



elasticsearch



Who?

```
$ curl http://localhost:9200/talk/speaker/dpilato
{
  "nom" : "David Pilato",
  "jobs" : [
    { "boite" : "SRA Europe (SSII)", "mission" : "bon à tout faire", "date" : "1995" },
    { "boite" : "SFR", "mission" : "touche à tout", "date" : "1997" },
    { "boite" : "e-Brands / Vivendi", "mission" : "chef de projets", "date" : "2000" },
    { "boite" : "DGDDI (douane)", "mission" : "mouton à 5 pattes", "date" : "2005" },
    { "boite" : "IDEO Technologies", "mission" : "CTO", "date" : "2012" },
    { "boite" : "elastic", "mission" : "développeur", "date" : "2013" } ],
  "passions" : [ "famille", "job", "deejay" ],
  "blog" : "http://david.pilato.fr/",
  "twitter" : [ "@dadoonet", "@elasticfr", "@scrutmydocs" ],
  "email" : "david@pilato.fr"
}
```



apache



3 documents found in 78 milliseconds

[NOTICE.txt](#)

by The **Apache** Software Foundation (<http://www.apache.org/>).



[Apache.pdf](#)

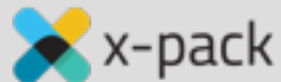
Configurer **Apache** mars 2007 | Ce dossier est public et libre au sens de la licence GNU-FDL : Permission serveur Web **Apache** [10] : installation et configuration de base, notion de Directive, création d'un répertoire serveur Web **Apache** est un logiciel Open source [6] et multiplate-forme : on peut l'installer aussi bien sur serveur **Apache** doit être installé en tant que ((service)) ; – Si vous faites cohabiter ce serveur avec serveur **Apache** doit alors écouter les requêtes HTTP [5] sur un autre port, conventionnellement le port

[LICENSE.txt](#)

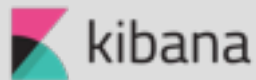
Apache License Version 2.0, January 2004 <http://www>
apply the **Apache** License to your work. To apply the **Apache** License to your work, attach the following under the **Apache** License, Version 2.0 (the "License"); you may not use this file except in compliance

The Elastic Stack

Plugins



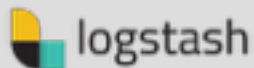
User Interface



Store, Index & Analyze



Ingest



Hosted Service



Old school search

```
SELECT
    doc.*, pays.*
FROM
    doc, pays
WHERE
    doc.pays_code = pays.code AND
    doc.date_doc > to_date('2011-12', 'yyyy-mm') AND
    doc.date_doc < to_date('2012-01', 'yyyy-mm') AND
    lower(pays.libelle) = 'france' AND
    lower(doc.commentaire) LIKE '%produit%' AND
    lower(doc.commentaire) LIKE '%david%';
```

User Interface

Power Search:

