



Drupagora

ELASTICSEARCH & DRUPAL

Retour d'expérience

David Pilato

Technical advocate



elasticsearch.

@dadoonet

Florian Le Moing

Technical project manager

Alter Way

@gastons78

Alter Way







elasticsearch.

elasticsearch



kibana



logstash



Marvel





elasticsearch.

Training (public and on-site)





elasticsearch.

Training (public and on-site)

Development support

Production support

we've got your back, 24/7

Accelerate your return on investment while mitigating your risk



24/7 response time available

Meet the SLAs in your production environment



Support from the core team that built Elasticsearch, Kibana, and Logstash

Get access to the people who wrote the code



Expert-level solutions

Use best practices in your mission-critical deployments



Hotfixes and emergency patches

Get immediate solutions to your issues

RECHERCHE DRUPAL

Recherche drupal native

- En base de données
- Lent
- Limité fonctionnellement

Recherche avec moteur open source

Lucene



elasticsearch.

Contexte

- Retour d'expérience sur notre utilisation dans un projet Drupal 7
- Projet corporate avec une grosse problématique de contenu
- Des catégorisations dans tous les sens
- Elasticsearch = ! Apache Solr => pas de bonne solution, dépend du contexte

ELASTICSEARCH



search = like % ?

```
SELECT
  doc.*, country.*
FROM
  doc, country
WHERE
  doc.country_code = country.code AND
  doc.date_doc > to_date('2011-12', 'yyyy-mm') AND
  doc.date_doc < to_date('2012-01', 'yyyy-mm') AND
  lower(country.name) = 'france' AND
  lower(doc.comment) LIKE '%product%' AND
  lower(doc.comment) LIKE '%david%';
```



Search engine ?

1964

Alphabetical **INDEX**

A	Accessory Groups.....	63
	Additional Accessories.....	62
	Adjustable Steering Wheel.....	54
	Air Cleaner.....	95
	Water-Heater System.....	52, 53
	Brake Control.....	71, 95
	Clamp Timer.....	51
	ing System.....	88, 89
		91, 101

B	Back-Up Lights.....	9
	Battery.....	70, 97
	Body Construction.....	82A
	Body Insulation.....	87
	Body Sealing.....	87
	Body Styles.....	26-49
	Brake Drums.....	76, 77, 102
	Brake Lining.....	77, 102
	Brakes.....	76, 77, 102
	Brakes, Parking.....	76, 77, 102
	Brakes, Power.....	76, 77, 102
	Bucket Seat Console.....	25
	Bucket Seats.....	35-39, 45
	Bumpers.....	9

		69, 94
		71, 95
		72-81
	View.....	72A
	partment.....	48, 49
		71, 95
	amber.....	68
		52, 53
	Ratio.....	91, 92
	Connecting Rods.....	93
	Console, Bucket Seat.....	25
	Constant Velocity (CV) Joints.....	74, 75, 100
	Controlled Differential.....	58
	Cooling.....	95, 96



plug & play

scalable

elasticsearch ?

REST/JSON

Lucene

plug & play

scalable

elasticsearch

Apache 2 license

REST/JSON

Lucene

Start...

```
$ wget https://download.elasticsearch.org/elasticsearch/  
elasticsearch/elasticsearch-1.4.0.tar.gz  
$ tar -xf elasticsearch-1.4.0.tar.gz  
$ ./elasticsearch-1.4.0/bin/elasticsearch  
[INFO ][node           ][Ghost Maker] {1.4.0}[5645]: initializing
```



... and play!

```
$ curl -XPUT localhost:9200/sessions/session/1 -d '{
  "title"      : "Elasticsearch + Drupal",
  "subtitle"   : "Retour d'experience",
  "date"       : "2014-11-14T14:00:00",
  "tags"       : [ "elasticsearch", "drupal", "drupagora" ],
  "speakers"  : [{
    "name"     : "Florian Le Moing"
  }, {
    "name"     : "David Pilato"
  }]
}'
```

