anrt
agence nationale de réglementation des télécommunications
الوكالة الوطنية لتقنين المواصلات

INPT
المعهد الوطني للبريد والمواصلات
Institut National des Postes et Télécommunications

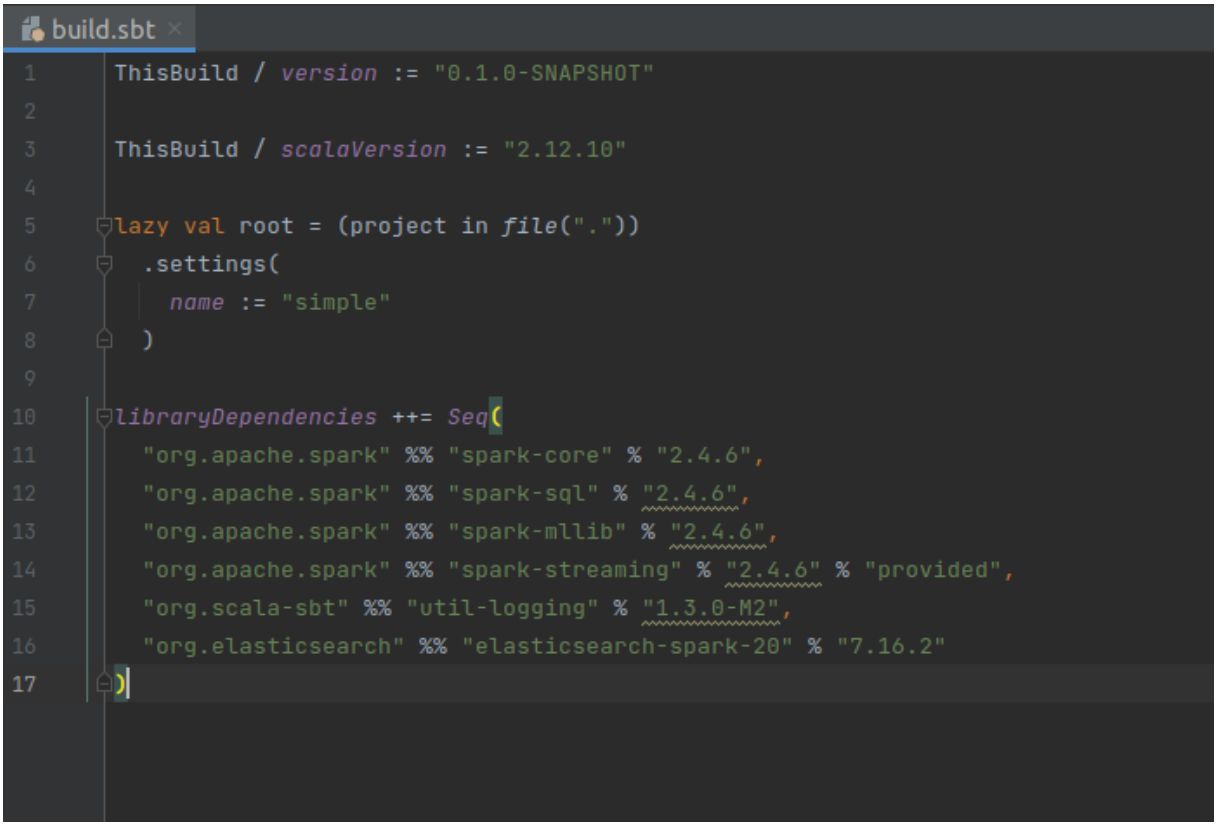
Projet scala :

Traitement en temps réel avec spark streaming

Réaliser par :

BENAICH Ayyoub

Nous ajoutons spark dépendances pour l’utiliser comme une librairie dans notre application scala



Importation du librairies nécessaire pour le projet

```
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.types._
import org.apache.spark.streaming
import org.apache.spark.sql.functions._
import org.apache.log4j._
import org.elasticsearch.spark.sql._
```

Création d'une spark session

```
object Main {

  def main(args: Array[String]): Unit = {

    val spark = SparkSession.builder().master( master = "local[*]").appName( name = "first")
      .config("spark.es.nodes","localhost")
      .config("spark.es.port","9200")
      .getOrCreate()
    Logger.getLogger( name = "org").setLevel(Level.ERROR)
```

