ElasticSearch

La formation « ElasticSearch » vous permettra de maîtriser l'installation, la configuration et la gestion d'un cluster, y compris sa sécurisation. Vous apprendrez à indexer, rechercher et visualiser des données avec Kibana, à utiliser les aggrégations, mappings dynamiques et documents imbriqués. Vous saurez également effectuer des recherches complexes (Query DSL, filtres, recherches géospatiales), résoudre les problèmes courants liés aux performances, et développer des applications avec les API et fonctionnalités avancées d'Elasticsearch. Le programme est donné à titre indicatif et sera adapté à vos besoins et votre niveau après audit. N'hésitez pas à nous contacter pour toute demande spécifique.

Pré-requis

Une bonne connaissance des systèmes d'information

Public concerné

Analystes, développeurs, chefs de projet…

Durée et tarif de la formation

La durée de la formation varie en fonction des besoins et des objectifs déterminés après audit. Les tarifs sont disponibles sur devis.

Contenu de la formation

Présentation d’Elasticsearch

Historique – License

Moteur Lucene : Comprendre le rôle de Lucene comme fondation d’Elasticsearch

API REST : Intégration avec l'API RESTful pour interagir avec les indices

Architecture d’Elasticsearch : Sharding, répartition des données et clustering

Versioning et Rolling Upgrades : Gestion des mises à jour sans interruption du service

Elastic Stack : Introduction au Stack Elastic, y compris Kibana, Beats, et Logstash

Installation

Installation via deb et rpm (ajout de détails sur la version 8.x)

Configuration : Fichiers de configuration, chemin de stockage, nom du cluster, paramètres de nœud

Elastic Agent et Fleet : Installation et configuration pour centraliser la gestion des agents dans le Stack Elastic

Indexation de documents

Ajout, mise à jour, suppression des documents

Mise à jour des mappings : Gestion dynamique des mappings et re-indexation

Types de données et mapping : Utilisation des mappings dynamiques et gestion des types de données dans Elasticsearch 7.x/8.x

Format des données

JSON : Structure des documents JSON

Mapping : Gestion des types de champs, utilisation des analyzers, et gestion des données structurées

Re-indexation et mappings dynamiques : Comment effectuer des mises à jour sans affecter les données existantes

Recherche de données

Query String et DSL : Utilisation des requêtes simples

Agrégations

Agrégations de Facets et Multi-Buckets : Agrégation de données par type (métriques, statistique, etc.)

Agrégations Nested : Gestion des données imbriquées et des sous-documents

Métriques et Statistiques : Statistiques approximatives et mesures sur des ensembles de données

Agrégations combinées : Combinaison de filtres et agrégations pour des analyses plus poussées

Relations entre les documents

Nested Documents : Utilisation des documents imbriqués pour les relations entre les données

Parent-enfant : Optimisation des requêtes avec des relations hiérarchiques

Dénormalisation : Approches pour éviter les relations complexes

Recherches et agrégations sur les documents imbriqués : Analyse des sous-documents à l'aide d'agrégations

Scaling / montée en charge

Sharding et Réplication : Meilleures pratiques pour la gestion des indices

Allouer les Shards : Stratégies de contrôle des allocations et équilibrage de charge

Scaling des écritures et lectures : Optimisation des performances pour la montée en charge

Index Lifecycle Management (ILM) : Mise en place de stratégies de gestion du cycle de vie des indices

Tuning des performances

API de performance : Suivi des performances via des APIs dédiées

Gestion du cache : Types de caches disponibles et leur utilisation optimale

Index Lifecycle Management (ILM) : Automatisation de la gestion des indices

Optimisation des segments : Rafraîchissement, flush, et optimisation des segments

Query warmers : Pré-chargement des requêtes pour

avec Query String et du Query DSL avancé

Filtres et Requêtes complexes : Combinaison de requêtes, utilisation des filtres (filtres booléens, term, range, etc.)

Search Templates : Utilisation des modèles de recherche pour améliorer les performances

Scoring et Pertinence des résultats : Calcul du score, TF-IDF, Boosting, utilisation de Function_score query et Scripting

API explain : Analyser et comprendre le processus de scoring

Analyse de données pour le stockage

Fonctionnement de Lucene et Elasticsearch : Comment les données sont indexées et stockées

Index inversé : Structure de stockage pour optimiser les requêtes

API d’analyse : Utilisation des Tokenizers, Token filters, et des Custom analyzers

Shingles, Ngrams, Edge Ngrams, et Stemming : Techniques pour améliorer la recherche textuelle

améliorer les temps de réponse

Administration du cluster

Surveillance et monitoring : Utilisation des outils de monitoring intégrés dans Elasticsearch (ex : Elastic Stack Monitoring, Elastic Agent)

Rolling Restart : Mise à jour du cluster sans interruption de service

Sécurité

Sécuriser l’accès aux données : Contrôle d'accès basé sur les rôles (RBAC), authentification (LDAP, SAML)

X-Pack : Fonctionnalités avancées de sécurité et de gestion des utilisateurs

Chiffrement des communications : Sécuriser les échanges via TLS

Backup et restauration : Sauvegarde et restauration des indices

Autres fonctionnalités

Auto-complétion avec Completion suggester : Utilisation des fonctionnalités de suggestion pour améliorer l'expérience utilisateur

Highlighting : Mise en valeur des termes recherchés dans les résultats

Données géolocalisées : Indexation et recherche sur des données géospatiales (mapping des coordonnées géographiques)

Machine Learning : Détection d’anomalies, prévisions, et automatisation des workflows analytiques