Elk Bootcamp

Durée : 5 Jour(s)
Nombre d'heure : 35 Heure(s)
Prix : 2500 €/Participant

Description

La stack ELK est composée d'une suite d’outils pour mettre en place rapidement un moteur de recherche et optimiser l'accès aux données. Cette formation permet aux participants de maîtriser en profondeurs les outils de base de la stack ELK : elasticsearch, Kibana et logstash

Public concerné

Développeurs, Architectes, Administrateurs systèmes, DevOps, Data Engineer, Data Scientist, Data Analyste

Pré-Requis

les participants doivent avoir des notions en développement informatique.

Objectifs pédagogiques

À l’issue de la formation, le participant sera en mesure de :

- Mettre en place un cluster Elasticsearch
- Administrer un cluster Elasticsearch
- Évaluer les possibilités offertes par Elasticsearch, logstash et Kibana
- Adapter la structuration de l’index Elasticsearch à la problématique métier
- Personnaliser l’indexation aux besoins métier
- Normaliser et nettoyer des données avec logstash
- Écrire des requêtes complexes pour chercher des données dans Elasticsearch
- Mettre en place un pipeline d’indexation et de recherche avec logstash, elasticsearch et kibana
- Intégrer la stack ELK dans un projet data
- Simplifier la remontée de la donnée vers Elasticsearch

Programme

Présentation et installation d'un cluster Elasticsearch, logstash et Kibana

- Les différents composants d’un cluster d'Elasticsearch
- Adapter le cluster aux besoins
- La configuration d'Elasticsearch
- Administration d'Elasticsearch
- Intégrer logstash et Kibana
- Introduire le rôle des Agents Beats dans la simplification de la gestion des données
- Kibana un outil d’administration

Pratique : Mettre en place un cluster elasticsearch avec 3 nœuds et deux clients logstash et Kibana.

Indexation des données avec Elasticsearch

- Présentation rapide des liens entre Apache Lucene et Elasticsearch
- Adapter le format du document JSON à la problématique
- Gérer le cycle de vie de la données dans l’index ILM
- Découvrir le module d’indexation d’elasticsearch
- Mapping Explicite VS Mapping Dynamique
- Personnaliser le mapping Dynamique
- Configurer la gestion des données Textuelles
- Simplifier l’indexation avec les templates

Pratique : Nettoyer et normaliser un grand jeu de données issue de l’open Data. Comprendre le jeu de données, préparer l’index, mettre en place un pipeline logstash, indexer le jeu de données dans le cluster d’elasticsearch

Recherche les données dans Elasticsearch

- Query DSL
- Rechercher par valeurs : Term, Terms, Range, Wildcard
- Recherche Full-text : Match, Match phrase, Query string, Mutli-match
- Composition de requêtes : Boolean, Boosting, Disjunction max
- Les requêtes de jointure : Nested, Parent ID, Has child, Has parent
- Recherche géolocalisée : Geo-bounding box, Geo-distance
- Les agrégations : Bucket aggregations, Metrics aggregations, Pipeline aggregations

Pratique : Écrire des requêtes simples et complexes

Optimisation des recherches textuelles

- Adapter l’analyse
- Enrichissement des données
- Personnaliser la similarité
- L'utilisation des scripts Painless en indexation et en recherche
- Gérer la pertinence des résultats (scoring)
- Ajouter l’incertitude dans les requêtes
- L’intégration du Machine learning pour optimiser la gestion des données textuelles
- Les agrégations : Bucket aggregations, Metrics aggregations, Pipeline aggregations
- Mise en valeur des résultats

Pratique : Écrire des Recherches multicritères dans elasticsearch avec une gestion des pertinences et une mise en valeur des résultats.