Création du dataframe

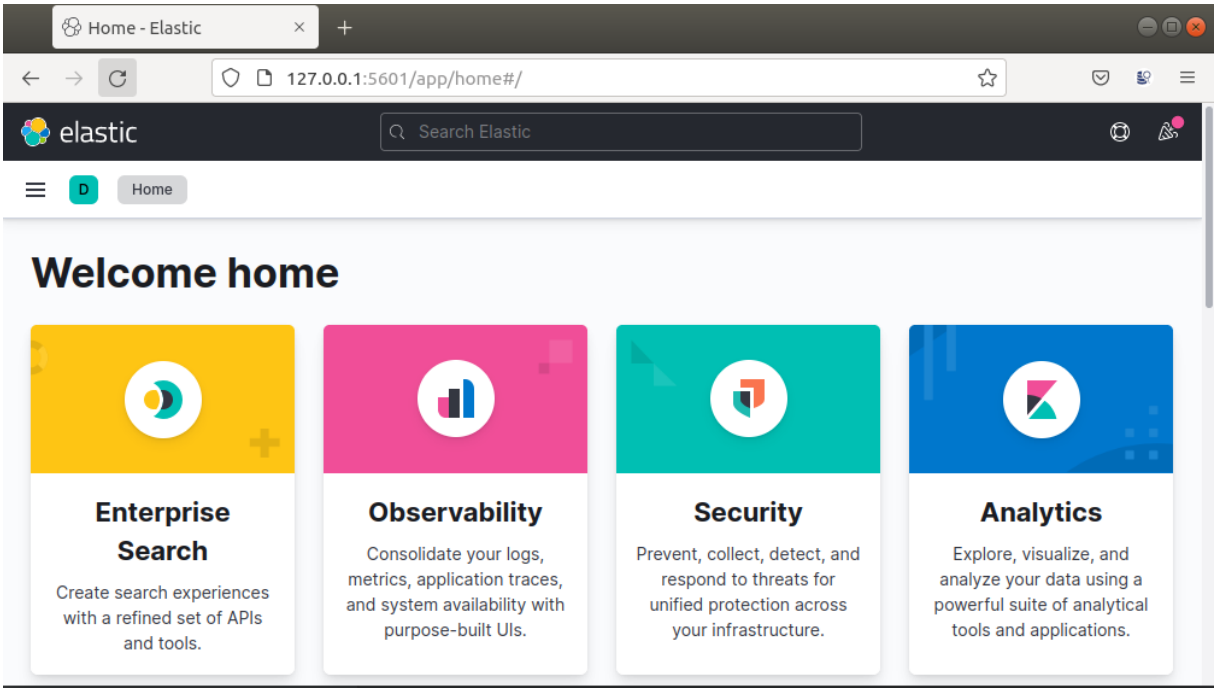
```
  //Create DataFrame
  val StreamDF = spark
    .readStream.option("delimiter", " ")
    .schema(schema)
    .csv( path = "/home/benaich/Documents/simple/logs")
  StreamDF.createOrReplaceTempView( viewName = "SDF")
  val outDF = spark.sql( sqlText = "select * from SDF")
```

Transfère du dataframe vers elasticSearch en mode streaming

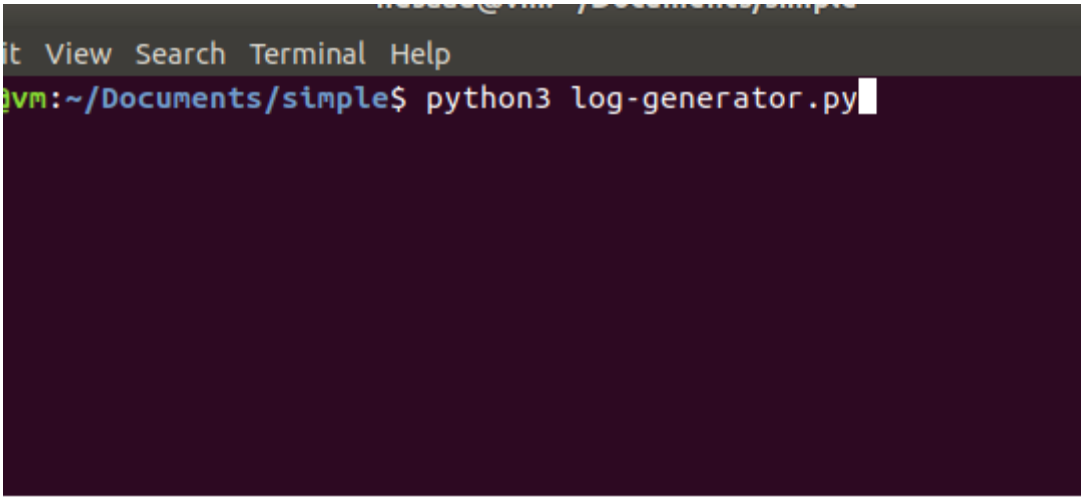
```
//write DF to elasticSearch
var query = outDF.writeStream
.outputMode( outputMode = "append")
.queryName( queryName = "writing_to_es")
.format( source = "org.elasticsearch.spark.sql")
.option("checkpointLocation", "/tmp/")
.option("es.resource", "logs/doc")
.option("es.nodes", "localhost")
.start()

query.awaitTermination()
```

