



# RAPPORT DU MINI PROJET:

**« ANALYSE ET SURVEILLANCE EN TEMPS RÉEL DES  
TWEETS ET AFFICHAGE SOUS FORME DE TABLEAU DE  
BORD »**



---

## I. Description du mini projet

---

Le but du projet est de concevoir un tableau de bord qui affichera des informations analysées sur les tweets de twitter en temps réel.

On a choisi comme thème pour notre mini-projet : Covid19 et spécialement les vaccins qui sont conçus pour lui. On envisage faire une analyse de sentiments à l'égard du vaccin afin d'avoir un aperçu sur les différentes opinions sur ce sujet.

Pour créer le tableau de bord, il existe plusieurs outils qui se mettent en œuvre, mais puisqu'on va créer un tableau de bord dynamique « en temps réel » dont les données s'actualisent à chaque fois qu'on l'ouvre, il faut premièrement faire une collecte de données en temps réel. Pour ce faire, on a choisi la suite ELK . C'est une suite qui se charge de tout ce processus.

Avant d'entamer les étapes qu'on a suivies pour accomplir ce projet, voyons tout d'abord l'intérêt de créer un Dashboard dynamique et puis quelles sont les différents outils utilisés et faisons un comparatif entre eux.

### **1. L'INTÉRÊT DES DASHBOARD DYNAMIQUES :**

- **Mises à jour en temps réel des données :**

Obtenir des mises à jour en temps réel sur les données est peut-être l'avantage le plus important de la mise en œuvre de tableaux de bord et de rapports dans vos opérations. Le but d'un tableau de bord est de surveiller les données sous-jacentes et de mettre à jour immédiatement les modifications. Cela vous permet de déterminer les mouvements dans les mesures clés dès qu'ils se produisent sans avoir à attendre des rapports occasionnels.

En obtenant des mises à jour en temps réel, les entreprises peuvent détecter rapidement les fraudes et les erreurs afin de minimiser les

---

dommages et d'affiner leurs contrôles internes. Le repérage immédiat des activités frauduleuses vous aide également à améliorer la précision des prévisions et des prévisions commerciales.

L'analyse des données en temps réel rend également les processus plus agiles, ce qui entraîne finalement une augmentation de la productivité de l'entreprise.

- **Visibilité améliorée :**

Obtenir une visibilité complète de vos opérations est essentiel pour augmenter l'efficacité et réagir rapidement aux changements de l'entreprise. Les avantages de l'utilisation du tableau de bord dans l'analyse incluent la fourniture des données appropriées de différents départements et la connexion des dirigeants via des flux de données perspicaces. Cela aide les chefs d'entreprise à éviter les erreurs dues aux estimations et à prendre les meilleures décisions pour l'ensemble de leurs opérations.

- **Meilleures prévisions :**

Réaliser de meilleures prévisions commerciales est l'un des avantages des rapports de tableau de bord qui est trop important pour être ignoré. Les rapports de tableau de bord vous permettent d'analyser les données historiques et de trouver des modèles pouvant être utilisés pour prédire les tendances et les changements futurs. Cela vous permet de devenir plus proactif face aux changements de l'entreprise et d'éliminer le potentiel d'échec et de pertes avant qu'ils ne surviennent.

- **Les données complexes sont simplifiées :**

Opérer dans la quatrième révolution industrielle a incité les entreprises à collecter plus de données que jamais. Cependant, des enquêtes montrent que 44% des organisations ne peuvent pas tirer pleinement parti de leurs données commerciales en raison de leur complexité. Les rapports de tableau de bord prennent toutes ces données et éliminent l'encombrement pour vous donner accès aux informations qui comptent.

---

Des tableaux de bord puissants peuvent transformer des informations complexes en visualisations de données à la fois simples et pertinentes à l'aide de graphiques faciles à comprendre. Cela permet aux chefs d'entreprise d'avoir un aperçu rapide de leurs performances globales afin de les optimiser pour plus d'efficacité. Il vous permet également de créer des rapports personnalisés qui fournissent les données correctes à la bonne personne au bon moment.

- **Intégrer des données de multiples sources**

L'intégration des données est un autre avantage qui vient de tirer parti des tableaux de bord et des rapports. Les tableaux de bord dynamiques permettent de centraliser les informations de plusieurs flux de données sur un seul tableau de bord. En connectant plusieurs sources de données, les entreprises peuvent exploiter la valeur des informations et accroître la collaboration entre les membres de l'équipe et les parties prenantes.

Cela peut être particulièrement utile à l'ère de la mondialisation rapide, où l'extraction et l'analyse instantanées de données sont vitales pour améliorer rapidement les processus. Les tableaux de bord peuvent également éliminer la mauvaise qualité des données et garantir l'intégrité des données pour une compétitivité accrue.

## **2. BENCHMARK DES OUTILS DE CRÉATIONS DES DASHBOARDS :**

Suite à notre recherche sur les différentes méthodes pour créer un tableau de bord, on a constaté qu'il y a une multitude de méthodes qui sont soit open-source, soit payante.

Pour les solutions open source il y a : Grafana, Freeboard, Mozaïk, Kibana, tableau ....

Voilà un comparatif entre chacune de ces outils en montrant leurs avantages et inconvénients :

---

## **GRAFANA :**

Grafana est un outil de visualisation de données puissant et riche en fonctionnalités qui permet aux utilisateurs de créer, d'explorer et de partager des tableaux de bord. Écrit principalement en Go et Typescript, il est principalement utilisé pour surveiller la santé du serveur / de l'architecture, mais peut également être utilisé pour différents types de visualisation de données ou de tableaux de bord de métriques. Il est utile pour tout propriétaire d'entreprise ou individu à la recherche d'un outil de visualisation pour analyser les mesures commerciales, mais il est le plus couramment utilisé pour la surveillance de l'infrastructure.

### **Avantages**

---

- Il fournit une bibliothèque officielle avec une pléthore de modèles de tableau de bord et de plugins.
- Il prend en charge une grande variété de sources de données et de bases de données, notamment Prometheus, Graphite, ElasticSearch, OpenTSDB, InfluxDB et bien d'autres.
- Il offre aux utilisateurs la possibilité de personnaliser leur bureau pour toute entreprise.
- Il fournit un contrôle utilisateur et un mécanisme d'authentification intégrés.
- Il fournit un ensemble diversifié de fonctionnalités, notamment des instantanés, des annotations de données et bien plus encore.
- Il permet aux utilisateurs des alertes et des notifications personnalisées.

### **Les inconvénients**

---

- Grafana est uniquement un outil de visualisation et ne fournit pas de support pour la collecte de données ou le stockage de données.
- Le tableau de bord demande un temps considérable à mettre en place et n'est pas aussi efficace qu'une solution rapide.

---

## **FREEBOARD :**

Freeboard est un outil de tableau de bord simple et open source qui permet aux utilisateurs de créer et de personnaliser des tableaux de bord interactifs en temps réel. Il est principalement conçu pour être utilisé avec des appareils IoT et est idéal pour les organisations et les individus travaillant avec des applications Web, des appareils externes ou des capteurs. Nous avons vraiment aimé l'interface facile à utiliser par glisser-déposer fournie par Freeboard ainsi que les options hébergées gratuites et payantes pour l'accès personnel et d'équipe aux tableaux de bord.

---

### **Avantages :**

- Vous pouvez facilement créer des sources de données et ajouter des widgets à vos tableaux de bord.
- Le tableau de bord fonctionne comme une page Web statique d'une seule page, supprimant le besoin d'un serveur.
- Intégration facile avec n'importe quelle API Web que vous utilisez.
- Licencié sous MIT pour que les utilisateurs utilisent d'autres Freeboards comme point de départ pour leurs propres tableaux de bord. Nous pensons que cela peut être une fonctionnalité très utile pour les débutants et les professionnels.
- Freeboard vous fournit également un tableau de bord prêt pour la production sous forme d'URL unique et facilement partageable. Une fonctionnalité qui nous a beaucoup plu.
- Vous obtenez des tableaux de bord gratuits illimités qui peuvent être utilisés sur autant d'appareils que vous le souhaitez.

---

### **Les inconvénients :**

- La version gratuite ne prend pas en charge les tableaux de bord privés. Nous recommandons l'édition premium pour la gestion des données sensibles.

- 
- Un manque de capacités de visualisation des données par rapport à ses concurrents.
  - Freeboard nécessite que vous vous familiarisiez avec l'écosystème JavaScript pour créer vos plugins. Mais, il fournit également une architecture de plugin facile à suivre dans Plugin Example.

## Mozaïk :

Mozaïk est un outil de tableau de bord hautement personnalisable basé sur le trifecta JavaScript de Node, React et D3 pour créer des tableaux de bord magnifiquement conçus. Les tableaux de bord qui sont conçus pour être facilement évolutifs et extensibles à l'aide de modules, sont compatibles avec plusieurs appareils. Nous avons aimé l'interface simple et conviviale et les widgets facilement intégrables fournis par Mozaïk.

### Avantages

---

- Il offre aux utilisateurs la possibilité de créer et d'utiliser plusieurs tableaux de bord.
- Communications backend optimisées et support en temps réel à l'aide de WebSockets.
- Il fournit un système de grille pour aider les utilisateurs à définir la disposition de leur tableau de bord. Une fonctionnalité intéressante qui permet aux utilisateurs de concevoir facilement leurs tableaux de bord.
- Il est livré avec 6 thèmes hautement personnalisables et permet aux utilisateurs de créer leurs propres thèmes.
- Il fournit des tableaux de bord avec une mise en page évolutive qui prend en charge plusieurs appareils.
- Il fournit un large éventail de widgets qui permettent aux tableaux de bord d'accéder à de nombreux services.

