TP9 : Géolocalisation

Dans ce TP, nous mettons en place une pipeline destinée à traiter les traces d'accès d'un serveur web. La pipeline utilisera entre autres le processeur *geo_ip* qui géolocalise à partir d'une adresse IP.

Une fois la pipeline mise en place, nous ingérons un fichier de log complet et effectuerons des requêtes d'agrégations

Installation des plugins ***ingest-geoip*** et ***ingest-user-agent*** (Pas nécessaire si v7.x)

Mise en place de la pipeline :

```
PUT _ingest/pipeline/access_log
{
  "description" : "Ingest pipeline for Combined Log Format",
  "processors" : [
    {
      "grok": {
        "field": "message",
        "patterns": ["%{IPORHOST:clientip} %{USER:ident} %{USER:auth} \\[%
{HTTPDATE:timestamp}\\] \"%{WORD:verb} %{DATA:request} HTTP/%
{NUMBER:httpversion}\" %{NUMBER:response:int} (?:-|%{NUMBER:bytes:int}) %
{QS:referrer} %{QS:agent}"]
      }
    },
    {
      "date": {
        "field": "timestamp",
        "formats": [ "dd/MMM/YYYY:HH:mm:ss Z" ]
      }
    },
    {
      "geoip": {
        "field": "clientip"
      }
    },
    {
      "user_agent": {
        "field": "agent"
      }
    }
```

```
  ]
}
```

Tester la pipeline avec :

```
POST _ingest/pipeline/access_log/_simulate
{
  "docs": [
    {
      "_source": {
        "message": "212.87.37.154 - - [12/Sep/2016:16:21:15 +0000] \"GET
/favicon.ico HTTP/1.1\" 200 3638 \"-\" \"Mozilla/5.0 (Macintosh; Intel Mac OS X
10_11_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116
Safari/537.36\""
      }
    }
  ]
}
```

Utiliser le programme fourni ou Ecrire son propre programme qui lit chaque ligne du fichier de log fourni et le fournit à elastic search dans un index nommé ***apache_access***

Visualiser l'index et écrire :

- des requêtes de type filtre par *geo_bounding_box*
- des requêtes d'agrégations par *geo_distance*