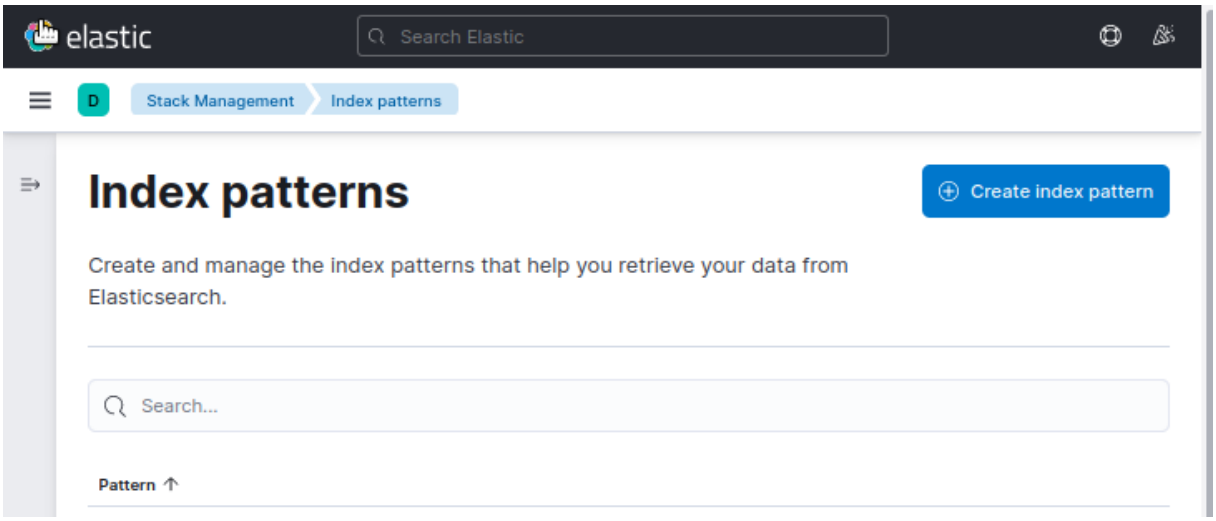
elasticSearch visualisation



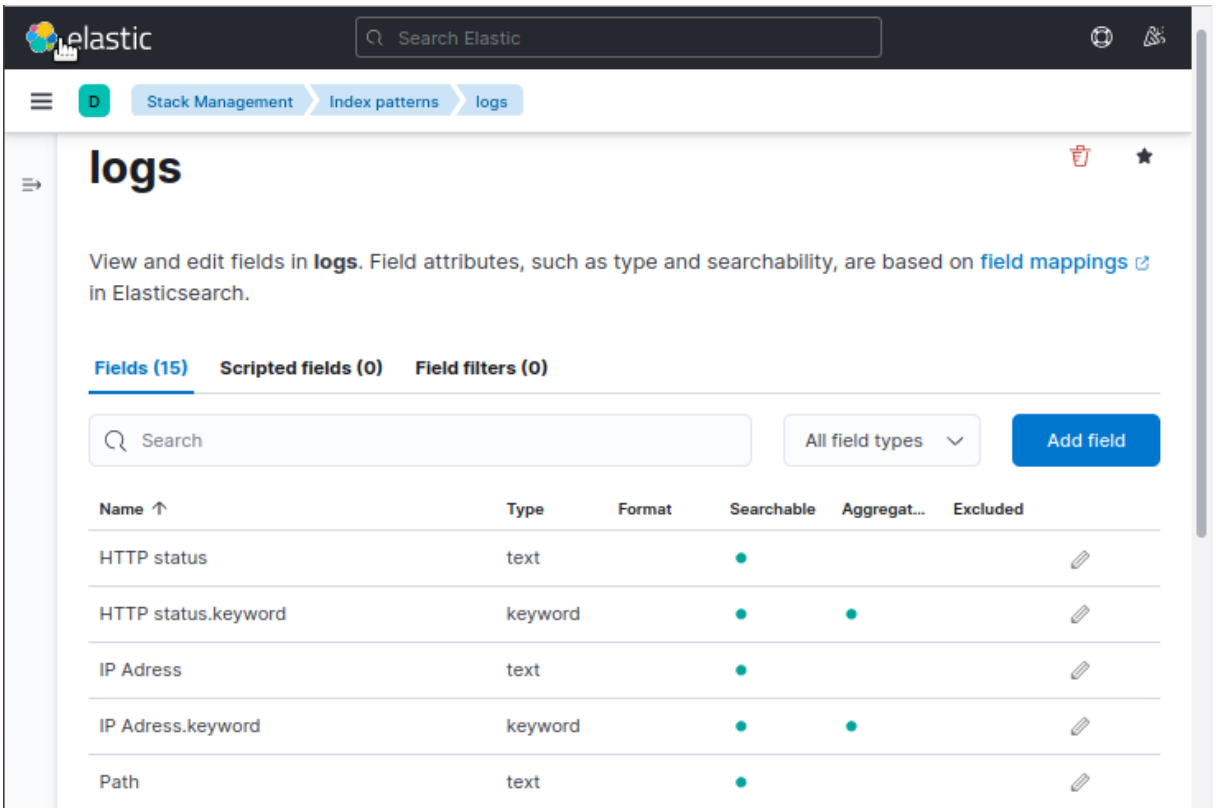