### Les inconvénients :

---

- La création de widgets qui ne sont pas déjà disponibles est une tâche difficile et nécessite une bonne dose de savoir-faire en matière de codage.

- 
- Un manque de capacités de manipulation de données et de mesures commerciales avancées.

## KIBANA :

Représentant le K dans la principale plate-forme de gestion de journaux au monde, la pile ELK, Kibana est une puissante application de navigation et de visualisation de données open source. Il offre aux utilisateurs la possibilité de surveiller et de gérer leurs données ElasticStack. C'est un excellent outil pour tous ceux qui souhaitent utiliser Elasticsearch ou la pile ELK pour gérer leurs données.

### Avantages :

---

- Il permet aux utilisateurs de créer leurs propres façons de représenter leurs données et leurs chiffres avec les outils de visualisation standard.
- Il permet aux utilisateurs d'explorer les relations dans leurs données.
- Il ne nécessite aucun codage ou infrastructure supplémentaire car il s'exécute sur un serveur Web Node.js.
- Il fournit de puissantes fonctionnalités de détection d'anomalies pour détecter facilement les problèmes dans les données.
- Il offre aux utilisateurs des capacités de partage de données, y compris certaines capacités d'exportation de données aux formats PDF et CSV.
- Il fournit une interface utilisateur unique efficace et facile à utiliser.

### Les inconvénients :

---

- Les fonctionnalités d'exportation de données sont encore limitées.
- Puisqu'il fonctionne avec ElasticStack, les limitations d'ElasticStack affectent également Kibana.

Il existe le fameux tableau qui se trouve comme opensource et aussi en version payante

---

# TABLEAU :

## Avantages :

- Crée rapidement des visualisations interactives
- Facilité de mise en œuvre
- Tableau peut gérer de grandes quantités de données
- Utilisation d'autres langages de script dans Tableau
- Support mobile et tableau de bord réactif

## Inconvénients :

- Tableau ne fournit pas la fonctionnalité d'actualisation automatique des rapports à l'aide de la planification.
- Aucune importation visuelle personnalisée. Tableau n'est pas un outil ouvert complet. Ainsi, tous les nouveaux visuels doivent être recréés au lieu d'être importés.
- Formatage personnalisé dans Tableau:

La mise en forme conditionnelle de Tableau et les affichages de tableau de 16 colonnes limités sont des problèmes pour les utilisateurs. De plus, pour implémenter le même formatage dans plusieurs champs, il n'y a aucun moyen pour un utilisateur de le faire pour tous les champs directement. Les utilisateurs doivent le faire manuellement pour chaque champ, ce qui prend beaucoup de temps.

- Paramètres statiques et à valeur unique:

Les paramètres de Tableau sont statiques et une valeur unique peut toujours être sélectionnée à l'aide d'un paramètre. Chaque fois que les données sont modifiées, ces paramètres doivent être mis à jour manuellement à chaque fois. Il n'existe aucun moyen pour un utilisateur d'automatiser la mise à jour des paramètres.

- 
- Résolution d'écran sur les tableaux de bord Tableau:

La disposition des tableaux de bord est perturbée si la résolution d'écran du développeur du tableau est différente de la résolution de l'écran de l'utilisateur final, par exemple si les tableaux de bord ont été créés dans une résolution d'écran de 1920X1080 et sont affichés en 2560X1440, alors la disposition des tableaux de bord sera déformée. Un petit peu. De plus, leurs tableaux de bord ne sont pas réactifs. Vous devrez donc créer un tableau de bord pour mobile et ordinateur de bureau.

- Prétraitement limité des données.

Tableau est strictement un outil de visualisation. Tableau Desktop vous permet d'effectuer un prétraitement très basique. Cela inclut la jonction et la fusion des données. En outre, vous avez la possibilité de modifier les types de données. Dans un monde idéal, la plupart des données seraient exportées dans des tableaux parfaits. Cependant, le nettoyage des données est une étape nécessaire. Dans la plupart des cas, un analyste doit créer un modèle de données récurrent pour formater les données. Cela nécessite un outil tel qu'Altyrex, Power BI, Python ou même Excel pour prétraiter les données avant le chargement. En 2018, a lancé son propre outil de préparation de données appelé Tableau Prep. Vous pouvez lire les résultats de Tableau Prep vs Altyrex

- Mise à l'échelle et tarification pour l'entreprise :

C'est le plus gros problème avec Tableau, c'est un produit très coûteux à mettre à l'échelle dans une grande organisation. Par rapport aux outils BI moins chers et plus complets. Tableau l'une des options les plus chères. Pour la sécurité et le partage, la seule option est Tableau Server qui peut 175 000 \$ pour une option à 8 coeurs et 35 \$ par utilisateur. Vous pouvez également utiliser Tableau Online qui est limité mais coûte 35 \$ par utilisateur. Cela peut s'accumuler si vous essayez d'avoir un grand nombre d'utilisateurs accèdent aux rapports.

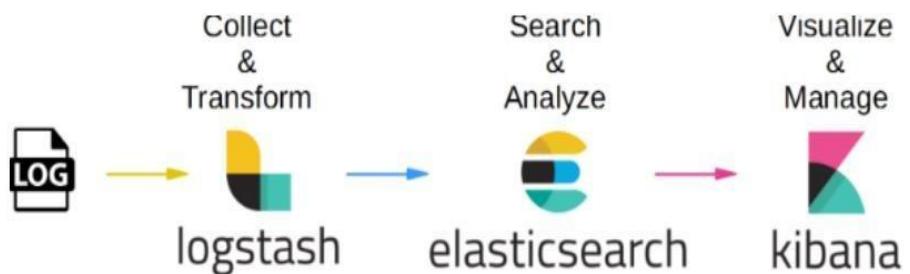
# Justification du choix de la suite ELK :

Après avoir vu ces différents outils, on a trouvé que la suite ELK sera un bon outil pour réaliser notre projet car il intègre les différents étapes pour le réaliser .Alors c'est quoi au juste cette suite ?



## **LA SUITE ELK ?**

"ELK" est un acronyme pour trois projets en open source : Elasticsearch, Logstash et Kibana. Elasticsearch est un moteur de recherche et d'analyse. Logstash est un pipeline côté serveur, destiné au traitement des données (type ETL). Sa mission ? Ingérer simultanément des données provenant de multitude de sources, puis les transformer et les envoyer vers un système de stockage comme Elasticsearch. Kibana permet aux utilisateurs de visualiser des données avec des tableaux et des graphes dans Elasticsearch.



#### ***REQUIS POUR NOTRE PROJET :***

Pour pouvoir accomplir notre projet, on aura besoin de :

Twitter Streaming API

L'installation et configuration de la suite ELK

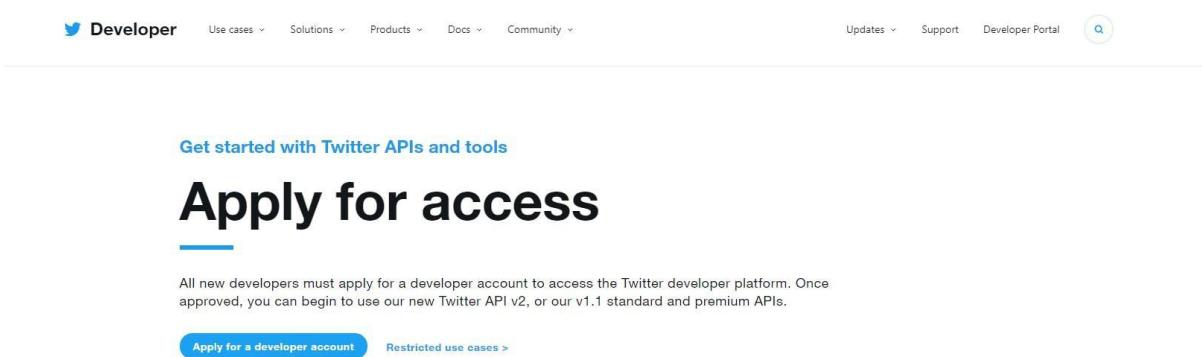
---

## II. Twitter Streaming API :

---

Il permet d'ingérer les données de Twitter.

Le processus d'ingestion de données commence par la demande d'un compte développeur Twitter. Les développeurs qui souhaitent accéder à l'API Twitter doivent demander un compte développeur. Vous ne pouvez commencer à utiliser l'API que lorsque la demande de compte développeur a été approuvée.



The screenshot shows the Twitter Developer website. At the top, there's a navigation bar with links for 'Developer', 'Use cases', 'Solutions', 'Products', 'Docs', 'Community', 'Updates', 'Support', 'Developer Portal', and a search icon. Below the navigation, a blue button says 'Get started with Twitter APIs and tools'. The main section is titled 'Apply for access' in large, bold, black font. A sub-section below it states: 'All new developers must apply for a developer account to access the Twitter developer platform. Once approved, you can begin to use our new Twitter API v2, or our v1.1 standard and premium APIs.' There are two buttons at the bottom: a blue one labeled 'Apply for a developer account' and a smaller grey one labeled 'Restricted use cases >'.

Une fois la demande approuvée, vous devez créer une application Twitter. Une application Twitter peut être créée via la page du tableau de bord de l'application Twitter avec un compte de développeur approuvé. Avec cette application, vous pouvez générer toutes les informations d'identification nécessaires à l'authentification et au streaming des données avec l'API. Les informations d'identification nécessaires incluent API KEY, API SECRET, ACCESS TOKEN et ACCESS TOKEN SECRET.

---

## III. INSTALLATIONS ET CONFIGURATIONS :

---

On a installé la suite ELK sous Ubuntu. Donc, on a créé une machine virtuelle avec VIRTUALBOX.

