

Durée : 3 jours soit 21 heures

Référence : IF-ECSH

Public visé :

Public cible : Développeurs, Architectes de données, Data Engineers, Chefs de projets techniques.

Pré-requis :

Prérequis : Bonne connaissance de Linux (ligne de commande), pratique d'un langage de programmation (Python, Java, PHP...), et compréhension du format JSON.

Objectifs pédagogiques :

Objectifs pédagogiques :

- Déployer et configurer un cluster Elasticsearch hautement disponible.
- Concevoir des mappings optimisés pour la volumétrie Big Data.
- Maîtriser le Query DSL pour des recherches complexes et pertinentes.
- Utiliser les agrégations pour l'analyse de données en temps réel.
- Gérer le cycle de vie des données (ILM) et les performances du cluster.

Compétences acquises à l'issue de la formation :

- Déployer et configurer un cluster Elasticsearch hautement disponible.
- Concevoir des mappings optimisés pour la volumétrie Big Data.
- Maîtriser le Query DSL pour des recherches complexes et pertinentes.
- Utiliser les agrégations pour l'analyse de données en temps réel.
- Gérer le cycle de vie des données (ILM) et les performances du cluster.

Modalités pédagogiques :

Session dispensée en présentiel ou téléprésentiel, selon la modalité inter-entreprises ou intra-entreprises sur mesure.

La formation est animée par un(e) formateur(trice) durant toute la durée de la session et présentant une suite de modules théoriques clôturés par des ateliers pratiques validant l'acquisition des connaissances. Les ateliers peuvent être accompagnés de Quizz.

L'animateur(trice) présente la partie théorique à l'aide de support de présentation, d'animation réalisée sur un environnement de démonstration.

En présentiel comme en téléprésentiel, l'animateur(trice) accompagne les participants durant la réalisation des ateliers.

Moyens et supports pédagogiques :

Cadre présentiel

Salles de formation équipées et accessibles aux personnes à mobilité réduite.

- Un poste de travail par participant
- Un support de cours numérique ou papier (au choix)
- Un bloc-notes + stylo
- Vidéo-projection sur tableau blanc
- Connexion Internet
- Accès extranet pour partage de documents et émargement électronique

Cadre téléprésentiel

Session dispensée via notre solution iClassroom s'appuyant sur Microsoft Teams.

- Un compte Office 365 par participant
- Un poste virtuel par participant
- Un support numérique (PDF ou Web)
- Accès extranet pour partage de documents et émargement électronique

Informations sur l'accessibilité :



Description / Contenu

Module 1 : Architecture distribuée et Installation

- Concepts clés : Noeuds (Master, Data, Ingest), Shards et Réplicas.
- Installation en mode cluster et sécurisation native (RBAC).

Module 2 : Stratégies d'indexation et Mapping

- Gestion des types de champs et normalisation.
- Analyse de texte (Analyzers, Tokenizers) et multi-langue.
- Utilisation des Aliases et Templates d'index.

Module 3 : Maîtrise du Query DSL (Recherche)

- Requêtes Full-text vs Term queries.
- Logique Booléenne (must, should, filter).
- Gestion de la pertinence (Scoring) et du Highlighting.

Module 4 : Analyse de données (Agrégations)

- Metrics aggregations (moyenne, stats) et Bucket aggregations (terms, range).
- Agrégations imbriquées et pipelines.

Module 5 : Administration avancée et Performances

- Gestion du cycle de vie (Hot-Warm-Cold architecture).
- Sauvegardes (Snapshots) et restauration.
- Tuning des performances d'indexation et de recherche.