Lancement du log



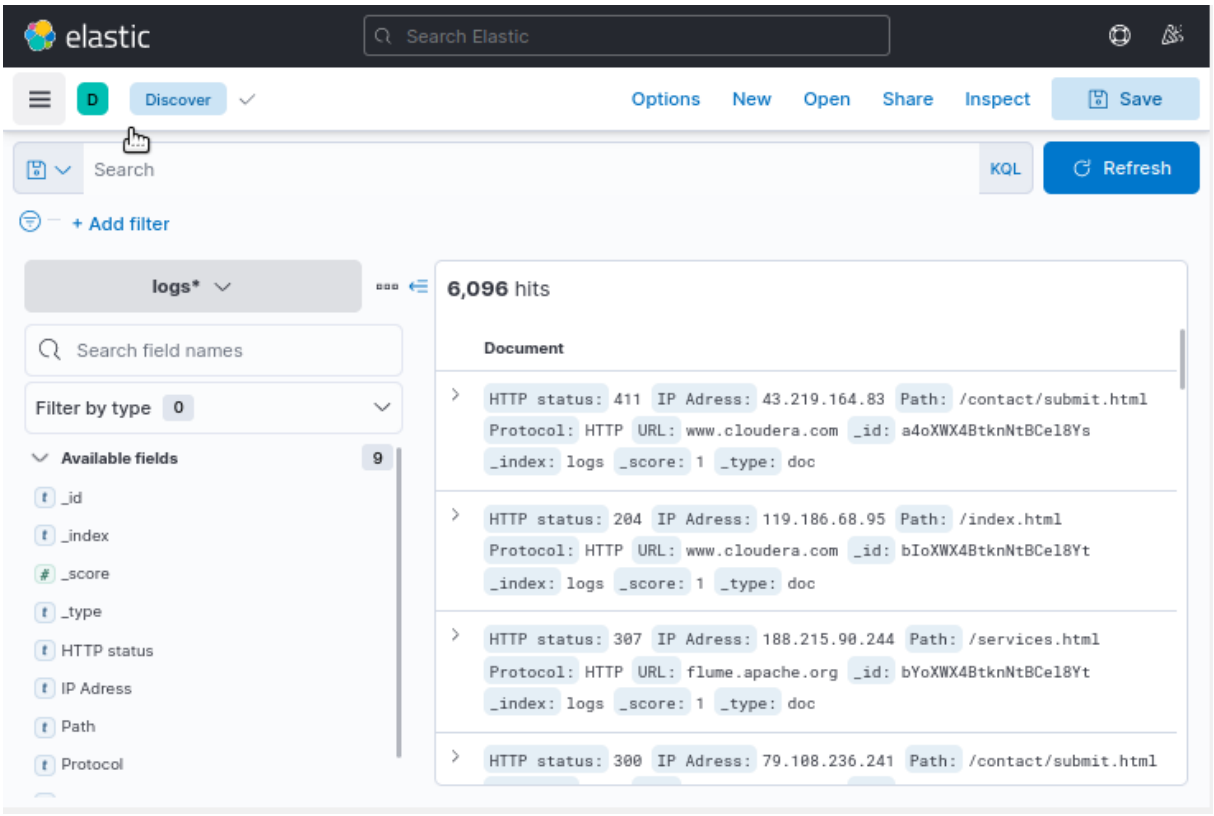
Puis en transfère les log vers le répertoire