Avant d'installer la suite ELK il faut installer :

- JAVA 8 si ce n'est pas encore installé
- Serveur web NGINX

---

## 1. INSTALLATION DE JAVA :

On installe java 8 en tapant la commande suivante :

```
hind@hind-VirtualBox:~$ sudo apt-get install openjdk-8-jdk
[sudo] Mot de passe de hind :
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Les paquets supplémentaires suivants seront installés :
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java
  libatk-wrapper-java-jni libgif7 libice-dev libpthread-stubs0-dev libsm-dev
```

```
hind@hind-VirtualBox:~$ java -version
openjdk version "1.8.0_292"
OpenJDK Runtime Environment (build 1.8.0_292-8u292-b10-0ubuntu1~18.04-b10)
OpenJDK 64-Bit Server VM (build 25.292-b10, mixed mode)
```

JAVA s'est bien installé

## 2. INSTALLATION DE NGINX :

**NGINX** est un serveur web open-source qui, depuis son succès initial en tant que serveur web, est maintenant aussi utilisé comme reverse proxy, cache HTTP, et load balancer.

NGINX est conçu pour offrir une **faible utilisation de la mémoire** et une grande simultanéité. Plutôt que de créer de nouveaux processus pour chaque requête Web, NGINX utilise une approche asynchrone et événementielle où les requêtes sont traitées dans un seul thread.

Avec NGINX, un processus maître peut contrôler plusieurs processus de travailleurs. Le maître gère les processus du travailleur, tandis que les travailleurs effectuent le traitement proprement dit. Comme NGINX est asynchrone, chaque requête peut être exécutée simultanément par le travailleur sans bloquer les autres requêtes.

Quelques caractéristiques communes vues dans NGINX incluent :

- Proxy inversé avec mise en cache
- IPv6

- 
- Équilibrage de charge
  - Support FastCGI avec mise en cache
  - WebSockets
  - Gestion des fichiers statiques, des fichiers d'index et de l'indexation automatique
  - TLS/SSL avec SNI

Dans notre cas, Nginx est utilisé pour configurer un accès contrôlé par mot de passe au tableau de bord Kibana.

Pour l'installer on procède de la façon suivante :

```
hind@hind-VirtualBox:~$ sudo apt-get install nginx
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Les paquets supplémentaires suivants seront installés :
  libnginx-mod-http-geoip libnginx-mod-http-image-filter
  libnginx-mod-http-xslt-filter libnginx-mod-mail libnginx-mod-stream
  nginx-common nginx-core
```

Pour démarrer le service Nginx et voir les services de Nginx:

```
ind@hind-VirtualBox:~$ sudo systemctl start nginx
ind@hind-VirtualBox:~$ sudo ufw app list
Applications disponibles :
  CUPS
  Nginx Full
  Nginx HTTP
  Nginx HTTPS
```

On remarque qu'il manque SSH, alors on va l'installer.

```
hind@hind-VirtualBox:~$ sudo apt-get install openssh-server
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Les paquets supplémentaires suivants seront installés :
  ncurses-term openssh-sftp-server ssh-import-id
Paquets suggérés :
  molly-guard monkeysphere rssh ssh-askpass
Les NOUVEAUX paquets suivants seront installés :
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
0 mis à jour, 4 nouvellement installés, 0 à enlever et 400 non mis à jour.
Il est nécessaire de prendre 637 ko dans les archives.
```

SSH est installé :

```
...
hind@hind-VirtualBox:~$ sudo ufw app list
Applications disponibles :
  CUPS
  Nginx Full
  Nginx HTTP
  Nginx HTTPS
  OpenSSH
```

### 3. INSTALLATION ET CONFIGURATION DE ELASTICSEARCH

- L'AJOUT DE ELK REPOSITORY:

**POUR INSTALLER ELK, IL FAUT AJOUTER ELK REPOSITORY.**

Les ELASTIC repository permettent d'accéder à tous les logiciels open source de la pile ELK. Pour les ajouter, on commence par importer la clé GPG.

```
hind@hind-VirtualBox:~$ wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
OK
```

OK signifie le succès de l'opération

- Après ,l'installation du paquet apt-transport-https

---

```
hind@hind-VirtualBox:~$ sudo apt-get install apt-transport-https
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Le logiciel suivant sera installé :
apt-transport-https
```

- L'ajout de Elastic repository à la liste des repository de notre système :

```
hind@hind-VirtualBox:~$ echo "deb https://artifacts.elastic.co/packages/7.x/stable main" | sudo tee -a /etc/apt/sources.list.d/elastic-7.x.list
deb https://artifacts.elastic.co/packages/7.x/apt stable main
```

Avant de commencer l'installation de elasticsearch, il faut faire une mise à jour des repositories :

```
hind@hind-VirtualBox:~$ sudo apt-get update
Atteint :1 http://ma.archive.ubuntu.com/ubuntu bionic InRelease
Réception de :2 http://ma.archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
Réception de :3 https://artifacts.elastic.co/packages/7.x/apt stable InRelease [10.4 kB]
Réception de :4 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
Réception de :5 http://ma.archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
Réception de :6 https://artifacts.elastic.co/packages/7.x/apt stable/main i386 Packages [54.6 kB]
Réception de :7 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 Packages [54.6 kB]
```

On installe elasticsearch en tapant la commande suivante :

```
hind@hind-VirtualBox:~$ sudo apt-get install elasticsearch
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Les NOUVEAUX paquets suivants seront installés :
  elasticsearch
0 mis à jour, 1 nouvellement installés, 0 à enlever et 400 non mis à jour.
Il est nécessaire de prendre 327 Mo dans les archives.
Après cette opération, 545 Mo d'espace disque supplémentaires seront utilisés.
Réception de :1 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 elasticsearch amd64 7.13.2 [327 MB]
```

```
Préparation du dépaquetage de .../elasticsearch_7.13.2_amd64.deb ...
Creating elasticsearch group... OK
Creating elasticsearch user... OK
Dépaquetage de elasticsearch (7.13.2) ...
Paramétrage de elasticsearch (7.13.2) ...
Created elasticsearch keystore in /etc/elasticsearch/elasticsearch.keystore
Traitement des actions différées (« triggers ») pour ureadahead (0.100.0-21) ...
.
Traitement des actions différées (« triggers ») pour systemd (237-3ubuntu10.33)
...
hind@hind-VirtualBox:~$
```

- Configuration de ELASTICSEARCH :

Pour configurer Elasticsearch, il faut modifier le fichier elasticsearch.yml

En tapant la commande suivante :

```
$$ sudo vim /etc/elasticsearch.yml
```

On élimine le caractère **#** avant « **network.host** » et « **http.port** » comme suit :

```
network.host: 192.168.0.1
#
# By default Elasticsearch listens for HTTP traffic on the first free port it
# finds starting at 9200. Set a specific HTTP port here:
#
http.port: 9200
```

Et on remplace ses valeurs comme suit :

```
network.host: localhost
#
# By default Elasticsearch listens for HTTP traffic on the first free port it
# finds starting at 9200. Set a specific HTTP port here:
#
http.port: 9200
```

on cherche la partie Discovery et on ajoute :

**discovery.type :single-node**

```
"----- Discovery -----  
#  
# Pass an initial list of hosts to perform discovery when this node is started:  
# The default list of hosts is ["127.0.0.1", "[::1]"]  
#  
#discovery.seed_hosts: ["host1", "host2"]  
#  
# Bootstrap the cluster using an initial set of master-eligible nodes:  
#  
#cluster.initial_master_nodes: ["node-1", "node-2"]  
#single node Elastic stack  
discovery.type: single-node  
# For more information, consult the discovery and cluster formation module documentation.  
#
```

Par défaut, la taille de tas JVM est définie sur 1 Go. Nous vous recommandons de ne pas dépasser la moitié de la taille de notre mémoire totale. On ouvre le fichier suivant pour le modifier :

```
hind@hind-VirtualBox:~$ sudo vim /etc/elasticsearch/jvm.options
```

Puis on cherche les lignes commençant par -Xms et -Xmx.

Ici , la taille maximale (-Xmx) et minimale (-Xms) sont définis sur 512 Mo.

```
## See https://www.elastic.co/guide/en/elasticsearch/reference/current/heap-size.html  
## for more information  
##  
#####  
-Xms512m  
-Xmx512m  
#####  
## Expert settings  
#####
```

Démarrons maintenant ElasticSearch :

```
hind@hind-VirtualBox:~$ sudo systemctl start elasticsearch.service  
hind@hind-VirtualBox:~$
```

---

Activer Elasticsearch pour démarrer au démarrage :

```
hind@hind-VirtualBox:~$ sudo systemctl enable elasticsearch.service
Synchronizing state of elasticsearch.service with SysV service script with /lib
/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable elasticsearch
Created symlink /etc/systemd/system/multi-user.target.wants/elasticsearch.servi
ce → /usr/lib/systemd/system/elasticsearch.service.
Afficher les applications ~$
```

Testons maintenant Elasticsearch :

```
hind@hind-VirtualBox:~$ curl -X GET "localhost:9200"
{
  "name" : "hind-VirtualBox",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "9usTV2yQRx0e0uzFFJfa8A",
  "version" : {
    "number" : "7.13.2",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "4d960a0733be83dd2543ca018aa4ddc42e956800",
    "build_date" : "2021-06-10T21:01:55.251515791Z",
    "build_snapshot" : false,
    "lucene_version" : "8.8.2",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
```

