



Elasticsearch

Durée 3 jour(s) (ELASTICSEARCH-03)

Savoir utiliser et configurer Elasticsearch, le moteur de recherche seconde génération

Description

Elasticsearch est un moteur de recherche et d'indexation Open Source nouvelle génération. Basé sur la librairie Apache Lucene, ce moteur de recherche offre des fonctionnalités avancées telles que les recherches par coordonnées géographiques, l'analyse et la catégorisation par agrégations, le filtrage de résultats ou encore la recherche sur plusieurs index et types de documents différents. Taillé pour le Cloud, Elasticsearch a été spécialement conçu pour indexer de très gros volumes de données tout en assurant une montée en charge performante et une forte tolérance aux pannes. Cette formation aborde les concepts clés des moteurs de recherche avant de détailler les fonctionnalités de base d'ElasticSearch et ses principes d'architecture. En s'appuyant sur des exemples concrets et des cas réels d'utilisation, la formation apporte toutes les connaissances nécessaires pour utiliser et exploiter ElasticSearch de façon efficace et résoudre les problématiques les plus souvent rencontrées.

Objectifs

- Identifier les enjeux et les clés de succès d'un moteur de recherche
- Apprendre à utiliser l'API d'elasticsearch pour indexer et rechercher des documents
- Savoir installer elasticsearch et effectuer sa configuration en cluster
- Intégrer elasticsearch dans une application existante
- Comprendre comment elasticsearch simplifie l'utilisation d'Apache Lucene
- Utiliser le client et l'API Java d'elasticsearch pour indexer et rechercher des documents
- Monitorer elasticsearch et comprendre les statistiques d'utilisation

Public

- Architecte
- Développeur

Prérequis

- Avoir des connaissances de ce que sont REST/HTTP
- Avoir des connaissances du format Json, Yaml
- Avoir des connaissances minimum de Linux et des lignes de commandes et notamment Curl, wget ...
- Avoir des connaissance minimum sur le réseau
- **Obligatoire pour le chapitre Java:** Connaissances du langage java et de son écosystème (maven, IDE)

En cas de non connaissance de Java, le chapitre peut être ignoré si d'autres n'en font pas la demande.

Les connaissances non obligatoires ralentissent fortement les travaux pratiques lorsqu'elles sont toutes à acquérir.

Répartition

60% Théorie, 40% Pratique

Programme

Introduction aux moteurs de recherche

- Généralités (Besoins, usages, fonctionnalités clefs des moteurs de recherche)
- Processus de mise en oeuvre d'un moteur de recherche
- Panorama des solutions de recherche Open Source

Premiers pas avec Elasticsearch

- Principes de fonctionnement en cluster
- Notions de node, d'index, de partitionnement, de document...
- Installation
- Configuration
- Vue générale de l'API REST
- Première indexation
- Première recherche

Indexation de documents

- Conception de l'index et de ses documents
- Indexer ou supprimer des documents avec l'API Rest
- Indexation en masse
- Version et gestion optimiste de la concurrence
- Présentation du stockage Lucene et refresh
- Autres fonctionnalités (routing, consistency, document enfant, ...)

Mapping

- Définition et rôle du mapping
- Types de champs
- Champs prédéfinis
- Méta données d'un Index

Analyse et extraction de texte

- La base de l'extraction et analyse de texte
 - Analyzers
 - Char filters
 - Tokenizers
 - Token Filters
- Configuration et utilisation des Analyzers prédéfinis ou customisés
- API de verification de l'analyse de texte
- Cas d'utilisation
 - Suppression de mots
 - Extraction d'adresses email et d'urls
 - Suppression de balises HTML dans du texte
 - Analyse de texte en plusieurs langues
 - Racinisation

Recherche de documents

- Rechercher des documents avec l'API Rest
- Gestion des résultats
- Les requêtes de recherche
 - Requête de type match_all
 - Requête de type query_string/simple_query_string
 - Requête de type match (et dérivées)
 - Requête de type term et terms
 - Requête de type wildcard
 - Requête de type range
 - Requête de type fuzzy
 - Requête de type bool
 - Requête de type exist/missing
 - Requête de type and, or et not
 - Requête de type _type et _id
- Différentier requêtes et filtres
- Avantages des filtres

Fonctionnalités avancées de recherche

- Gestion de la pertinence et du score
- Suggestions
- Autocomplétion
- Highlighting
- Recherche Objets / Nested / Parent-Enfant
- More Like This
- Recherche géospatiale

Clients et API Java

- Types de clients disponibles
- Utilisation de l'API Java
 - Indexer et recherche des documents
 - Gérer les index
 - Gérer les mappings
- Intégrer elasticsearch dans une application Java (embarqué, noeud client, transport client)

Cloud et Clusterisation

- Principe de fonctionnement d'un cluster elasticsearch
- Prévenir les split brains
- Configurer un cluster elasticsearch
- Partitionnement et scalabilité et volumétrie de données
- Supervision d'un cluster elasticsearch :
 - l'API health
 - l'API state