ID Number

Web Title

Url

Category

Web Description

Keywords

Contact Name

Contact Email

Featured Links 🍌

Cool Links 🍷

Bold Links

Icon

Rating Average ★★★★★

Number of Votes

Total Hits

Hits Today

IP Address

Submission Software Name

 ▼ ▼ ▼

☐ ⚠	☐ 😬	☐ 💡
☐ 📺	☐ 🔧	☐ 🏠

 ▼

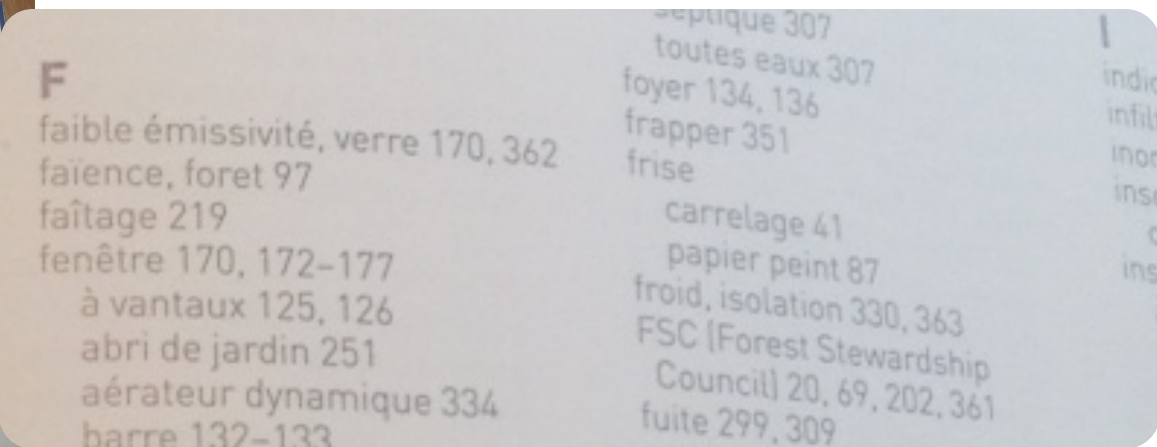
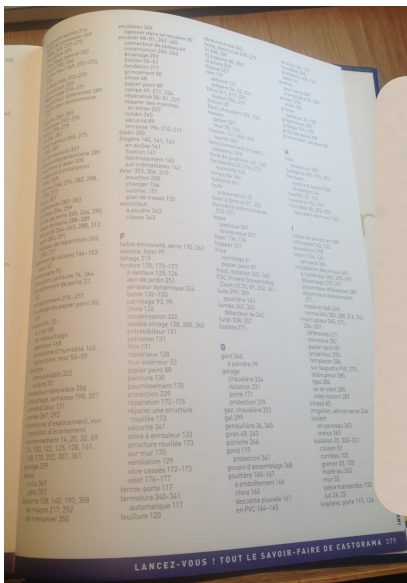
between and

between and

between and

Search engine?

- Moteur d'indexation de documents



- Moteur de recherche dans les index



Apache Lucene
HTTP / REST / JSON
Distribué, Scalable

Think document!

forget relational model

```
{
  "text": "Bienvenue au #BBL #elasticsearch",
  "created_at": "2012-04-06T20:45:36.000Z",
  "source": "Twitter for iPad",
  "truncated": false,
  "retweet_count": 0,
  "hashtag": [ { "text": "bbl", "start": 14, "end": 17 },
               { "text": "elasticsearch", "start": 19, "end": 32 } ],
  "user": { "id": 51172224, "name": "David Pilato",
            "screen_name": "dadoonet", "location": "France",
            "description": "Developer | Evangelist\r\nDeeJay 4 times a
year, just for fun !" }
}
```

Index a document

CRUD

```
$ curl -XPUT localhost:9200/talks/talk/1 -d '{
  "text": "Bienvenue au #BBL #elasticsearch",
  "created_at": "2012-04-06T20:45:36.000Z",
  "source": "Twitter for iPad",
  "truncated": false,
  "retweet_count": 0,
  "hashtag": [ { "text": "bbl", "start": 14, "end": 17 },
                { "text": "elasticsearch", "start": 19, "end": 32 } ],
  "user": { "id": 51172224, "name": "David Pilato",
            "screen_name": "dadoonet", "location": "France",
            "description": "Developer | Evangelist\r\nDeeJay 4 times a
year, just for fun !" }
}'
```

Search for documents

The unstructured way

```
$ curl localhost:9200/talks/talk/_search?q=elasticsearch
{
  "took" : 5, "timed_out" : false,
  "_shards" : { "total" : 5, "successful" : 5, "failed" : 0 },
  "hits" : {
    "total" : 1,
    "max_score" : 0.06780553,
    "hits" : [ {
      "_index" : "talks",
      "_type" : "talk",
      "_id" : "1",
      "_score" : 0.06780553,
      "_source" : {
        "text" : "Bienvenue au #BBL #elasticsearch",
        "created_at" : "2012-04-06T20:45:36.000Z", [...]
```

Search for documents

The structured way

```
$ curl localhost:9200/talks/talk/_search -d '{
  "query": {
    "bool": {
      "filter": {
        "term": { "user.name": "david" }
      },
      "must_not": {
        "range": { "hashtag.start": { "gte": 0, "lte": 10 } }
      },
      "should": [ {
        "match": { "user.location": "france" }
      }, {
        "match": { "text": "elasticsearch bienvenue" }
      }
    ]
  }
}
```



beats



logstash

Aggregations

Make sense of your data!
(in near real time)



Results (15,905)

25

50

75

100

Sort by

Relevance ▾



Search tools ▾

1

2

3

4

5

of

160



File type ▾

Orientation ▾

Size ▾

Collection ▾

Price ▾

Recency ▾

People ▾

Options ▾

Color ▾

Category ▾



✓ All

≤ 2 years (~ 7,700)