#### 4. INSTALLATION ET CONFIGURATION DE KIBANA :

- On installe Kibana en tapant la commande suivante :

```
hind@hind-VirtualBox:~$ sudo apt-get install kibana
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Les NOUVEAUX paquets suivants seront installés :
  kibana
0 mis à jour, 1 nouvellement installés, 0 à enlever et 398 non mis à jour.
Il est nécessaire de prendre 310 Mo dans les archives.
Après cette opération, 844 Mo d'espace disque supplémentaires seront utilisés.
Réception de :1 https://artifacts.elastic.co/packages/7.x/apt/stable/main/kibana amd64 7.13.2 [310 MB]
```

- configuration de kibana :

Pour configurer Kibana, il faut modifier le fichier kibana.yml

Donc ,on tape :

```
$$ sudo vim /etc/kibana/kibana.yml
```

---

on supprimez le signe # au début des lignes suivantes pour les activer :

```
#server.port: 5601 , #server.host: "your-hostname"
```

```
#elasticsearch.hosts: ["http://localhost:9200"]
```

Et on modifie ces lignes comme suit :

```
# Kibana is served by a back end server. This setting specifies the port to use
#server.port: 5601

# Specifies the address to which the Kibana server will bind. IP addresses and
# host names are both valid values.
# The default is 'localhost', which usually means remote machines will not be able to connect.
# To allow connections from remote users, set this parameter to a non-loopback address.
#server.host: "localhost"

# Enables you to specify a path to mount Kibana at if you are running behind a proxy.
# Use the `server.rewriteBasePath` setting to tell Kibana if it should remove the basePath
# from requests it receives, and to prevent a deprecation warning at startup.
# This setting cannot end in a slash.
#server.basePath: ""
```

```
# The maximum payload size in bytes for incoming server requests.
#server.maxPayload: 1048576

# The Kibana server's name. This is used for display purposes.
#server.name: "localhost"

# The URLs of the Elasticsearch instances to use for all your queries.
#elasticsearch.hosts: ["http://localhost:9200"]
```

Puis on démarre kibana et ,on l "enable pour qu'il démarre à chaque démarrage :

```
hind@hind-VirtualBox:~$ sudo systemctl start kibana
hind@hind-VirtualBox:~$ sudo systemctl enable kibana
Synchronizing state of kibana.service with SysV service script with /lib/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable kibana
hind@hind-VirtualBox:~$
```

- Autoriser le trafic sur le port 5601

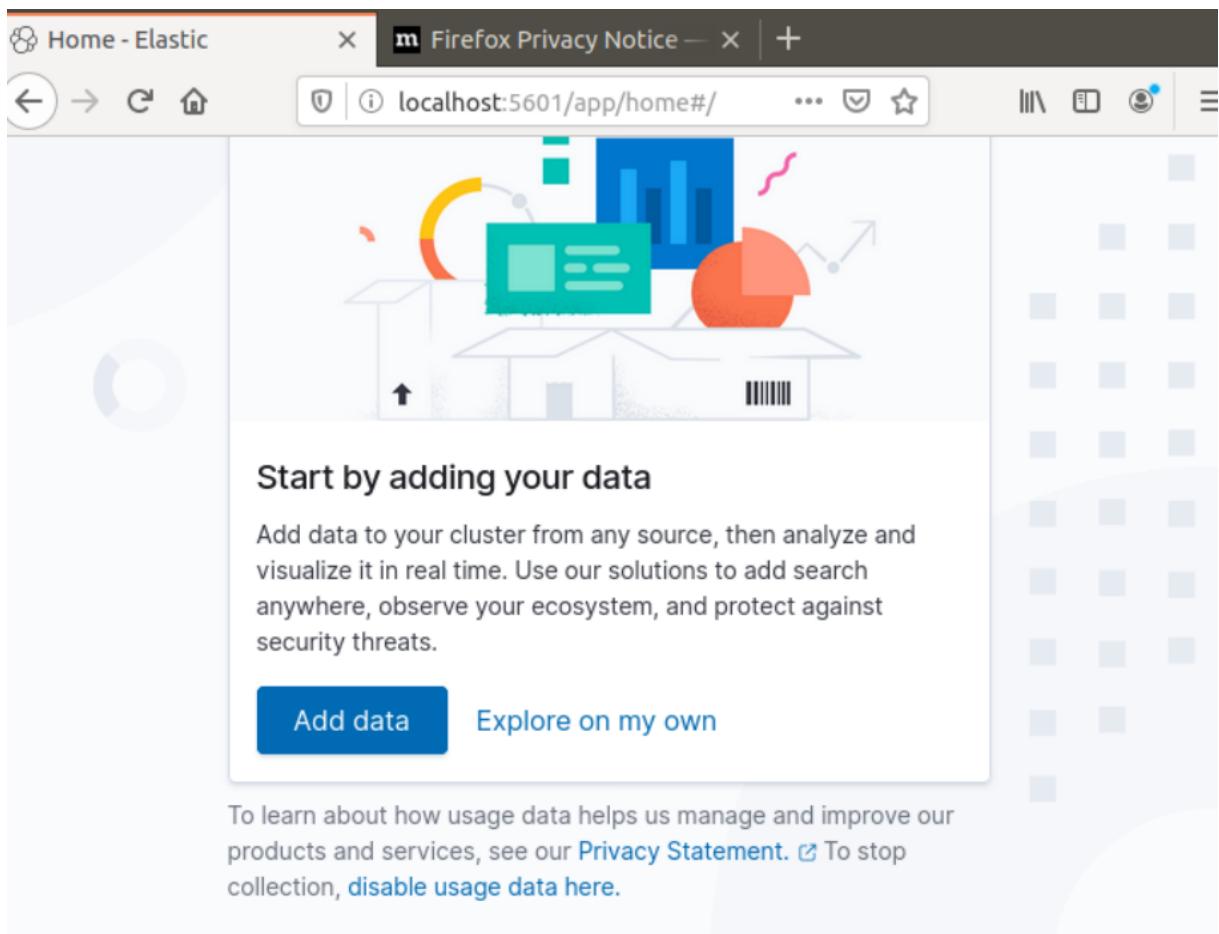
Si le pare-feu UFW est activé ,il faut autoriser le traffic sur le port 5601 pour accéder au tableau de bord de kibana

```
hind@hind-VirtualBox:~$ sudo ufw allow 5601/tcp
Les règles ont été mises à jour
Les règles ont été mises à jour (IPv6)
```

- Testons notre installation :

On accède à notre navigateur et on tape l'adresse :

<https://localhost:5601>



## 5. INSTALLATION ET CONFIGURATION DE LOGSTASH

- Installons Logstash :

On tape la commande suivante :

---

```
hind@hind-VirtualBox:~$ sudo apt-get install logstash
Lecture des listes de paquets... Fait
Construction de l'arbre des dépendances
Lecture des informations d'état... Fait
Les NOUVEAUX paquets suivants seront installés :
  logstash
0 mis à jour, 1 nouvellement installés, 0 à enlever et 398 non mis à jour.
Il est nécessaire de prendre 363 Mo dans les archives.
Après cette opération, 625 Mo d'espace disque supplémentaires seront utilisés.
Réception de :1 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64
  logstash amd64 1:7.13.2-1 [363 MB]
```

- Démarrons et activons Logstash :

```
hind@hind-VirtualBox:~$ sudo systemctl enable logstash
Created symlink /etc/systemd/system/multi-user.target.wants/logstash.service
/etc/systemd/system/logstash.service.
```

- Testons l'état de Logstash :

```
lines 1-11/11 (END)
hind@hind-VirtualBox:~$ sudo systemctl status logstash
● logstash.service - logstash
  Loaded: loaded (/etc/systemd/system/logstash.service; enabled; vendor preset
  Active: active (running) since Sat 2021-06-19 11:47:43 +01; 15s ago
    Main PID: 4330 (java)
      Tasks: 15 (limit: 4666)
     CGroup: /system.slice/logstash.service
             └─4330 /usr/share/logstash/jdk/bin/java -Xms1g -Xmx1g -XX:+UseConcMa
Jun 19 11:47:43 hind-VirtualBox systemd[1]: Started logstash.
Jun 19 11:47:43 hind-VirtualBox logstash[4330]: Using bundled JDK: /usr/share
Jun 19 11:47:44 hind-VirtualBox logstash[4330]: OpenJDK 64-Bit Server VM warn
```

## 6. INSTALLATION DU PLUGIN :

Dans notre mini projet, on a envisagé, comme déjà expliqué, d'effectuer une analyse de sentiments des tweets sur le vaccin de covid 19 afin d'avoir une idée sur les sentiments des gens à l'égard du vaccin puisque plusieurs rumeurs circulent sur ce sujet.

Pour accomplir cette analyse, on aura besoin d'installer un plugin et l'ajouter à Logstash. Ce plugin est développé par une personne, qu'on a trouvé sur GITHUB. Le plugin est nommé « sentimentalizer »

Pour l'installer, il faut se déplacer dans le répertoire

« `/usr/share/logstash/bin` » et taper la commande suivante :

```
hind@hind-VirtualBox:/usr/share/logstash/bin$ sudo ./logstash-plugin install logstash-filter-sentimentalizer
[sudo] Mot de passe de hind :
Using bundled JDK: /usr/share/logstash/jdk
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC
was deprecated in version 9.0 and will likely be removed in
a future release.
Validating logstash-filter-sentimentalizer
Installing logstash-filter-sentimentalizer
Installation successful
```

Listons maintenant la liste des plugins de Logstash :

```
hind@hind-VirtualBox:/usr/share/logstash/bin$ sudo ./logstash-plugin list
Using bundled JDK: /usr/share/logstash/jdk
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC was deprecated in version 9.0 and will likely be removed in a future release.
logstash-codec-avro
logstash-codec-cef
logstash-codec-collectd
logstash-codec-dots
logstash-codec-edn
logstash-codec-edn_lines
logstash-codec-es_bulk
```

```
logstash-filter-mutate
logstash-filter-prune
logstash-filter-ruby
logstash-filter-sentimentalizer
logstash-filter-sleep
logstash-filter-split
logstash-filter-syslog_pri
logstash-filter-throttle
```

Le plugin  
s'est bien  
ajouté

## 7. CONFIGURATION DE LOGSTACH :

Pour configurer Logstash, il suffit de créer un fichier d'extension « **.conf** » qui va contenir le processus d'extraction de données, de traitement et d'intégration dans kibana.