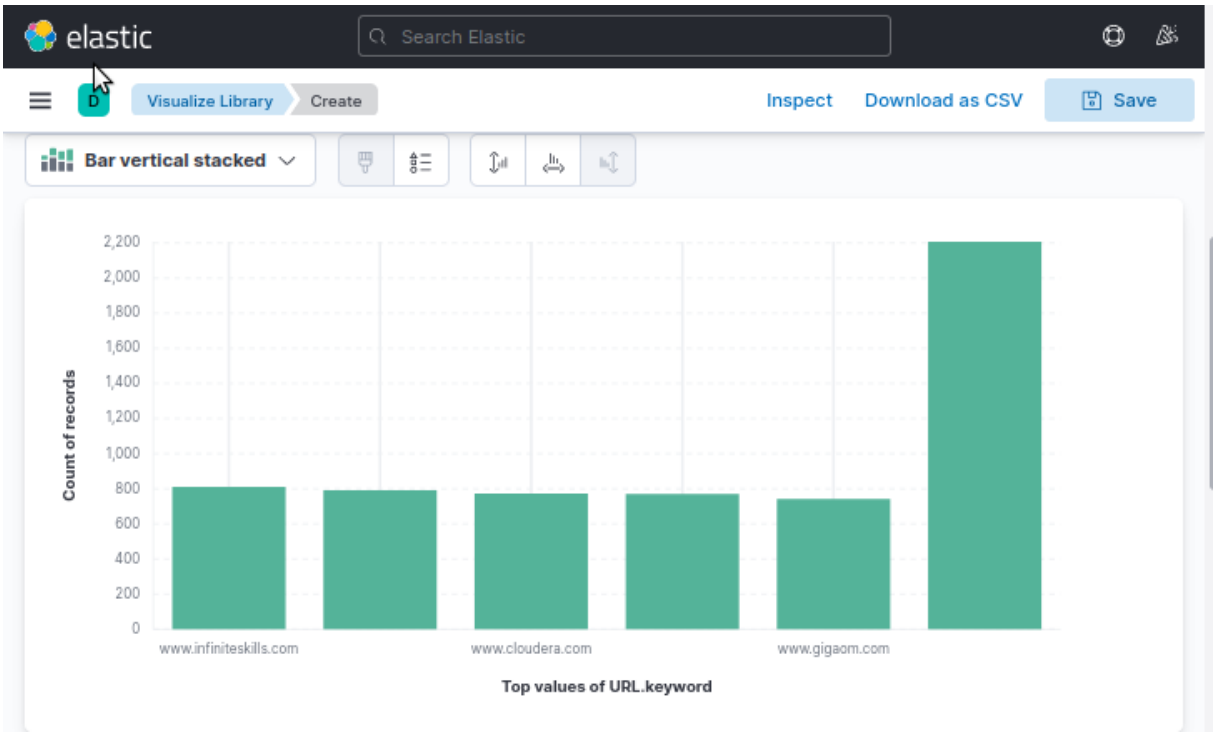
Récupération du donnes par sur kibana

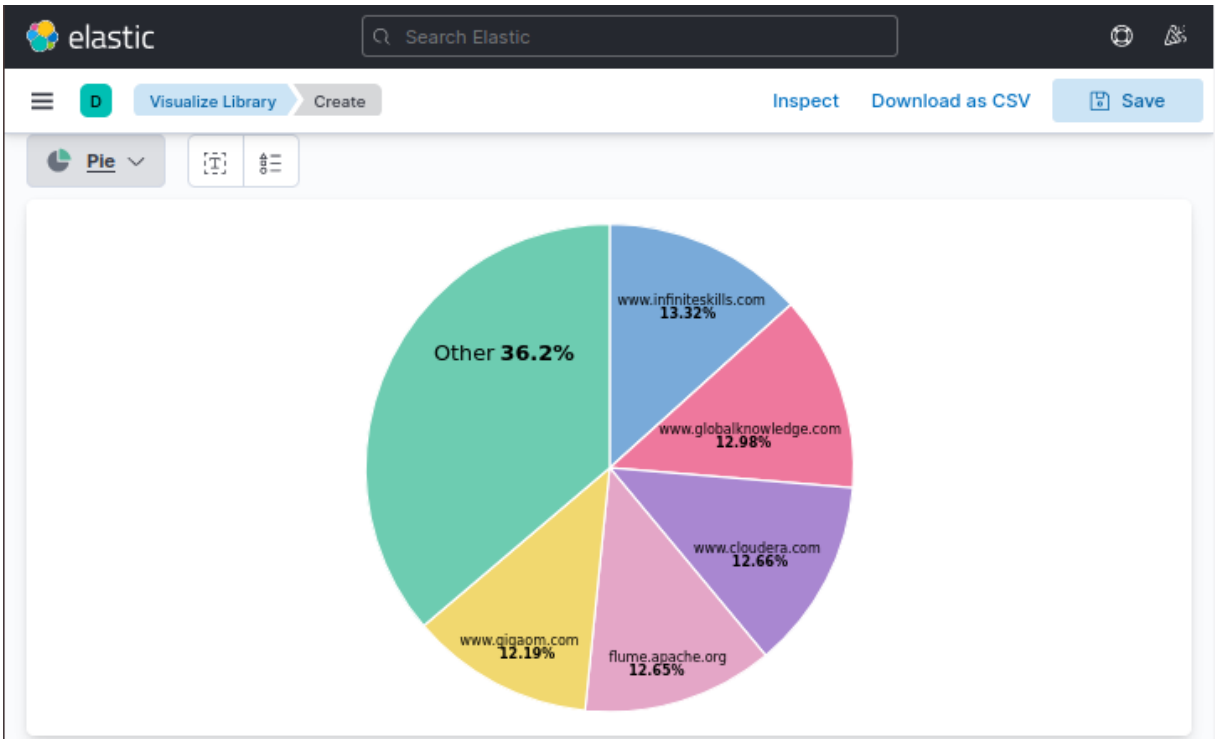


On a reçu 6096 enregistrements ,on va faire la visualisation de ces enregistrements

