
- Search Objects / Nested / Parent-Child
- More Like This
- Geospatial research

Clients and Java API

- Types of customers available
- Using the Java API
 - Index and search documents
 - Manage indexes
 - Manage mappings
- Integrate elasticsearch into a Java application (embedded, client node, client transport)

Cloud and Clustering

- Principle of functioning of an elasticsearch cluster
- Prevent split brains
- Configure an elasticsearch cluster
- Partitioning and scalability and volumetry of data
- Supervision of an elasticsearch cluster:
 - the health API
 - the state API
 - the stats API
 - the pending_task API
 - the API _nodes alias node info
 - the hot_threads API
 - supervision plugins
 - logs
 - the _cat API
 - Backup and restore an elasticsearch cluster

Advanced Features

- Other types of elasticsearch plugins
- Scripting
- Preheat your cluster
- Percolation