Search!

```
$ curl http://localhost:9200/sessions/session/_search -d'
{
  "query": {
    "multi_match": {
      "query": "elasticsearch drupagora florian",
      "fields": [ "title^3", "tags^2", "speakers.name" ]
    }
  },
  "post_filter": {
    "range": {
      "date": {
        "from": "2014-11-01",
        "to": "2014-12-01"
      }
    }
  }
}
```

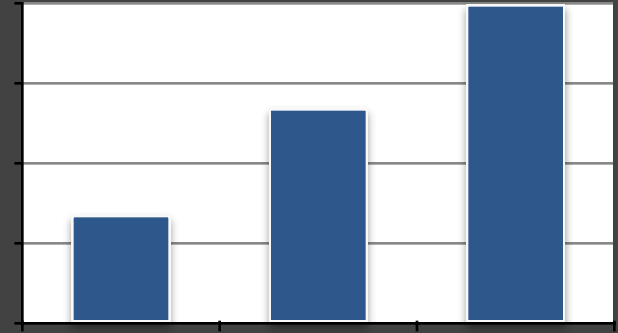
Compute?

<i>B</i>	Back-Up Lights.....	9
	Battery.....	70, 97
	Body Construction.....	82A
	Body Insulation.....	87
	Body Sealing.....	87
	Body Styles.....	26-49
	Brake Drums.....	76, 77, 102
	Brake Lining.....	77, 102
	Brakes.....	76, 77, 102
	Brakes, Parking.....	76, 77, 102
	Brakes, Power.....	76, 77, 102
	Bucket Seat Console.....	25
	Bucket Seats.....	35-39, 45
	Bumpers.....	9

Compute!

```
$ curl http://localhost:9200/sessions/session/_search -d'
```

```
{  
  "query": { ... },  
  "aggs": {  
    "by_date": {  
      "date_histogram": {  
        "field": "date",  
        "interval": "day",  
        "format": "dd/MM/yyyy"  
      }  
    }  
  }  
}'
```



```
"by_date": [  
  { "key_as_string": "03/04/2014", "doc_count": 1 },  
  { "key_as_string": "12/04/2014", "doc_count": 2 },  
  { "key_as_string": "16/04/2014", "doc_count": 3 }  
]
```



Let's make sense of ...

- logs
- twitter
- github
- marketing data
- CMS content
- ...
- your data
- your big data

Let's make sense of ...

- logs
- twitter
- github
- marketing data
- CMS content
- ..
- your data
- your big data

```
{  
  "name": "Pilato David",  
  "dateOfBirth": "1971-12-26",  
  "gender": "male",  
  "children": 3,  
  "marketing": {  
    "fashion": 334,  
    "music": 3363,  
    "hifi": 2351  
  },  
  "address": {  
    "country": "France",  
    "city": "Paris",  
    "location": [2.332395, 48.861871]  
  }  
}
```

démo

MAKE SENSE OF YOUR (BIG) DATA!

let's inject some marketing documents...