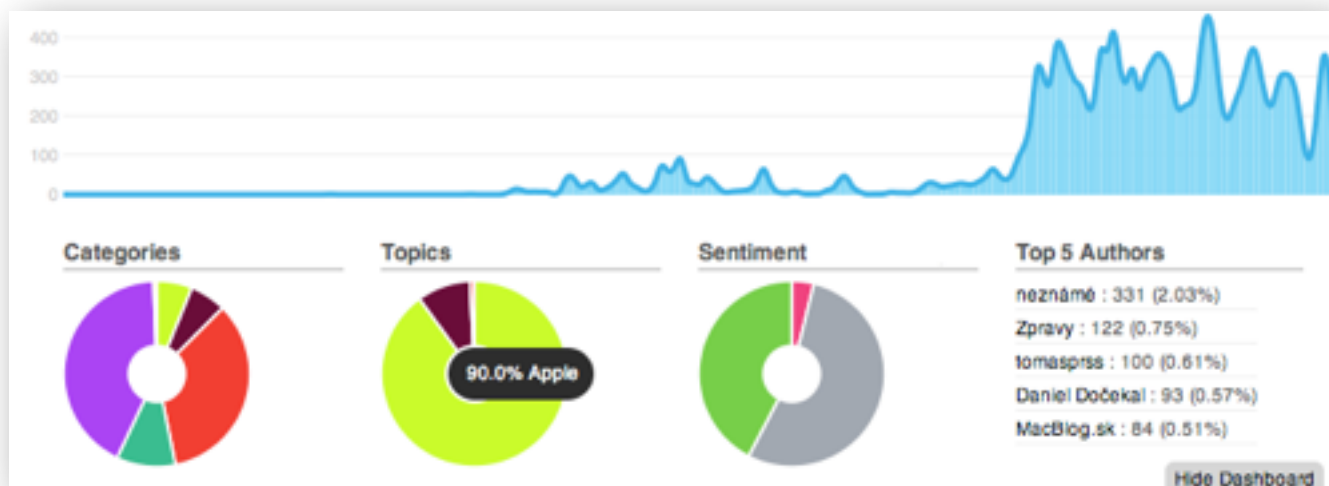
≤ 1 year (~ 4,500)

≤ 6 months (~ 2,100)

≤ 1 month (~ 360)

≤ 1 week (~ 20)





The one million tweet map

powered by **maptimize**

Tweets since page load

- - - 5 5 2

Legend



Filters

cluster view ☒ heatmap view

Q keywords filter

hashtags filter

5 most popular hashtags

job (23355) hiring (23417)

jobs (12371) careers (11896)

ica (3000)

