

Formation Elasticsearch

Durée

2 jours (14 heures)

Présentation

[Elasticsearch](#) est un moteur de recherche et d'indexation Open Source nouvelle génération. Sa nature distribuée et sa capacité à être résilient et hautement disponible ont déjà séduit des acteurs majeurs tels que Wikipedia, LinkedIn, Netflix, Ebay, ou WordPress. 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 des moteurs de recherche avant de détailler les fonctionnalités de base d'Elasticsearch. La formation apporte toutes les connaissances nécessaires pour utiliser et exploiter Elastic Search de façon efficace et résoudre les problématiques les plus souvent rencontrées.

La formation utilisera la dernière version stable en date du projet ([Elasticsearch 7](#) à ce jour).

Objectifs

- Apprendre à utiliser Elasticsearch, de l'installation à son intégration en passant par sa configuration
- Utiliser le client et l'API Java d'Elasticsearch pour indexer et rechercher des documents
- Monitorer elasticsearch et comprendre les statistiques d'utilisation

Public visé

Architecte, Développeur, Chef de projet, Big Data

Pré-requis

Connaissances en Java & Linux, avoir idéalement suivi [notre formation sur ElasticStack](#)

Pour aller plus loin

Programme de la formation Elasticsearch

Introduction aux moteurs de recherche

- Généralités
- Panorama des solutions de recherche Open Source
- Intégrer un moteur de recherche dans une application
- Enjeux et clés de succès
- Présentation des nouveautés de la [version 6 & 7](#)

Introduction à Elasticsearch

- Historique du projet
- Apache Lucene
- Ce qu'apporte Elasticsearch par rapport à Lucene

Les bases d'Elasticsearch

- Principes de fonctionnement en cluster
- Installation
- Configuration
- Notion de Node, d'Index et de Type
- Partitionnement de données
- Vue générale de l'API Rest

Indexation de documents

- Conception de l'index et de ses documents
- Indexer ou supprimer des documents avec l'API Rest
- Indexation en masse
- Version
- Pondération
- Autres fonctionnalités (routing, consistency, document enfant, ...)

Mapping

- Définition et rôle du mapping
- Type de champs
- Les champs prédéfinis
- Méta données d'un Index

Analyse et extraction de texte

- Base de l'extraction et analyse de texte
 - Analysers
 - Char filters
 - Tokenizers
 - Token Filters
- Cas d'utilisation
- Configuration et utilisation des Analyzers prédéfinis ou customisés
- Analyse de texte en plusieurs langues
- Suppression de mots
- Extraction d'adresses email et d'urls
- Suppression de balises HTML dans du texte
- Mise en place de la correction orthographique
- Indexation de fichiers binaires (utilisation d'Apache Tika)

Recherche de documents

- Rechercher des documents avec l'API Rest
- Gestion des résultats
- Types de requêtes
 - 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
- Types de filtres
- Combiner efficacement les filtres
- Pertinence

Fonctionnalités avancées de recherche

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

- Types de clients disponibles
- Intégrer elasticsearch dans une application Java
 - Embarqué
 - Noeud client
 - Transport client
- Utilisation de l'API Java
- Indexer et rechercher des documents
- Gérer les index
- Gérer les mappings

Cloud et Clusterisation

- Principe de fonctionnement d'un cluster elasticsearch
- Prévenir les split brains
- Configurer un cluster elasticsearch
- Scalabilité et volumétrie de données
- Sauvegarder et restaurer un cluster elasticsearch
- Supervision d'un cluster elasticsearch :
 - l'API health
 - l'API state
 - l'API stats
 - l'API pending_task
 - l'API _nodes alias node info
 - l'API hot_threads
 - l'API _cat
 - Plugins de supervision
 - Logs

Fonctionnalités avancées

- Autres types de plugins elasticsearch
- River: définition
- Scripting
- Pré-chauffer son cluster
- La percolation
- Maintenance des nodes
- Répartition des indexes sur différents nodes
- Requêtes et statistiques

MODULE COMPLÉMENTAIRE EN ANGLAIS SUR DEMANDE (+2 JOURS)

- Training language: English
- Course level : Beginner to intermediate

Cette formation permet de maîtriser les concepts de base de Elasticsearch et d'explorer l'ensemble des fonctionnalités. Elle apporte les connaissances nécessaires pour pouvoir utiliser Elasticsearch.

d'une façon efficace en s'appuyant sur des cas d'utilisations réelles. Lors de cette formation, on aborde les bonnes pratiques ainsi que les problèmes souvent rencontrés

Théorie : 60% Pratique : 40%

Audience :

- Software Developers
- Data Engineers
- Architects

Prerequisites :

- Knowledge of REST/HTTP, Json, Yaml are appreciated
- No knowledge required

Getting Started

- Elasticsearch Overview
- Key Features
- Basic Concepts
- Install Elasticsearch
- CRUD Operations
- First steps on Search API

Mapping and Analysis

- Introduction
- Data Types
- Main parameters
- Mapping API
- Analysis and Inverted Index
- Custom Analyzer
- Multi-Fields

Querying

- Search API Overview
- Terms Search
- Full Text Search
- Compound Queries

Aggregations

- Aggregations Overview

- Metrics, Aggregations
- Buckets Aggregations
- Pipelines Aggregations

Modeling data

- Denormalization
- Object relationship
- Nested relationship
- Parent/Child relationship

Dynamic mapping and Templates

- Dynamic field mapping
- Dynamic template
- Template API

Nodes and Cluster Management

- Multi-nodes cluster
- Nodes Types
- Cluster settings

Ingest API

- Ingest Node
- Pipeline API
- Scripting

Overview

- Script API
- Painless language

More Features

- Alias
- Watcher
- Highlighting
- Suggestion

- Reindex Data
- Update by Query and Delete by Query

Formation Kibana

Formation Elasticsearch administration et infrastructure

Formation Elastic Stack ELK

Sociétés concernées

Cette formation s'adresse aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiple permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.