ELASTICSEARCH

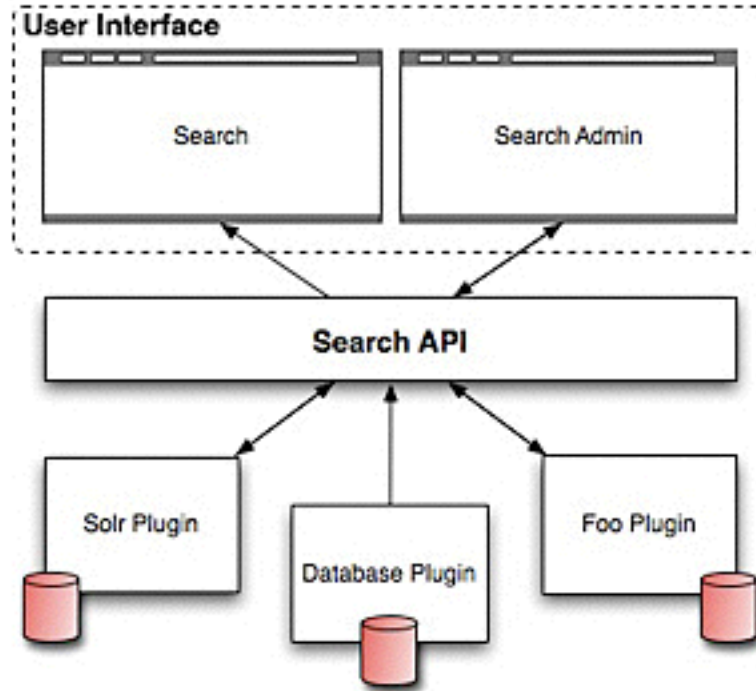


Choix d'elasticsearch

- Recherche avec facettes
- Indexation multi-site
- Une seule et même « stack » côté hébergement et côté production
- Rapports sur l'utilisation du contenu
- Indexation des documents

ARCHITECTURE DRUPAL

Search API



https://www.drupal.org/project/search_api

Architecture Drupal : Modules

- Beaucoup de modules ! Peu sont actualisés et fonctionnent avec la dernière version d'ES 1.4.0
- 2 modules principaux mis à jour :
 - https://drupal.org/project/search_api_elasticsearch (utilisé sur 319 sites)
 - https://drupal.org/project/elasticsearch_connector (utilisé par 189 sites)

Module : search_api_elasticsearch

- <https://github.com/ruflin/Elastica>
- <https://austin2014.drupal.org/session/using-elasticsearch-drupal-7-search>
- https://github.com/VeggieMeat/search_api_elasticsearch/tree/7.x-1.x-multiple-client-support
- On contribue (un peu) au module

AGRÉGATIONS / FACETTES

Module : facet_api

FILTRE PAR THÉMATIQUE :

- ☒ Domaines d'expertise (148)
- ☒ Missions (104)
- ☒ Marchés (11)
 - ☐ Communication (5)
- ☒ Formation (7)
 - ☐ Profession comptable (5)
- ☒ Exercice professionnel (17)
- ☒ L'Ordre (50)

RECHERCHE SIMPLE

Il y a 346 résultats

Modifier ma recherche : **Envoyer**

Article Mise à jour : 27/08/2014

L'Ordre > Régions

Les Conseils Régionaux

Les Conseils régionaux représentent l'Ordre dans chaque circonscription. Il existe vingt-trois Conseils régionaux en métropole, en Corse, Martinique, Guadeloupe et la Réunion, ainsi qu'un Comité...

Article Mise à jour : 01/09/2014

L'Ordre > Régions

Alsace

https://www.drupal.org/project/facet_api

Module : facetapi_bonus

FILTRE PAR THÉMATIQUE :

- ☐ Domaines d'expertise (148)
- ☒ Social (55)
 - ☐ Aides à l'emploi (4)
 - ☐ Chômage (4)
 - ☐ Cotisations et contributions sociales (7)
 - ☐ Rémunération (7)
- ☒ Représentants du personnel (8)
- ☒ Rupture du contrat de travail (12)
 - ☐ Aides à l'emploi (4)
- ☒ Fiscal (47)
- ☒ Juridique (25)
 - ☐ Comptable (19)
- ☒ Missions (104)
- ☒ Marchés (11)
 - ☐ Communication (5)
- ☒ Formation (7)
 - ☐ Profession comptable (5)
- ☒ Exercice professionnel (17)
- ☒ L'Ordre (50)

RECHERCHE SIMPLE

Il y a 346 résultats

Modifier ma recherche :

Envoyer

Article Mise à jour : 27/06/2014

L'Ordre > Regions

Les Conseils Régionaux

Les Conseils régionaux représentent l'Ordre dans chaque circonscription. Il existe vingt-trois Conseils régionaux en métropole, en Corse, Martinique, Guadeloupe et la Réunion, ainsi qu'un Comité...