No Time Filtering

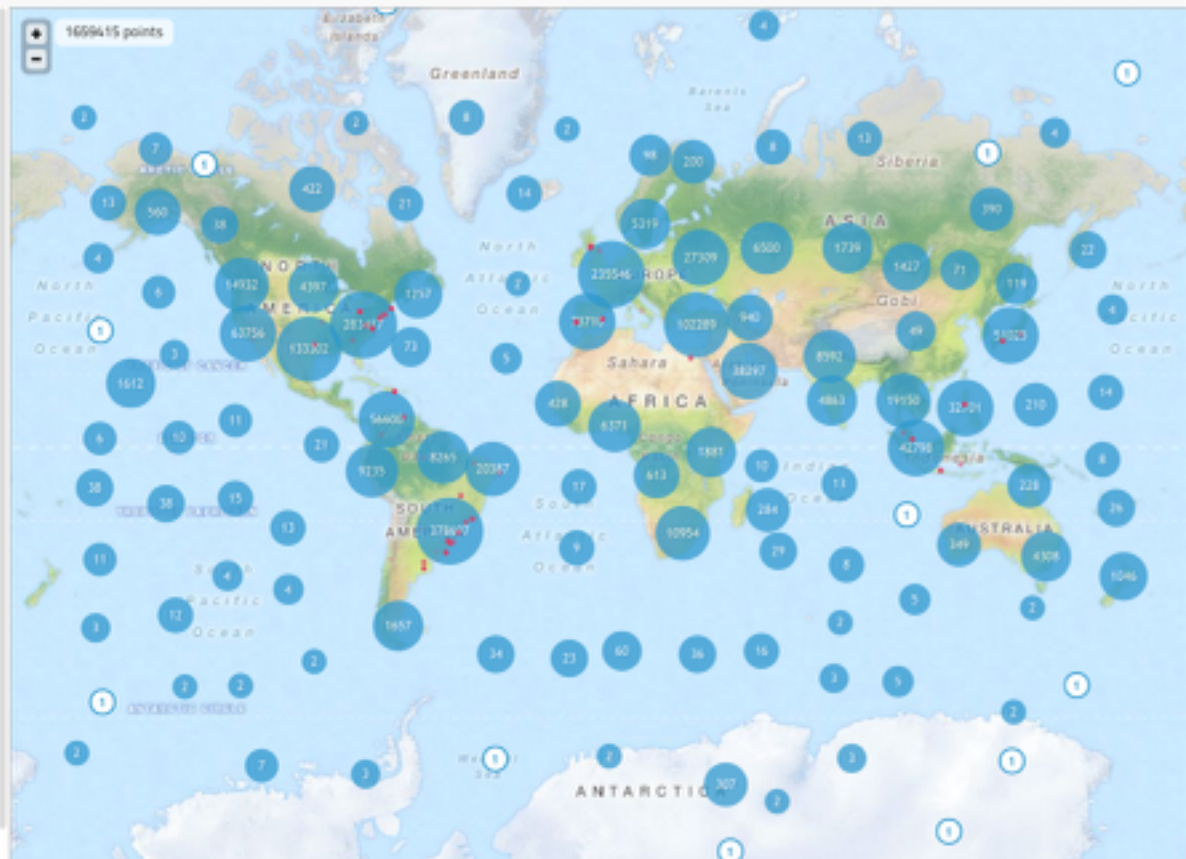
If you want to start with a blank map, click on reset map button below the page to see all tweets again.

Reset map

Maptimize

A geographical clustering engine for online maps to display and analyse big geolocalized data.

Get more info



Power Search:

ID Number

Web Title

Url

Category

Web Description

Keywords

Contact Name

Contact Email

Featured Links 🖱️

Cool Links 🚗

Bold Links

Icon

Rating Average ★★★★★

Number of Votes

Total Hits

Hits Today

IP Address

Submission Software Name

Select

Select ▼

Select ▼

Select ▼



Select ▼

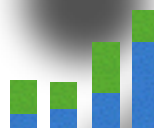
between and

between and

between and



Submission Software Name



La description de ce résultat n'est pas accessible à cause du fichier robots.txt de ce site. En savoir plus



elasticsearch



kibana



cloud

Demo time!

Percolation

aka inverted search

Record searches

```
$ curl -XPOST localhost:9200/twitter/.percolator/dadoonet -d '{
  "query" : { "term" : { "user.screen_name" : "dadoonet" } }
}'

$ curl -XPOST localhost:9200/twitter/.percolator/elasticsearch -d '{
  "query" : { "match" : { "hashtag.text" : "elasticsearch" } }
}'

$ curl -XPOST localhost:9200/twitter/.percolator/mycomplexquery -d '{
  "query": {
    "bool": {
      "filter": {
        "term": { "user.name": "david" }
      },
      "must_not": {
        "range": { "hashtag.start": { "gte": 0, "lte": 10 } }
      }
    }
  }
}'
```

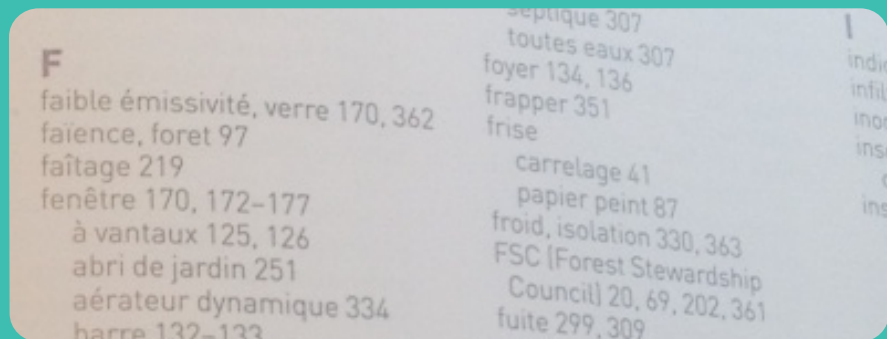
Percolate a document

```
$ curl localhost:9200/twitter/tweet/_percolate -d '{
  "doc": {
    "text": "Bienvenue à la conférence #elasticsearch pour #JUG",
    "created_at": "2012-04-06T20:45:36.000Z",
    "hashtag": [ { "text": "elasticsearch", "start": 27, "end": 40 } ],
    "user": { "screen_name": "dadoonet" } } }'

{
  "took" : 19,
  "_shards" : { "total" : 5, "successful" : 5, "failed" : 0 },
  "total" : 2,
  "matches" : [ { "_index" : "twitter", "_id" : "dadoonet" },
                 { "_index" : "twitter", "_id" : "elasticsearch" }
] }
```

Analysis & Mapping

Should we index everything?