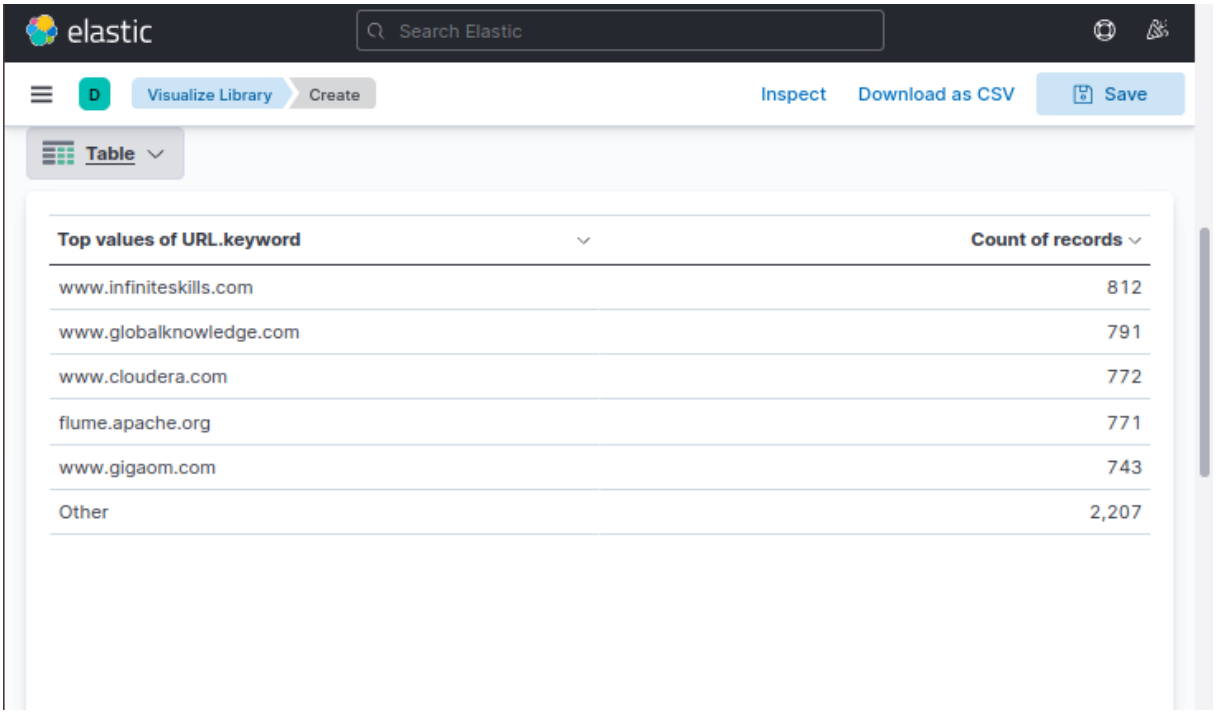


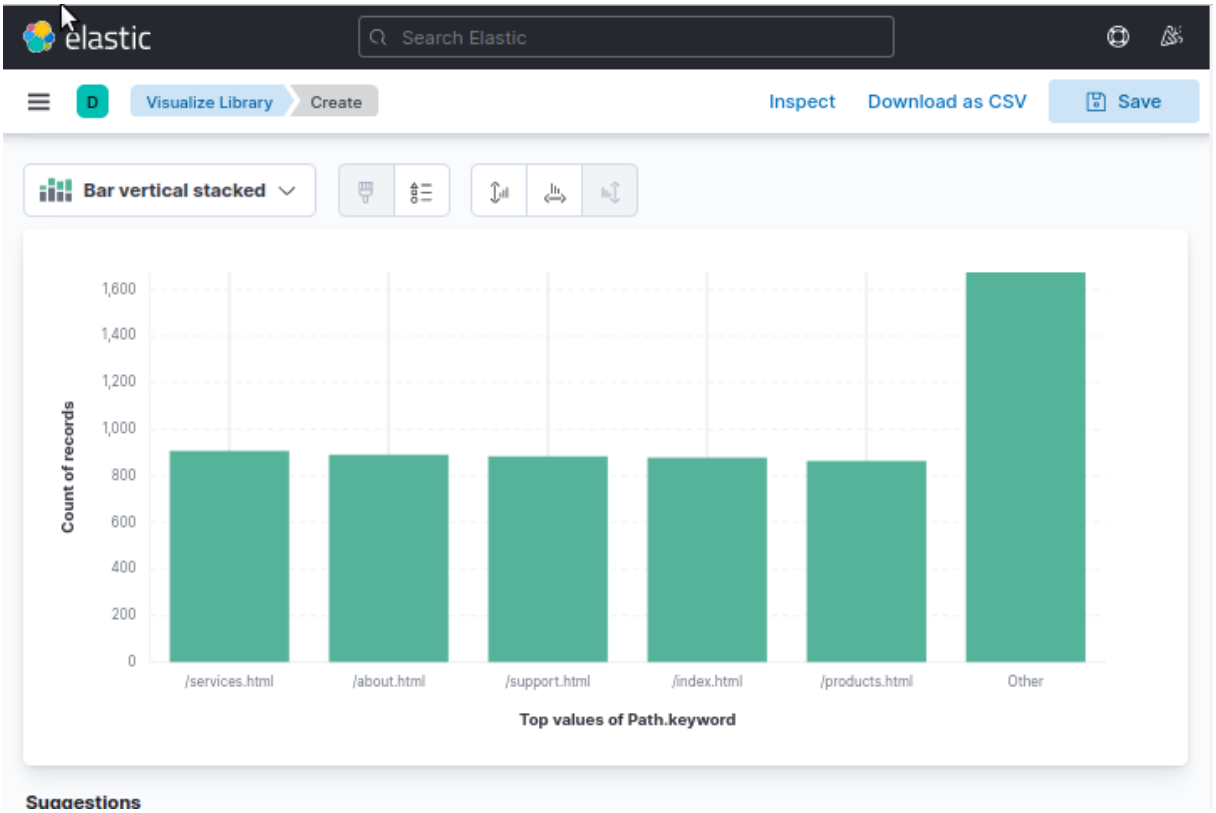
Visualisation :