Article Mise à jour : 01/09/2014

L'Ordre > Regions

Alsace

Le Conseil Régional CROEC Alsace Maison de la Comptabilité et de la Gestion...

Article Mise à jour : 01/09/2014

L'Ordre > Regions

Alsace

https://www.drupal.org/project/facetapi_bonus

MULTI INDEX

Multi index : exemple

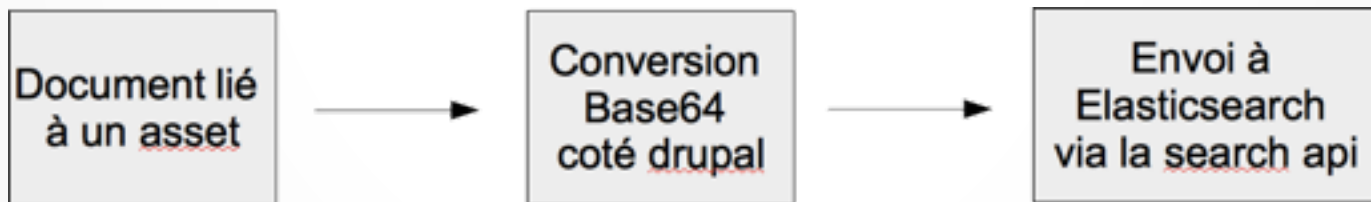
- Un index par besoin :
 - contenu de l'ordre => article/actualité/article pnp/événement/...
 - log => drupal et system et statistiques de recherche
 - document => index pour la recherche de document (mémorialiste) avec des filtres spécifiques
 - Kibana => index pour la gestion des dashboard dans Kibana
 - regions => un index par région
 - annuaire
 - site externe
 - ...

https://www.drupal.org/project/search_api_multi

INDEXATION DOCUMENTS

Indexation document

- Conversion base 64
- Envoi à elasticsearch
- Indexation des documents par elasticsearch via le plugin elasticsearch-mapper-attachments

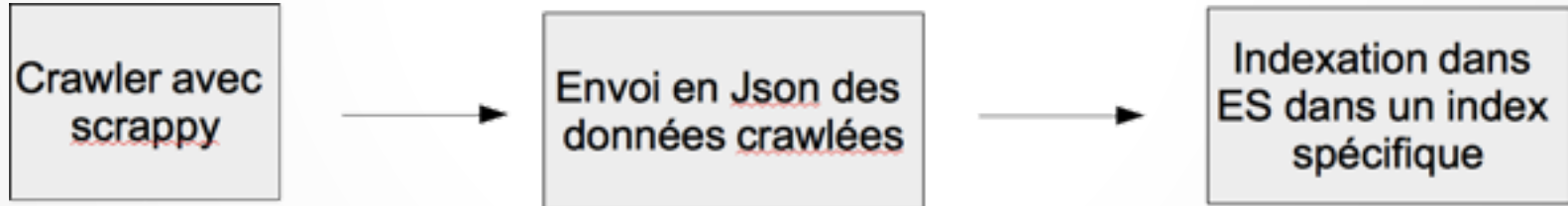


<https://github.com/elasticsearch/elasticsearch-mapper-attachments>

INDEXATION SITE EXTERNE

Crawling de site externe

- Scrappy
- cron hebdomadaire qui crawle les différents sites souhaités



<http://scrapy.org>

Type de document

- ☐ Checkbox 01
- ☐ Checkbox 02
- ☐ Checkbox 03

Thème

- ☒ Social
- ☐ Congés
- ☒ Fiscal
- ☐ Congés
- ☐ Aides
- ☒ Absences
- ☒ Maladie, maternité, accident du travail, maladie professionnelle
- ☐ Comptable
- ☐ Juridique
- ☐ Autre

Période

- > 2014
- < 2013
- janvier
 - février
- > 2012

☐ Rechercher sur l'ensemble des sites de l'ordre

Valider

Il y a 127 résultats pour la recherche "comptables"

Modifier ma recherche :

comptables

Envoyer

Billet Mise à jour: 26/04/2014

Fiscal > Absences > Maladie, maternité, accident du travail, maladie professionnelle

Tout savoir sur le mécénat culturel : aspects comptables, fiscaux et juridiques

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris volutpat commodo. Nunc viverra libero eget ante raesent iaculis euismod purus.

