Travaux Pratiques :

Querying Data

Note :
Le travail s'effectue par groupe de 3 personnes
Ce TP représente 40 % de la note finale (Projet : 60%)

1. Via un programme Python/Javascipt ou autre :
   - Vous allez indexer le fichier blog.csv sous un index nommé :
   « nom_de_votre_groupe »
   Cela nécessite de convertir le fichier JSON correspondant
   Crée la connexion avec le cluster/localhost
   Et charger la donnée

2. Écrivez une requête qui **match tous** les documents de l'index des blogs.
   Vous devriez avoir un total de 1594 hits:

3. Ajoutez le paramètre "taille" à votre demande précédente et définissez-le sur 100.
   Vous devriez maintenant voir 100 blogs dans les résultats et non 10.

4. Ecrivez et exécutez une requête qui affiche tous les articles de blog publiés en mai 2017. La réponse devrait vous indiquer qu'il y a 44 hits en tout, mais notez que vous ne récupérez que 10 hits car la taille par défaut de la taille d'une requête est 10.

5. Écrivez et exécutez une requête match pour les blogs qui ont le terme "elastic" dans le champ "title".
   Vous devriez obtenir 260 hits.

6. Maintenant, lancez une requête match pour "elastic stack" dans le champ de titre.
   Pourquoi le nombre de résultats a-t-il augmenté en ajoutant un terme à la requête match?

7. La recherche de " elastic stack " envoie un résultat très large. Les résultats les plus réussis ont l'air beau, mais la précision n'est pas bonne pour pas mal d'autres, en particulier pour les produits dont le nom est " elastic " mais pas " stack ".
   Modifiez l'opérateur de votre requête de match précédente eet puis réexécutez-la.
   Notez que cela augmente la précision, puisqu'il n'y a plus que 70 hits:

8. Ecrivez une requête pour chacune des recherches suivantes:
   les blogs qui ont le mot " search " dans leur champ " content ".

les blogs qui ont " search " ou " analytics " dans leur champ " content ".
les blogs qui ont " search " et " analytics " dans leur " content ".

9. Exécutez une recherche match_phrase pour les analyses de recherche dans le champ de contenu qui renvoie les 3 résultats les plus importants. Vous devriez obtenir 6 hits au total.

10. L'expression "search and analytics " est assez courante dans le contenu du blog.
Mettez à jour la requête match_phrase précédente de sorte qu'elle permette à 1 terme (n'importe quel mot - pas seulement " and ") d'apparaître entre " search " et " analytics ".
Combien de hits voyez-vous maintenant?

11. Exécutez une requête qui répond à la question: " Which blogs have performance or optimizations or improvements in the content field?" Vous devriez obtenir les hits suivants:

```
"hits": {
    "total": 374,
```

12. Exécutez une requête qui répond à la question: " Which blogs have a content field that includes at least 2 of the terms performance or optimizations or improvements?" Vous devriez obtenir les hits suivants cette fois-ci

13. Analysons les hits de la recherche " performance optimizations improvements ":
Quel était le score maximal?

    Si vous recherchiez «performance optimizations improvements» pour affiner votre déploiement, voyez-vous des problèmes avec certains des résultats du groupe de documents avec le score le plus élevé?

Détails :
Les noms de colonnes :

{title};
{seo_title};
{url};
{author};
{date};
{category};
{locales};
{content}