RER – Stack Elastic

**Contexte :**

Prendre en main la Stack Elastic avec des requêtes simples et d’autres agrégations plus complexes.

# **Problématique(s) :**

Comment réaliser des requêtes simples avec Stack Elastic ?

Comment utiliser Stack Elastic en local ?

En quoi Elasticsearch permet de se passer des requêtes SQL ?

Comment utiliser l’API REST de Elasticsearch pour indexer des données ?

# **Mots clés :**

* Elasticsearch : un système de recherche et d'analyse de données, principalement utilisé pour stocker, rechercher et analyser de grandes quantités d'informations de manière rapide et efficace. Il est souvent utilisé pour créer des moteurs de recherche, des systèmes de surveillance des journaux, des tableaux de bord d'analyse de données et d'autres applications nécessitant une recherche et une analyse puissantes de données non structurées ou semi-structurées. Elasticsearch est conçu pour être hautement évolutif et capable de gérer de gros volumes de données en temps réel. Il est souvent utilisé en combinaison avec d'autres produits de la suite Elastic, tels que Logstash et Kibana, pour fournir une solution complète d'analyse de données et de visualisation.
* Stack Elastic : ensemble de logiciels et d'outils développés par Elastic, la société derrière Elasticsearch. Cette stack comprend généralement trois composants principaux : Elasticsearch, Logstash, Kibana.

Ensemble, ces trois composants forment une stack puissante pour la recherche, l'analyse et la visualisation de données, souvent utilisée dans des applications telles que la surveillance des journaux, la gestion des données de sécurité, l'analyse des performances et bien d'autres domaines.

* Kibana : application web qui fait partie de la stack Elastic. Elle permet de visualiser et d'analyser les données stockées dans Elasticsearch, le moteur de recherche et d'analyse de données. En utilisant Kibana, les utilisateurs peuvent créer des tableaux de bord interactifs, des graphiques, des diagrammes et des visualisations pour explorer et comprendre leurs données. C'est un outil essentiel pour l'analyse de données en temps réel, la surveillance des journaux, la gestion des performances et d'autres applications nécessitant la visualisation de données. En résumé, Kibana facilite la transformation des données stockées dans Elasticsearch en informations compréhensibles et utiles.
* Beats : une famille d'agents légers développés par Elastic, conçus pour collecter, expédier et centraliser des données dans la stack Elastic, notamment Elasticsearch, Logstash et Kibana. Ces agents sont spécialement conçus pour différentes tâches de collecte de données, comme la surveillance des journaux (Filebeat), la collecte de données système (Metricbeat), la collecte de données de sécurité (Auditbeat), et d'autres tâches similaires.
* Logstash : un outil de traitement et de transformation de données qui fait partie de la stack Elastic. Il est principalement utilisé pour collecter, enrichir, filtrer et acheminer des données à partir de diverses sources vers Elasticsearch ou d'autres destinations.

En termes simples, Logstash agit comme un "tuyau" pour vos données, les aidant à passer d'une variété de sources (comme les fichiers journaux, les bases de données, les flux de réseau, etc.) à Elasticsearch, où elles peuvent être stockées, recherchées et analysées. Logstash permet également de nettoyer et d'enrichir les données en ajoutant des métadonnées, en appliquant des filtres et en effectuant d'autres opérations de transformation, ce qui en fait un composant essentiel pour la gestion des données et l'analyse au sein de la stack Elastic.

* Apache Lucene : une bibliothèque open-source de recherche et d'indexation de textes. Elle fournit des outils et des fonctionnalités pour indexer et rechercher efficacement des informations dans de grandes quantités de texte. Lucene est souvent utilisé comme composant de base pour construire des moteurs de recherche, des systèmes de recherche de texte et d'autres applications qui nécessitent une recherche rapide et précise dans des collections de documents.
* Curl : une commande ou un outil informatique utilisé en ligne de commande pour effectuer des requêtes réseau, généralement des requêtes HTTP, vers des serveurs web ou d'autres ressources en ligne. Il permet de récupérer des données à partir d'URL, d'envoyer des données vers des serveurs, et d'effectuer diverses opérations liées au protocole HTTP.
* Nodes (nœud) :
* Indexation : processus de création d'une liste organisée et structurée de références ou de données, permettant de faciliter la recherche, l'accès et la récupération d'informations ultérieures. Elle consiste à créer un index, qui est essentiellement une sorte de table de correspondance entre des éléments (comme des mots, des numéros, des termes) et les endroits où ces éléments sont stockés.

En d'autres termes, l'indexation transforme des données brutes en un système organisé qui permet aux utilisateurs de trouver rapidement des informations spécifiques, que ce soit dans un livre, une base de données, un site web ou d'autres types de collections de données. C'est un concept clé dans la recherche d'informations et la gestion de données, car elle rend les informations plus accessibles et plus efficaces à manipuler.