Dans notre cas, on configure Logstash comme suit :

---

```
input {
    twitter {
        consumer_key => "consumer_key"
        consumer_secret => "consumer_secret"
        oauth_token => "oauth_token"
        oauth_token_secret => "oauth_token_secret"
        keywords => [ "Coronavirus", "COVID-19", "vaccin", "pfizer", "astrazeneca", "covid19" ]
        full_tweet => true
    }
}
}

filter{
sentimentalizer{
source => "[extended_tweet][full_text]"
}
}

output {
    elasticsearch {
        hosts => ["localhost:9200"]
        index => "twitter_sentiment_analyse_idx"
    }
}
```

---

## IV. Visualisation avec Kibana :

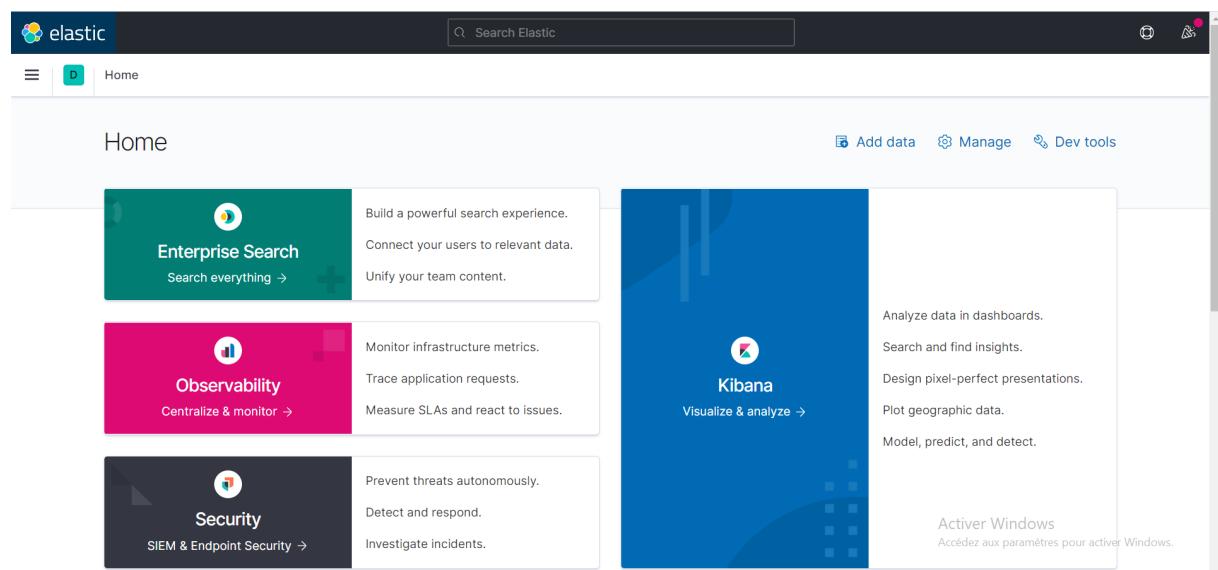
---

### 1. CRÉER UN INDEX PATTERN :

Après la configuration et lancement de Logstash on doit lancer kibana sur notre navigateur :

```
log  [09:45:18.538] [info][server][Kibana][http] http server running at http://localhost:5601  
log  [09:45:18.896] [info][plugins][reporting] Browser executable: C:\Program Files\Elastic\kiba  
[64\x-pack\plugins\reporting\chromium\chrome-win\chrome.exe
```

Avec URL <http://localhost:5601> on visualise l'interface graphique de Kibana comme suit :



The screenshot shows the Elastic Home interface. At the top, there's a navigation bar with the Elastic logo, a search bar labeled "Search Elastic", and three buttons: "Add data", "Manage", and "Dev tools". Below the navigation bar, the word "Home" is displayed. On the left side, there are three main service cards: "Enterprise Search" (teal background), "Observability" (pink background), and "Security" (dark grey background). On the right side, there's a large blue card for "Kibana" with the sub-section "Visualize & analyze" highlighted. Below the Kibana card, there's a button labeled "Activer Windows" with the French text "Accédez aux paramètres pour activer Windows.".

Pour ajouter notre index provenant de Logstash à Kibana, on doit se déplacer sur le menu :

The screenshot shows the Elastic Stack Home page. The top navigation bar includes the elastic logo, a search bar, and a user icon. The left sidebar has sections for Recently viewed (with items like "twitter sentiment extended tweet", "Twitter Analyse Dashboard", "New Dashboard", "test", and "covid19"), Administration, Management (Dev Tools, Fleet, Stack Monitoring, Stack Management), and a local navigation link. The main content area features three large cards: "Search" (Build a powerful search experience, Connect your users to relevant data, Unify your team content), "Infrastructure" (Monitor infrastructure metrics, Trace application requests, Measure SLAs and react to issues), and "Security" (Prevent threats autonomously, Detect and respond, Investigate incidents). A Kibana card on the right says "Visualize & analyze".

## Aller sur Stack management

The screenshot shows the Stack Management page. The top navigation bar includes the elastic logo, a search bar, and a user icon. The left sidebar has links for Rollup Jobs, Transforms, Remote Clusters, and sections for Alerts and Insights (Rules and Connectors, Reporting, Machine Learning Jobs), Kibana (Index Patterns, Saved Objects, Tags, Search Sessions, Spaces, Advanced Settings), and Stack (License Management, 8.0 Upgrade Assistant). A message in the center states: "A complete list of apps is in the menu on the left."

Puis on choisit index Management :

Ingest

Ingest Node Pipelines

Data

Index Management

Index Lifecycle Policies

Snapshot and Restore

Rollup Jobs

Transforms

Remote Clusters

Alerts and Insights

Rules and Connectors

Index patterns

Create and manage the index patterns that help you retrieve your data from Elasticsearch.

Search...

Pattern ↑

logstash-\* Default

idx\_twitter

logstash-2021.06.24-000001

+ Create index pattern

On clique sur “Create index pattern” et on écrit le nom de notre index qui existe sur le fichier de configuration Logstash :

Ingest

Ingest Node Pipelines

Data

Index Management

Index Lifecycle Policies

Snapshot and Restore

Rollup Jobs

Transforms

Remote Clusters

Alerts and Insights

Rules and Connectors

Create index pattern

An index pattern can match a single source, for example, filebeat-4-3-22, or multiple data sources, filebeat-\*.

[Read documentation](#)

Step 1 of 2: Define an index pattern

Index pattern name

index-name-\*

Use an asterisk (\*) to match multiple indices. Spaces and the characters \, /, ?, ., <, >, | are not allowed.

X Include system and hidden indices

Next step >

Pour être sûr que notre index a été bien créé, on ouvre index management et on visualise que notre index existe parmi les données :

Indices Data Streams Index Templates Component Templates

Update your Elasticsearch indices individually or in bulk. [Learn more.](#)

X Include rollup indices  X Include hidden indices

Search

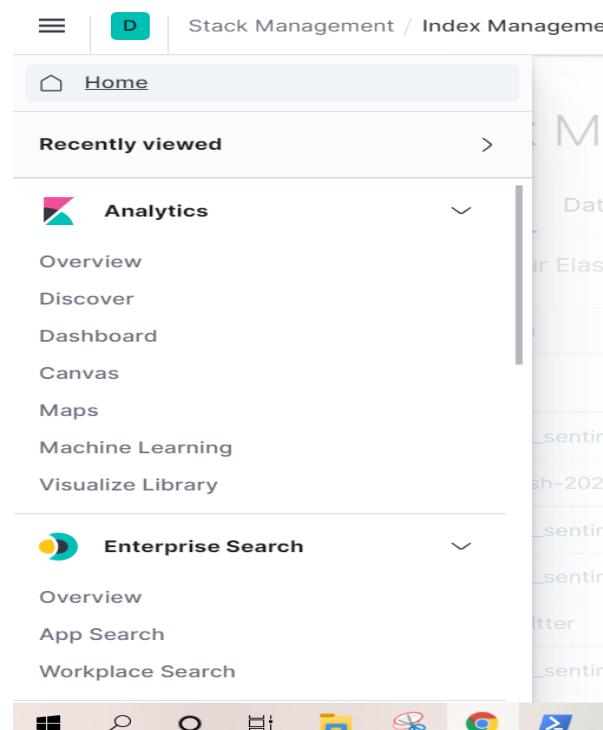
Lifecycle status Lifecycle phase Reload indices

Name	Health	Status	Primaries	Replicas	Docs count	Storage size	Data stream
twitter_sentiment	yellow	open	1	1	22599	106.6mb	
logstash-2021.06.24-000001	yellow	open	1	1	0	208b	
twitter_sentiment_anal	yellow	open	1	1	8041	40.4mb	
twitter_sentiment_analyse	yellow	open	1	1	16107	78.7mb	
idx_twitter	yellow	open	1	1	7443	41.7mb	
twitter_sentiment_analyse_idx	yellow	open	1	1	42698	191mb	

