

## Formation Elasticsearch, Indexation de contenu

### Présentation

Devenu extrêmement populaire depuis sa sortie en 2010, Elasticsearch est un moteur de recherche open source puissant permettant d'indexer et de rechercher des données.

Cette formation certifiante vous permettra d'apprendre à développer avec Elasticsearch.

Durée : 14,00 heures (2 jours)

Tarif INTRA : Nous consulter

### Objectifs de la formation

- Découvrir les enjeux et les concepts clés d'Elasticsearch dans la recherche en plein texte
- Monter des indices et indexer des documents
- Maîtriser les requêtes DSL et les agrégations pour la recherche en plein texte
- Appliquer les scripts et les boosts pour personnaliser les résultats de recherche
- Utiliser les fonctionnalités avancées pour répondre aux besoins courants : auto-complétion, géolocalisation
- Bien choisir la relation des documents dans la production
- Savoir administrer et configurer les clusters
- Coupler Java à Elasticsearch via l'API client

### Prérequis

- Connaissances solides en administration système Unix/Linux.
- Connaissances basiques en SGBD, JSON et REST.

### Public

- Ce cours est destiné aux futurs développeurs d'applications Big Data intégrant Elasticsearch comme moteur de recherche.

### Programme de la formation



**Concepts de base**

- Concept du moteur de recherche
- Introduction à Elasticsearch
- Concepts basics d'Elasticsearch
- Installation et configuration

**Indexation de documents**

- Vue générale de l'API REST
- CRUD API : Index, Get, Exist, Delete, Update, Multi Get, Bulk
- Le versionage des documents

**Elasticsearch vs Lucene : les valeurs ajoutées**

- Index inversé
- Trace de transaction
- Segment

**Analyse de texte**

- Le besoin d'une analyse de texte
- Anatomie d'un Analyzer
- Analyze API
- Configuration d'un Analyzer
- Les Tokenizers (Parseurs en Token)
- Filtre de Tokens
- Filtre de Synonyme
- ICU Tokenizers & FiltersSQL avec Spark

**Mapping**

- Les fondamentaux du Mapping
- Mapping API
- Les différents attributs et types de champs

**Recherche de documents**

- Recherche de documents

**Spark – SQL découverte**

- Les dataframes
- Chargement de données avec un schéma

- Agrégation
- Jointure
- SQL avec Spark

### **Spark – streaming, recherche avancée**

- Agrégation : Scope
- Agrégation métrique
- Agrégation bucket
- Pertinence
- Modèle d'espace vectoriel
- TF/IDF
- Booster la requête
- Scoring personnalisé
- Highlighting
- Suggestion

### **Relation des documents**

- Problématique
- Dénormalisation
- Nested Object :
  - ❓ Mapping
  - ❓ Requête, Filtre, Agrégations
- Parent / Child :
  - ❓ Mapping
  - ❓ Requête, Filtre, Agrégations

### **Fonctionnalités avancées**

- Percolator
- Geolocation

### **Vers la production**

- Configuration d'un cluster
- Scalabilité et volumétrie de données
- Bonnes pratiques matérielles
- Supervision

**Java et Elasticsearch**

- Requête DSL
- Requête full-text
- Requête aux termes
- Requête composée
- BitSet, Filtre et Lucene
- Pagination
- Ordonnement
- Types de recherche
- Warmer Api

**Organisation****Formateur**

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

**Moyens pédagogiques et techniques**

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

**Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation****En amont de la formation**

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

**Tout au long de la formation**

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

**A la fin de la formation**

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Évaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

NB : dans le cadre d'une Action collective, chaque stagiaire bénéficiaire sera contacté par un prestataire choisi par l'Opco Atlas afin d'évaluer « à chaud » la qualité de la formation suivie.

**Accessibilité**

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.