Analyse **L'Académie** Mise à jour: 26/04/2014

Social > Congés

69e Congrès, le monde change : oser, agir, conquérir

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo.

Pratique **Club développement durable** Mise à jour: 26/04/2014

Fiscal > Absences > Maladie, maternité, accident du travail, maladie professionnelle

Métier d'expert-comptable : image, perceptions, attentes

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris volutpat commodo. Nunc viverra libero eget ante

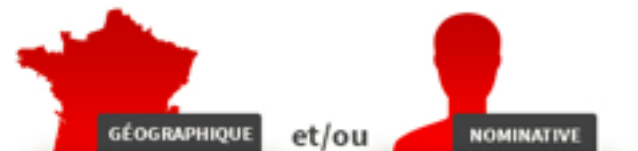
recherche multi index

Billet Mise à jour: 26/04/2014

Fiscal > Absences > Maladie, maternité, accident du travail, maladie professionnelle

CARTE + ELASTICSEARCH

Annuaire / Recherche



Code postal ou n° de département :

Nom de personne ou de cabinet :

Nom de ville ou de département :

Note : peut être combinée avec une précision géographique à gauche

Note : pour les recherches en France, métropole et outre-mer, si il n'y a pas de résultat dans la localité précisée, les plus proches géographiquement pourront être proposées.

Qualités :

- ☐ Experts comptables et cabinets d'expertise comptable
- ☐ Stagiaires
- ☐ AGC et salariés d'AGC 83T et 83Q
- ☐ Experts-comptable honoraires
- ☐ Sociétés de participations financières de professions libérales

Personnes :

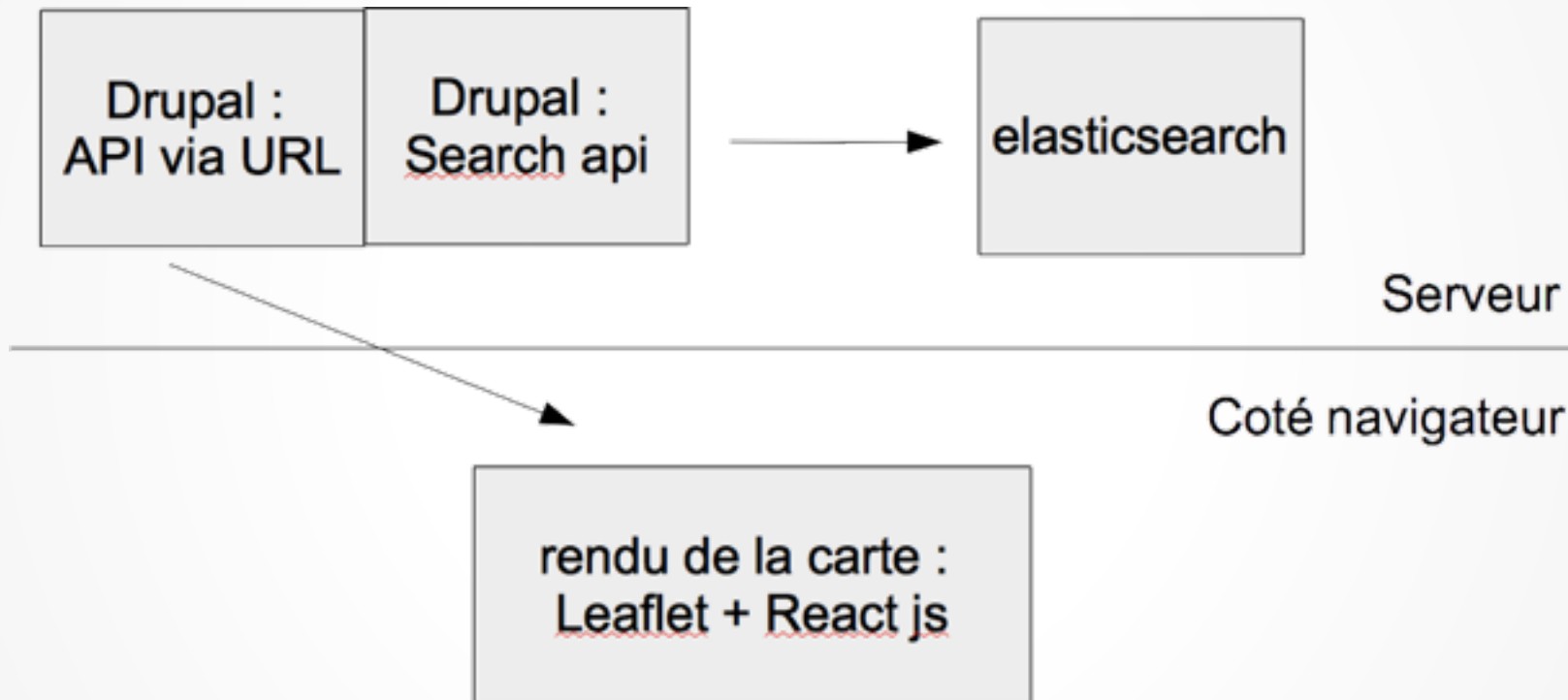
- ☐ Personnes physiques
- ☐ Personnes morales