Rows per page: 10 < 1 >

## 2. VISUALISER L'INDEX :

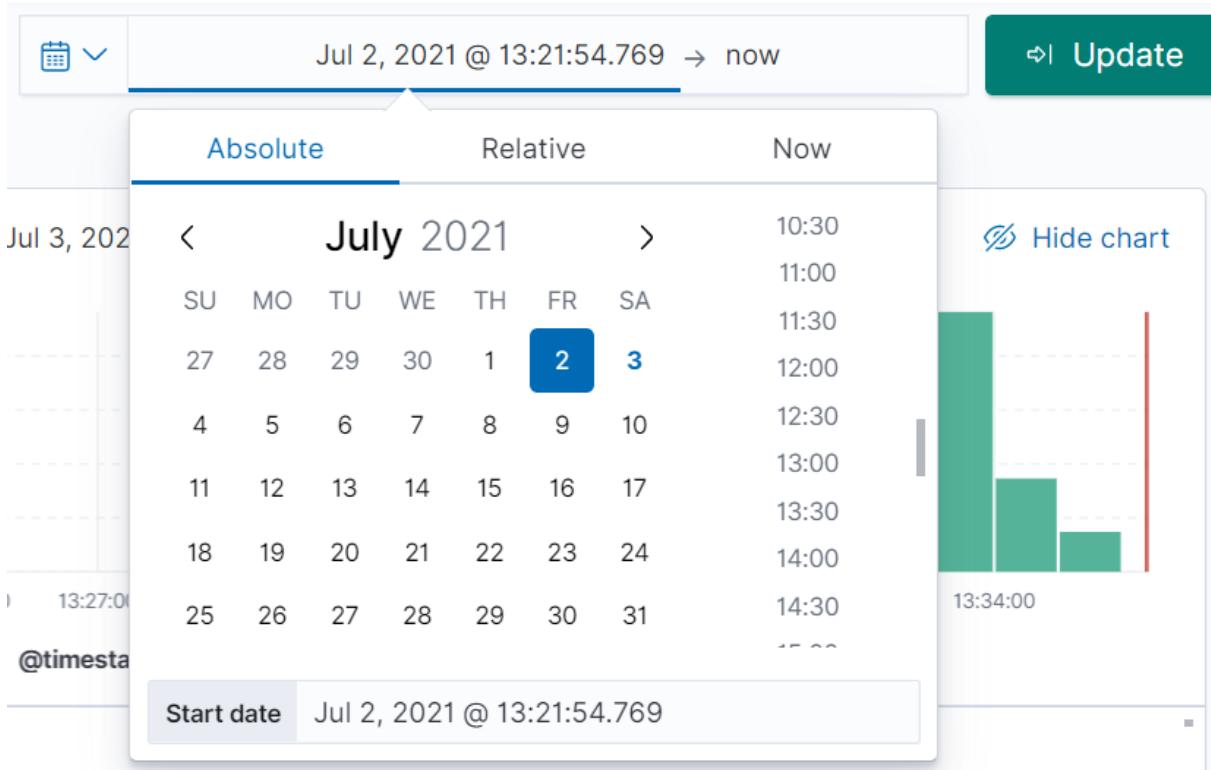
Pour visualiser notre index, on passe au menu Discover :



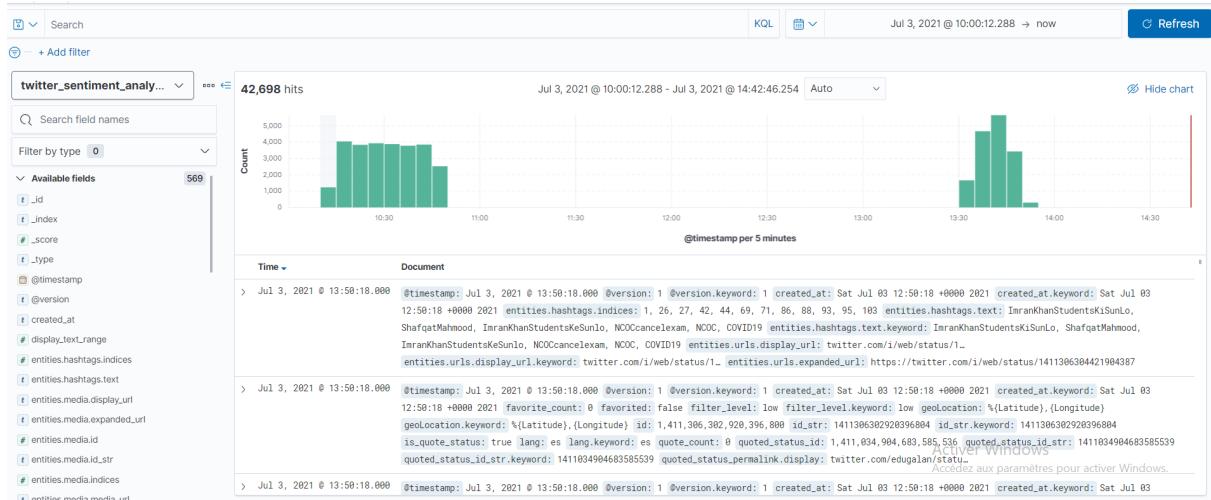
On doit vérifier d'abord que l'index choisi correspond bien à notre index.

A screenshot of the Elasticsearch Discover interface. The top bar includes 'Discover', 'Options', 'New', 'Save', 'Open', 'Share', and 'Inspect'. The search bar contains 'logstash-\*'. The results section displays a message: 'No results match your search criteria'. Below this, a note says 'Expand your time range' and explains that the current time range may not contain data. The left sidebar shows filter options: 'logstash-\*' (selected), 'Search field names', 'Filter by type' (0), and 'Available fields'.

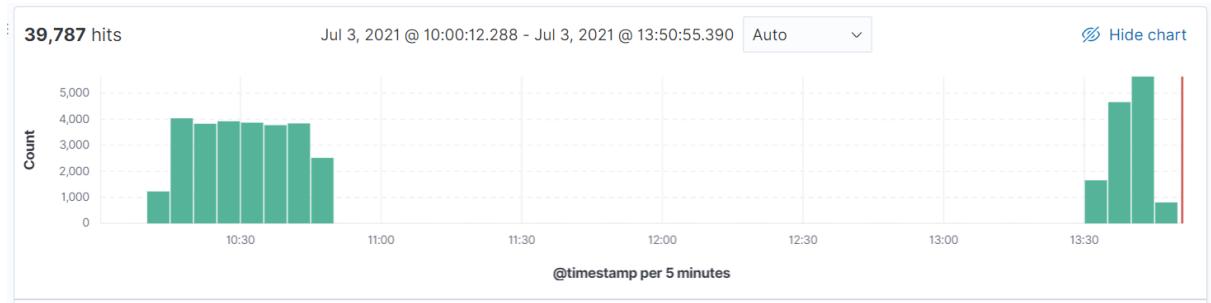
Et aussi on doit changer la date et l'heure pour afficher les données déjà enregistrés.



Le résultat final nous permet de visualiser toutes les données stockées et aussi celles en train d'être collecte depuis Elasticsearch :



Ce diagramme donne la distribution des enregistrements envoyés par Logstash d'une manière interactive.



Les différents champs créés sont stocké dans available fields.

On peut chercher facilement un champ spécifique soit en tapant son nom ou en appliquant des filtres sur les champs.

(?) — + Add filter

**twitter\_sentiment\_analy...** ▼

🔍 Search field names

Filter by type 0 ▼

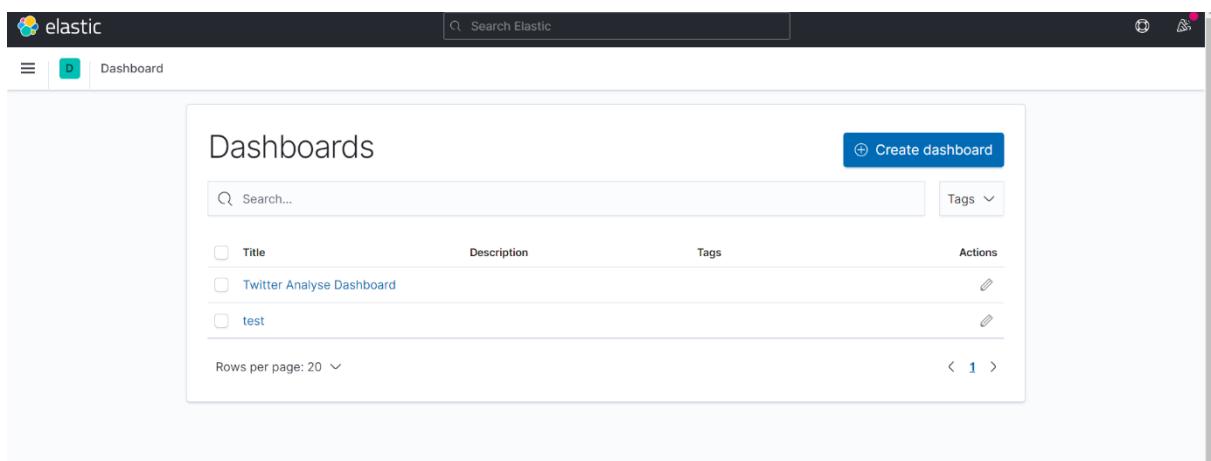
Available fields 574

- t \_id
- t \_index
- # \_score
- t \_type
- 📅 @timestamp
- t @version
- t created\_at
- # display\_text\_range
- # entities.hashtags.indices
- t entities.hashtags.text