Analysis

Standard Analyzer

```
$ curl -XPOST 'localhost:9200/test/_analyze?analyzer=standard&pretty=1' -d 'The quick brown fox jumps over the lazy Dog'
```

```
{
  "tokens" : [ {
    "token" : "quick",
    "start_offset": 4, "end_offset": 9, "type": "<ALPHANUM>", "position": 2
  }, {
    "token" : "brown",
    "start_offset": 10, "end_offset": 15, "type": "<ALPHANUM>", "position": 3
  }, {
    "token" : "fox",
    "start_offset": 16, "end_offset": 19, "type": "<ALPHANUM>", "position": 4
  }, {
    "token": "jumps",
    "start_offset": 20, "end_offset": 26, "type": "<ALPHANUM>", "position": 5
  }, {
    "token": "over",
    "start_offset": 27, "end_offset": 31, "type": "<ALPHANUM>", "position": 6
  }, {
    "token" : "lazy",
    "start_offset": 36, "end_offset": 40, "type": "<ALPHANUM>", "position": 8
  }, {
    "token" : "dog",
    "start_offset": 41, "end_offset": 44, "type": "<ALPHANUM>", "position": 9
  } ] }
```

Analysis

Whitespace Analyzer

```
$ curl -XPOST 'localhost:9200/test/_analyze?analyzer=whitespace&pretty=1' -d 'The quick brown fox jumps over the lazy Dog'
```

```
{
  "tokens" : [ {
    "token" : "The", ...
  }, {
    "token" : "quick", ...
  }, {
    "token" : "brown", ...
  }, {
    "token" : "fox", ...
  }, {
    "token" : "jumps", ...
  }, {
    "token" : "over", ...
  }, {
    "token" : "the", ...
  }, {
    "token" : "lazy", ...
  }, {
    "token" : "Dog", ...
  } ] }
```

Analyzer?

- whitespace

"the dog!" -> "the", "dog!"

- standard

"the dog!" -> "the", "dog"

- asciifolding

éléphant -> elephant

- stemmer french

elephants -> "eleph"

prenez -> "prendre"

- stopword french (le, la, un, une, être, avoir, ...)
- ngram ou edge ngram

eleph -> ["el", "ele", "elep", "eleph"]

Register your analyzer

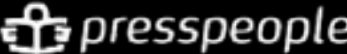
```
"analysis":{
  "analyzer":{
    "francais":{
      "type":"custom",
      "tokenizer":"standard",
      "filter":["lowercase", "stop_francais", "fr_stemmer", "asciifolding", "elision"]
    }
  },
  "filter":{
    "stop_francais":{
      "type":"stop",
      "stopwords":["_french_", "twitter"]
    },
    "fr_stemmer" : {
      "type" : "stemmer",
      "name" : "french"
    },
    "elision" : {
      "type" : "elision",
      "articles" : ["l", "m", "t", "qu", "n", "s", "j", "d", "lorsqu"]
    }
  }
}
```

Define your mapping

```
"tweet" : {
  "properties": {
    "description": { "type": "string", "analyzer": "français" },
    "username": { "type": "string", "analyzer": "ngram", "search_analyzer": "simple" },
    "city": {
      "type": "string",
      "analyzer": "français",
      "fields": {
        "ngram": {
          "type": "string",
          "analyzer": "ngram"
        },
        "raw": {
          "type": "string",
          "index": "not_analyzed"
        }
      }
    }
  }
}
```

Users & Community





francetvinfo



AramisAuto.com
Le Prix Le Choix Les Marques

Crédit Mutuel
ARKEA



bluekiwi
enterprise social software



SOCIETE
GENERALE



SeLogger
.com



viadeo





elasticfr



@elasticfr



elastic

discuss.elastic.co