Une image contenant texte, Police, logiciel, Page web

Description générée automatiquement

# **Hypothèses :**

* Adrien : On peut déployer Elasticsearch (la base de données) sur plusieurs machines
* Axel : On ne peut pas utiliser Elasticsearch pour collecter des données non structurées.

FAUX, Elasticsearch est principalement conçu pour indexer, rechercher et analyser des données semi-structurées ou non structurées.

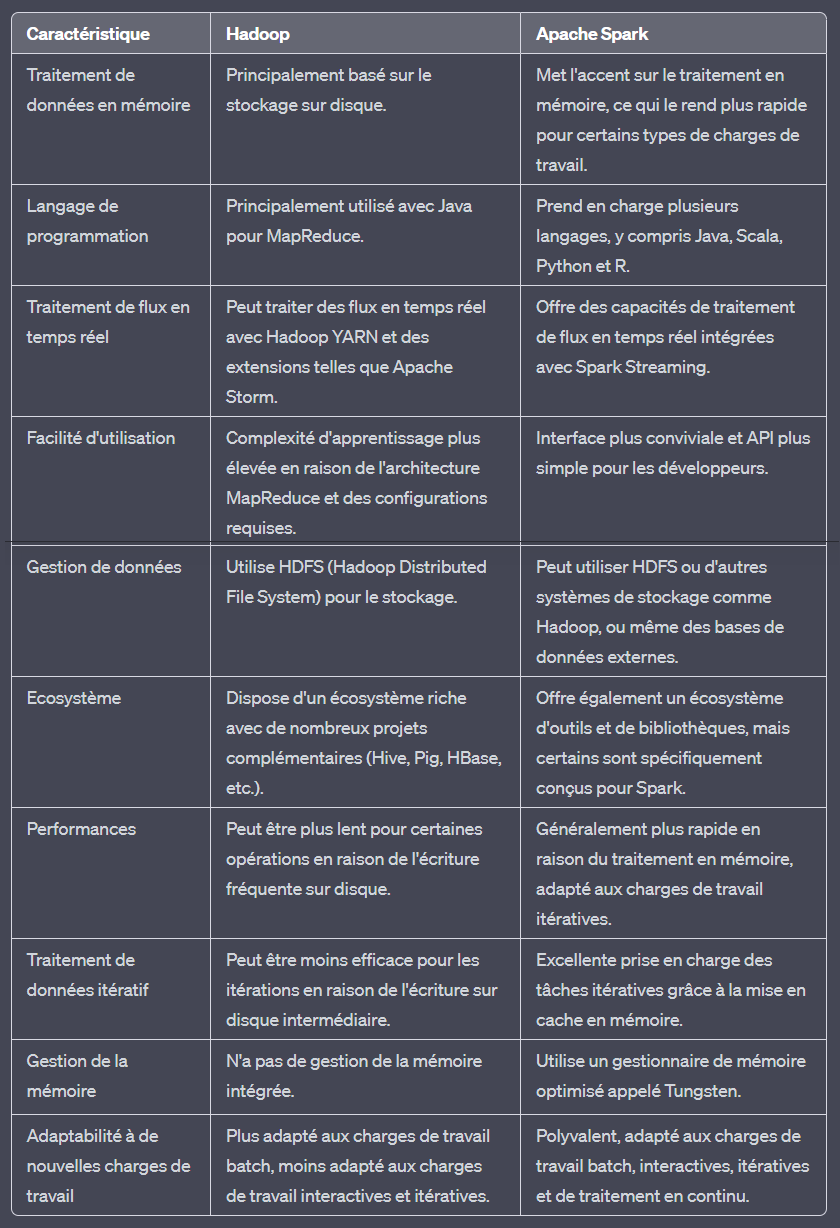
* Etienne : On peut faire des requêtes d’agrégation directement depuis l’API d’Elasticsearch
* Loic : on ne peut utiliser Elasticsearch que pour des données textuelles
* Briand : On doit faire les clusters de données avant de pouvoir utiliser Elasticsearch
* Seydou : Elasticsearch peut être utilisé pour implémenter des fonctionnalités de recherche sémantique
* Nicolas : On peut utiliser Grafana avec Elasticsearch
* Aude : Elasticsearch est une alternative à snowflake
* Jean-Paul : L’indexation de données s’appuie sur le dictionnaire de données.
* Tetyana : Pour afficher les données d’Elasticsearch on utilise un langage GO pour des requêtes
* Bassam : On peut utiliser des dockers pour lancer Elasticsearch

# Plan d'action :

* Explorer les ressources
* Traiter les mots clés
* Répondre aux problématiques
* Comparer Elasticsearch et kibana



* Comparer Hadoop et Apache Spark



* Rendre les livrables : RER et workshops