### 3. DASHBOARD :

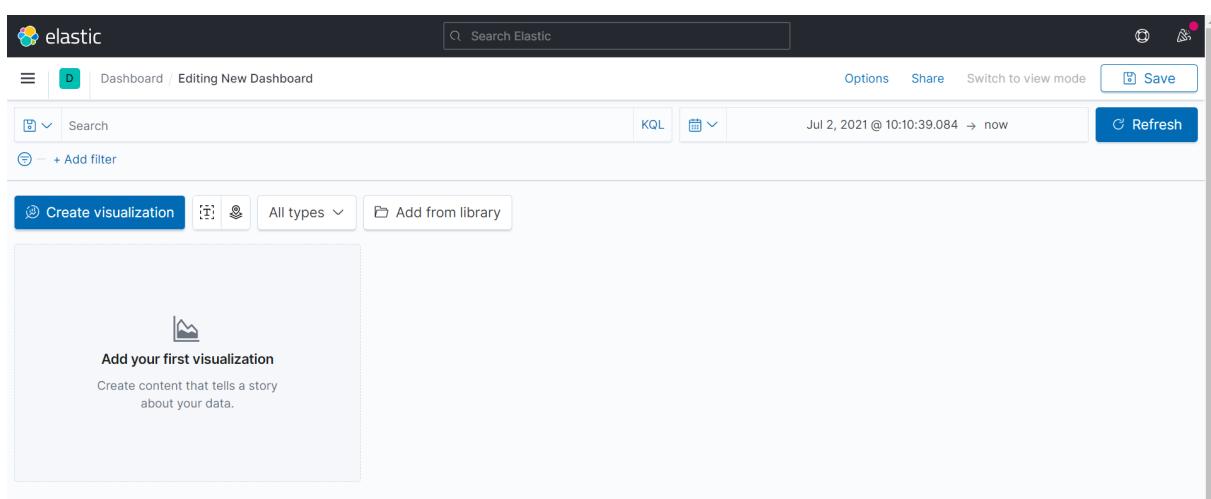
La meilleure façon de comprendre nos données est de les visualiser. Avec une Dashboard, on peut transformer nos données d'un index à des visualisations plus significatives et claires pour ensuite pouvoir sortir avec des conclusions pertinentes et correctes.

Sur Kibana, pour créer un dashboard, on doit aller sur le menu et cliquer sur “dashboard” puis une interface comme suit va apparaître :



The screenshot shows the Kibana interface with the title "Dashboards". At the top right is a blue button labeled "+ Create dashboard". Below it is a search bar and a "Tags" dropdown. A table lists three existing dashboards: "Twitter Analyse Dashboard" and "test". Each entry has an edit icon in the "Actions" column. At the bottom left is a "Rows per page" dropdown set to 20, and at the bottom right are navigation arrows.

On clique sur “Create dashboard” puis on peut maintenant créer les différentes visualisations.

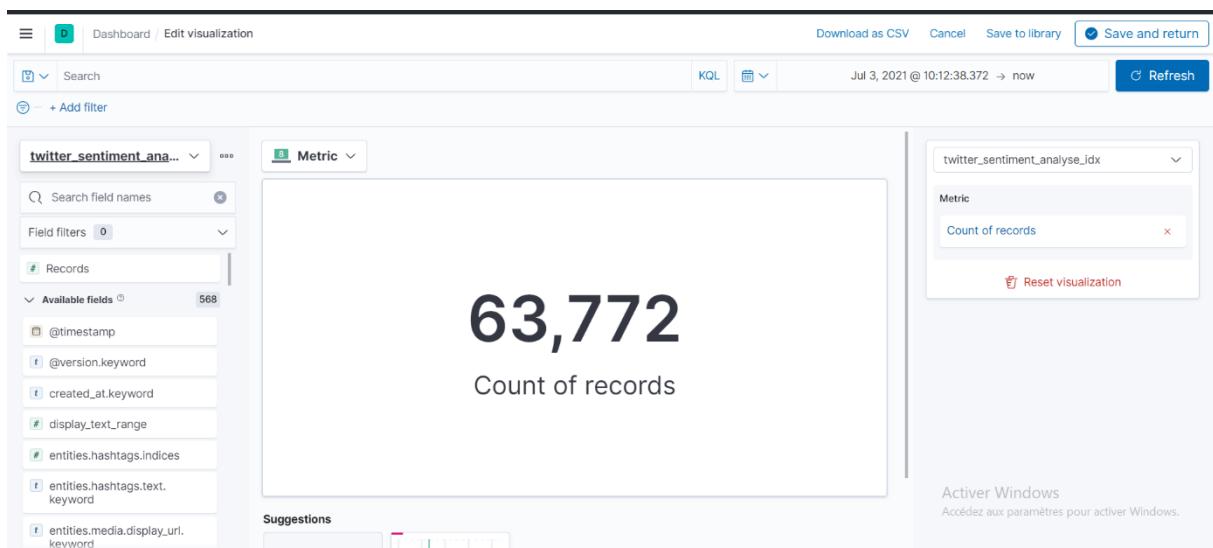


The screenshot shows the Kibana interface for creating a new dashboard. At the top, it says "Dashboard / Editing New Dashboard". It includes a search bar, date range selector, and time range from "Jul 2, 2021 @ 10:10:39.084 → now". There are buttons for "Options", "Share", "Switch to view mode", "Save", and "Refresh". Below the header are buttons for "Create visualization", "Search", "KQL", and "Add filter". A large central area is labeled "Add your first visualization" with the text "Create content that tells a story about your data." and a small chart icon.

À chaque fois qu'on crée une visualisation il faut vérifier que l'index utilisé correspond bien à notre index et ajuster l'intervalle du temps.

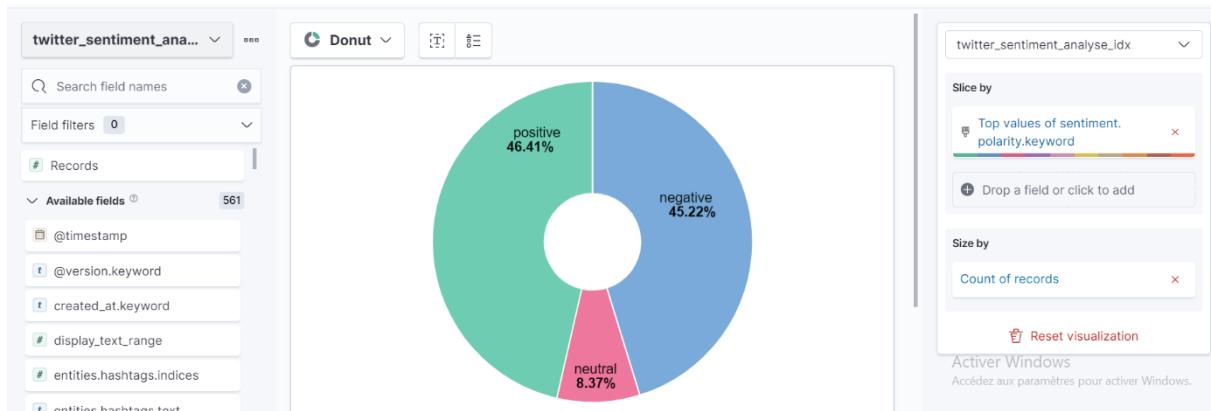
## METRIC

On peut visualiser le nombre des enregistrements collectés grâce à la visualisation Metric et un count sur les données.



## PIE CHART

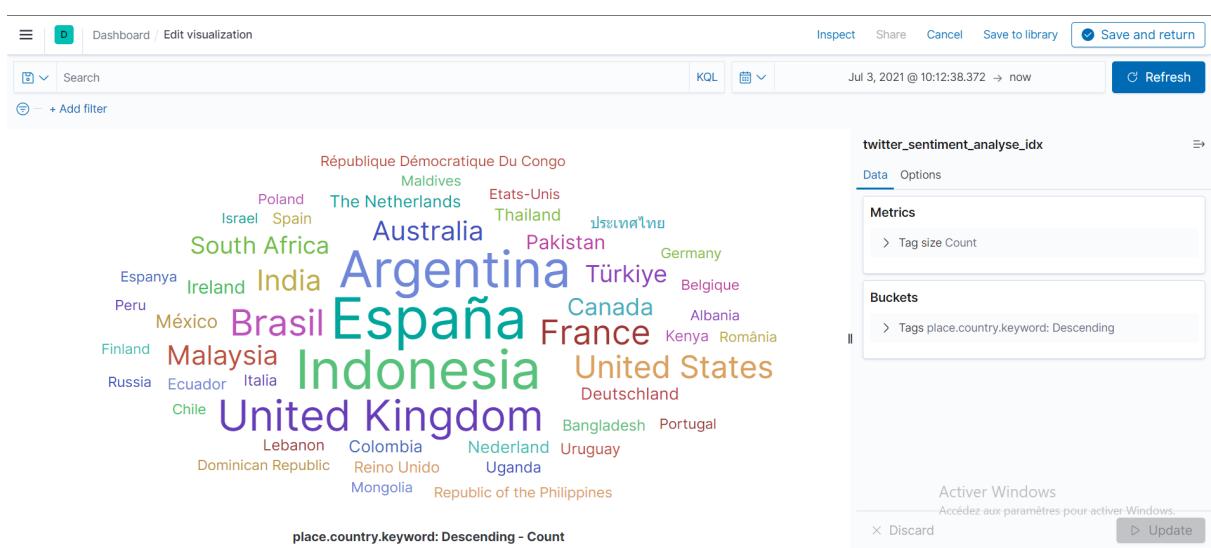
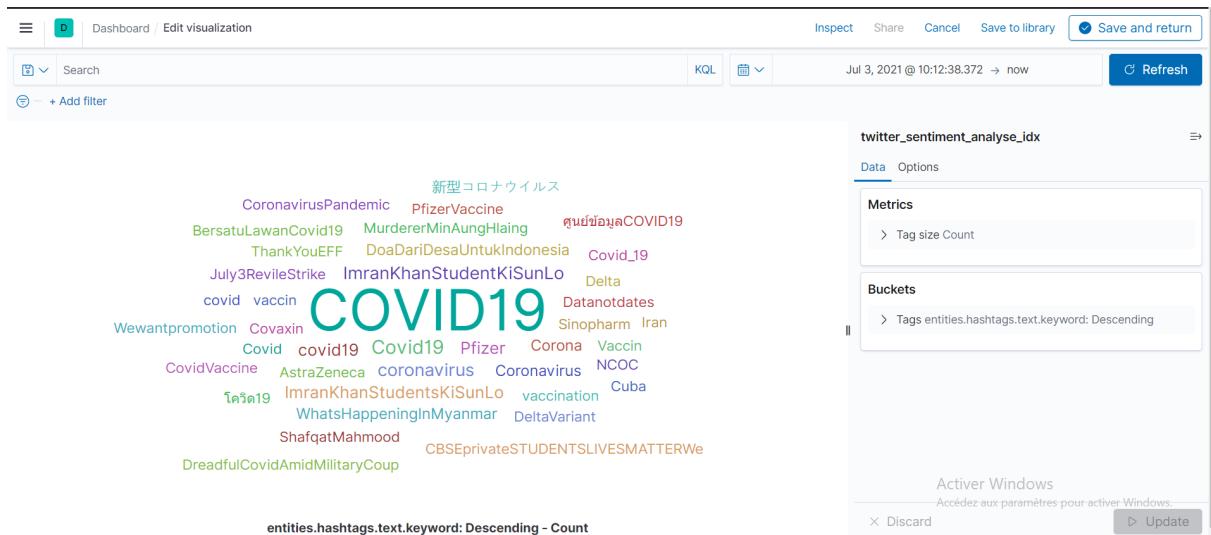
On visualise la distribution des différents sentiments en choisissant comme champs de découpage sentiment.polarity.keyword avec un count sur les enregistrements.