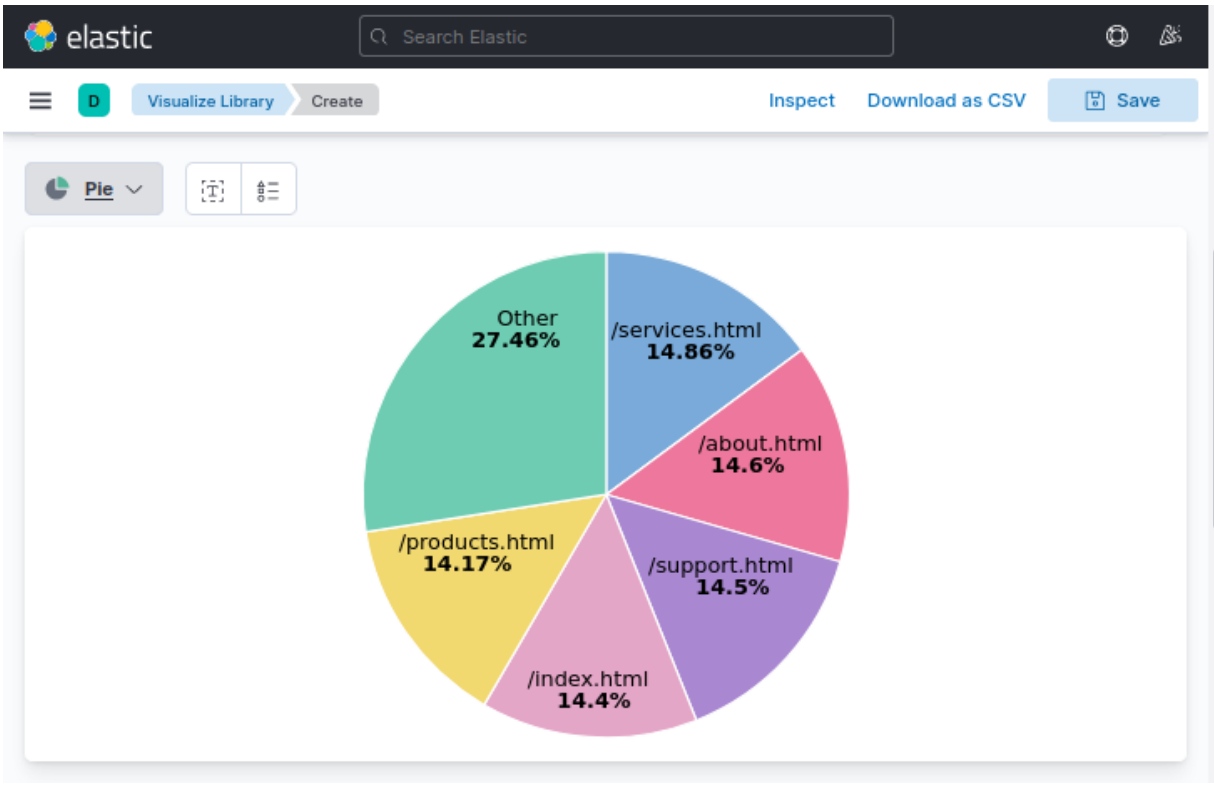


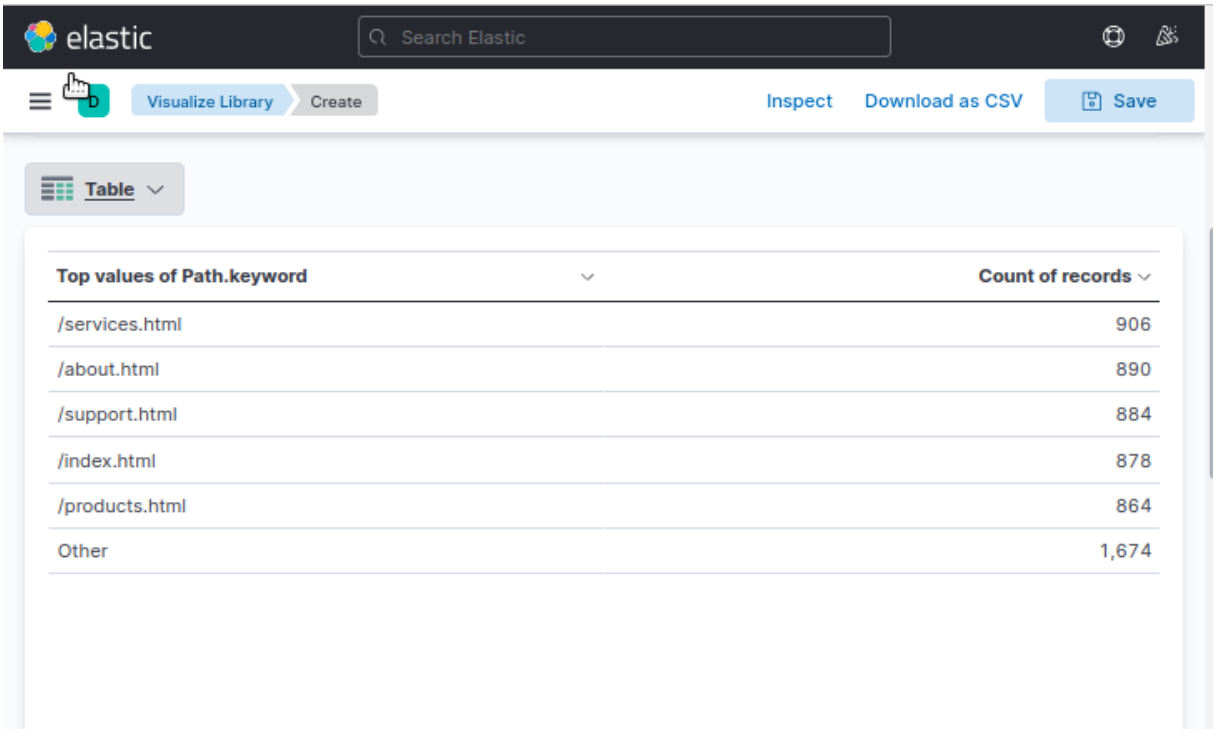


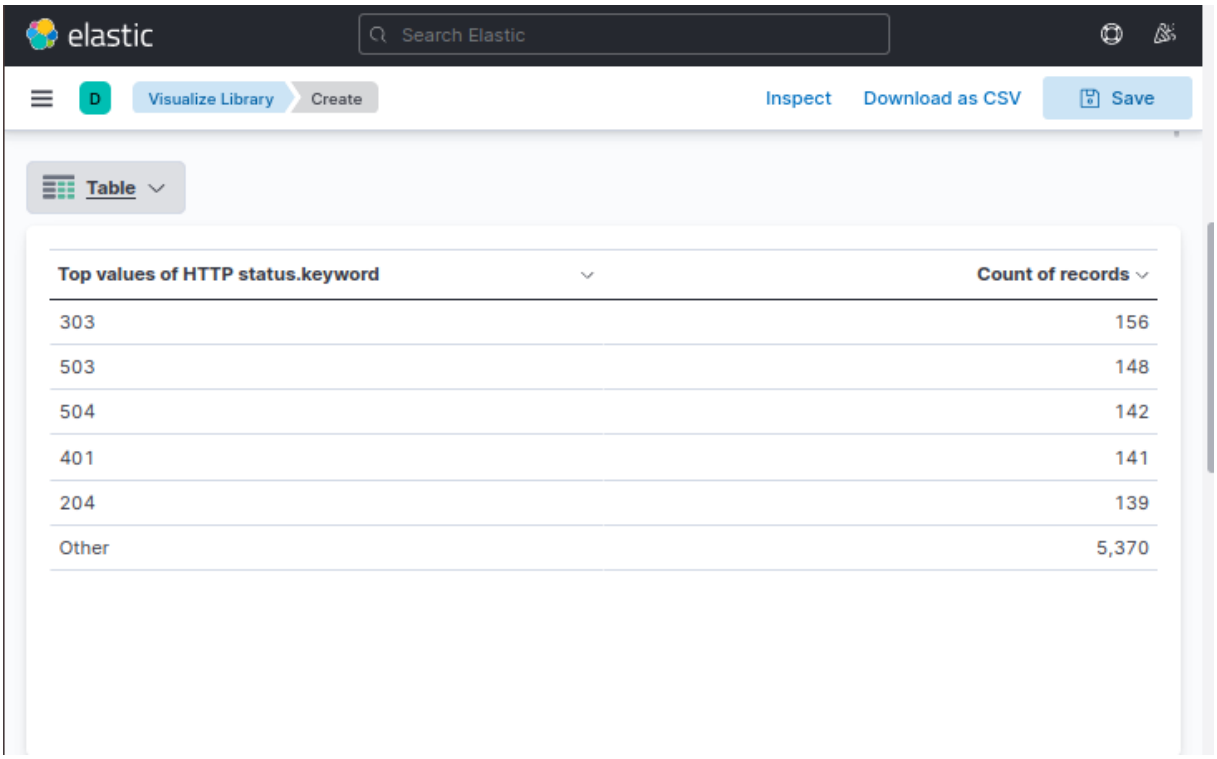
Nombres des enregistrement

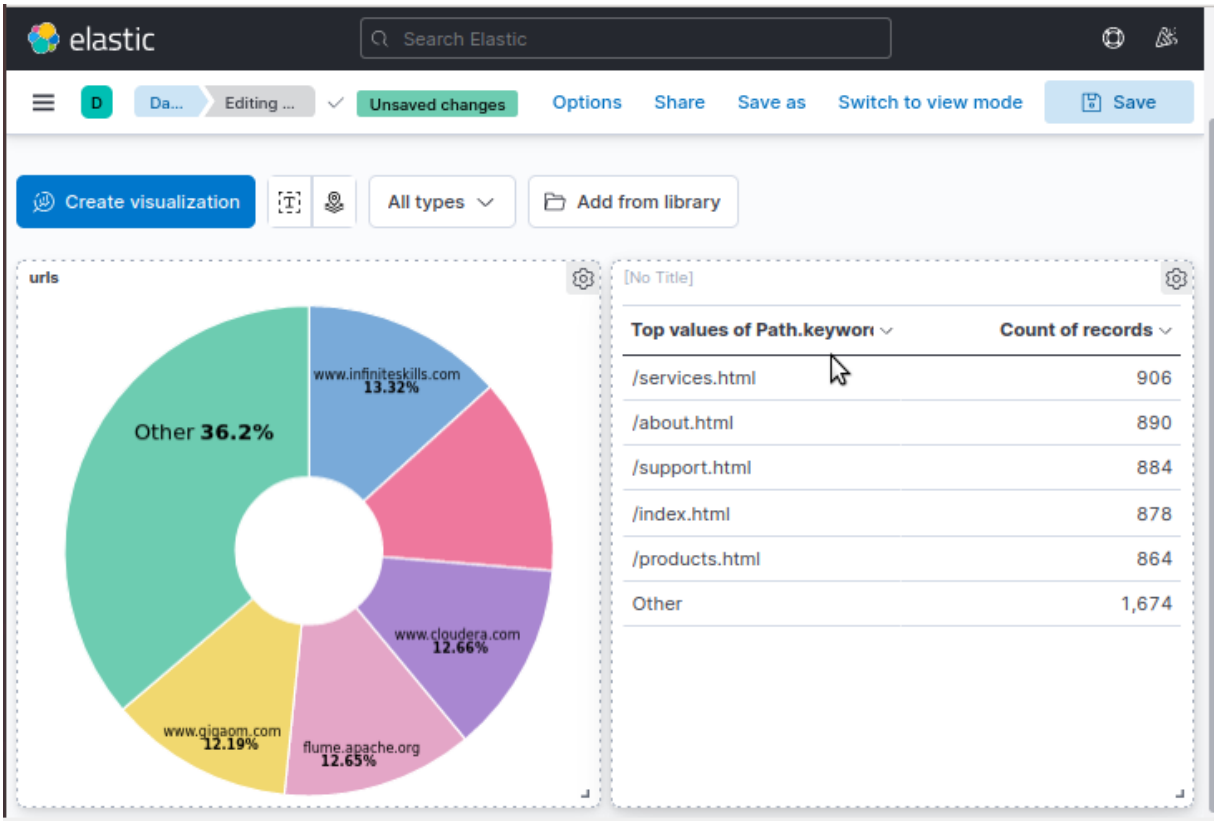





elastic
Search Elastic
D
Visualize Library
Create
Inspect
Download as CSV
Save
Pie
Other
27.46%
/services.html
14.86%
/about.html
14.6%
/support.html
14.5%
/index.html
14.4%
/products.html
14.17%


elastic
Search Elastic
D
Visualize Library
Create
Inspect
Download as CSV
Save
Table
Top values of Path.keyword
Count of records
/services.html 906
/about.html 890
/support.html 884
/index.html 878
/products.html 864
Other 1,674


elastic
Search Elastic
D
Visualize Library
Create
Inspect
Download as CSV
Save
Table
Top values of HTTP status.keyword
Count of records
303 156
503 148
504 142
401 141
204 139
Other 5,370


elastic
Search Elastic
D
Da...
Editing ...
Unsaved changes
Options
Share
Save as
Switch to view mode
Save
Create visualization
All types
Add from library
urls
www.infiniteskills.com 13.32%
Other 36.2%
www.cloudera.com 12.66%
www.gigaom.com 12.19%
flume.apache.org 12.65%
[No Title]
Top values of Path.keywor
Count of records
/services.html 906
/about.html 890
/support.html 884
/index.html 878
/products.html 864
Other 1,674