## TAG CLOUD

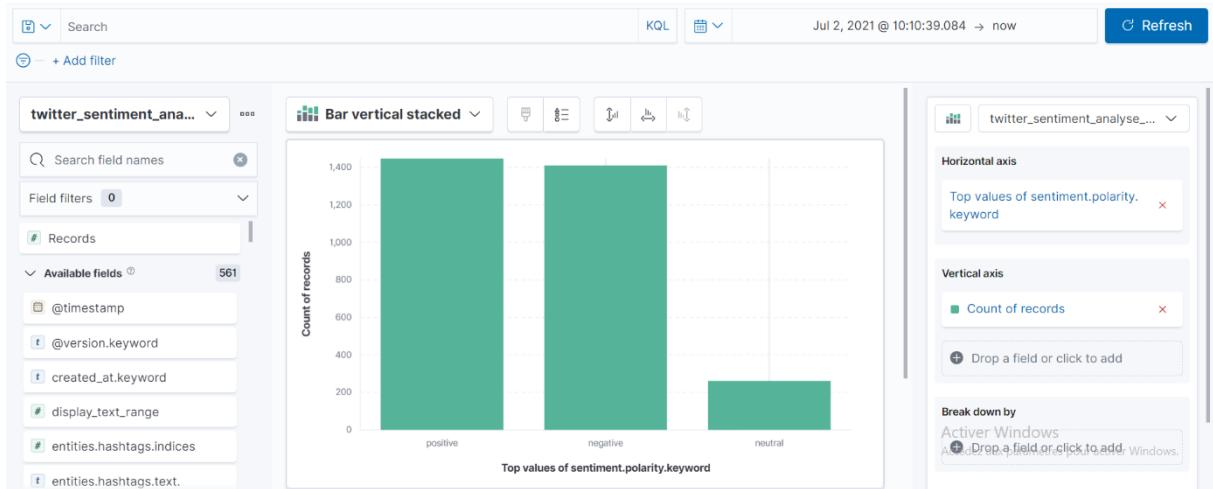
Est une nouvelle représentation visuelle de données textuelles, généralement utilisée pour décrire des métadonnées de mots clés (tags) sur des sites Web ou pour visualiser un texte de forme libre.

On a choisie de faire un tag cloud des hashtags keyword utilisés et aussi des pays d'où provient les tweets.



# VERTICAL BAR

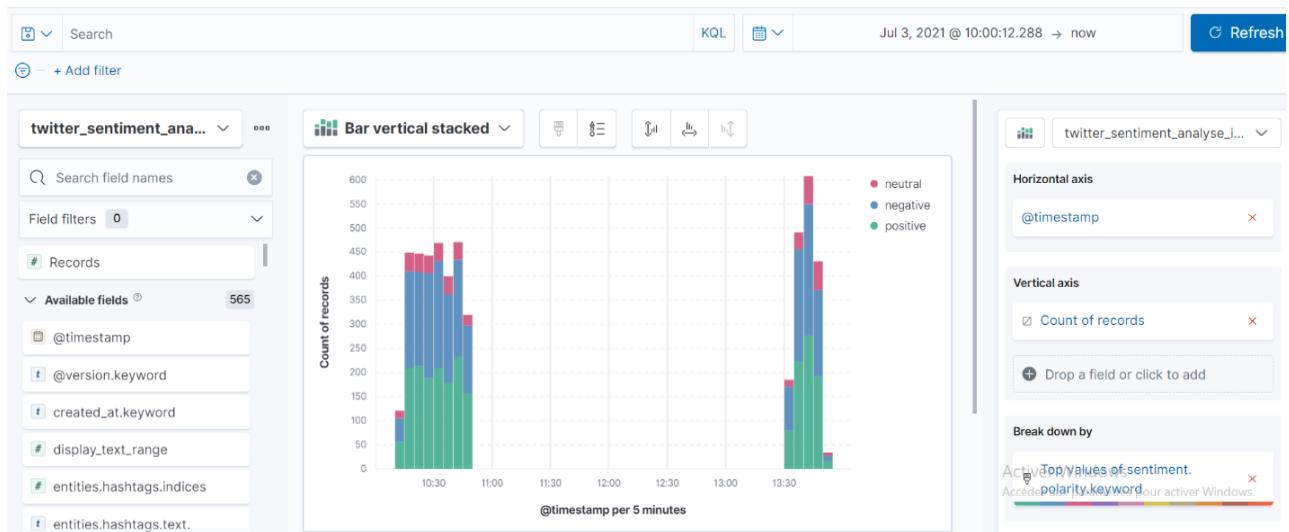
Pour créer verticaux bars, il suffit de choisir les champs à mettre dans les différents axes



## BAR VERTICAL STACK

Ce graphe permet de visualiser la distribution de nombre des enregistrements dans le temps et pour chaque bar déterminer le pourcentage des sentiments.

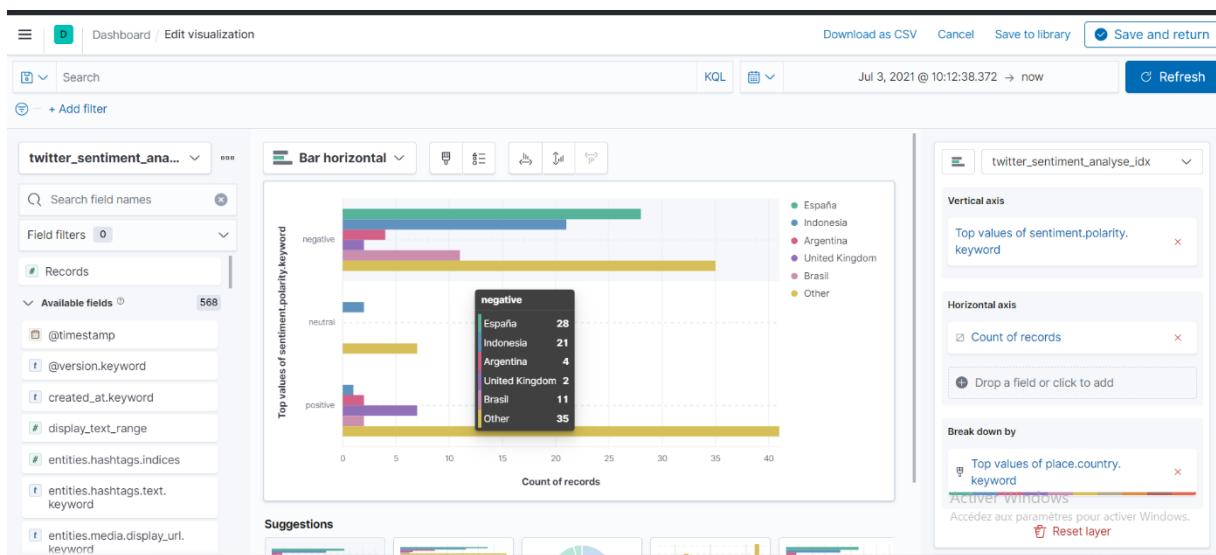
Pour cela en choisit comme axis horizontal le temps (@timestamp) et pour l'axis vertical on choisit un count sur les enregistrements avec le champ sentiment.polarity.keyword pour diviser les bars.



## BAR HORIZONTAL STACKED

Ce graphe permet de visualiser la distribution de nombre des enregistrements pour les différents pays et pour chaque barre déterminer le pourcentage des catégories sentimentale.

Pour cela on choisit comme axis vertical les différents sentiments et pour l'axis horizontal, on choisit un “count” sur les enregistrements avec le champ “place.country. keyword”, qui désigne les pays d'où provient les tweets, pour diviser les bars.



# DASHBOARD

On visualise la version complète de notre tableau de bord.



---

## V. CONCLUSION

---

Ce projet donne un aperçu général sur les sentiments générés sur Twitter à l'égard du vaccin et de COVID-19.

Cela peut n'être qu'un point de départ, car ELK ouvre de nombreuses possibilités pour l'analyse de données.