Annuler

Rechercher



Architecture

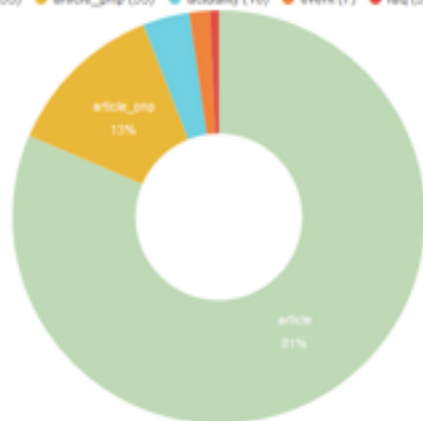


KIBANA

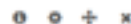
REPARTITION TYPE DE CONTENU



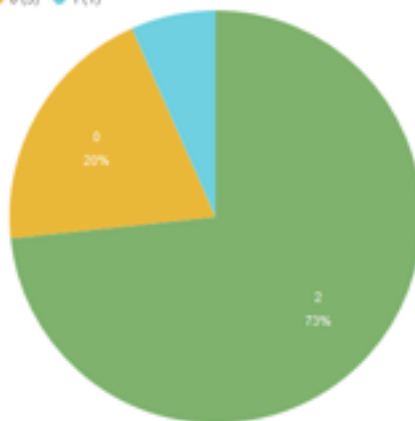
● article (355) ● article_gnp (55) ● actuality (16) ● event (7) ● faq (3)



TYPE D'ACTUALITE



● 2 (11) ● 0 (3) ● 1 (1)



CATÉGORIE FAQ



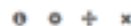
Term	Count	Action
385	3	Q @

NOMBRE DE CONTENU



● (436)

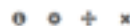
REGION DES EVENEMENTS



● 384 (5) ● 2025 (1) ● 1744 (1)



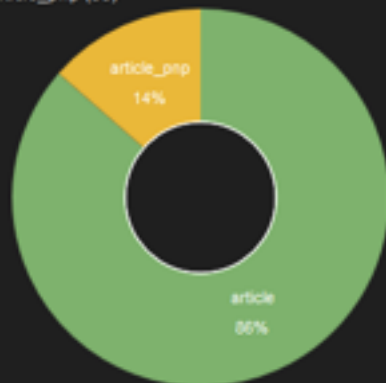
TYPE D'EVENEMENTS



Term	Count	Action
2	7	Q @

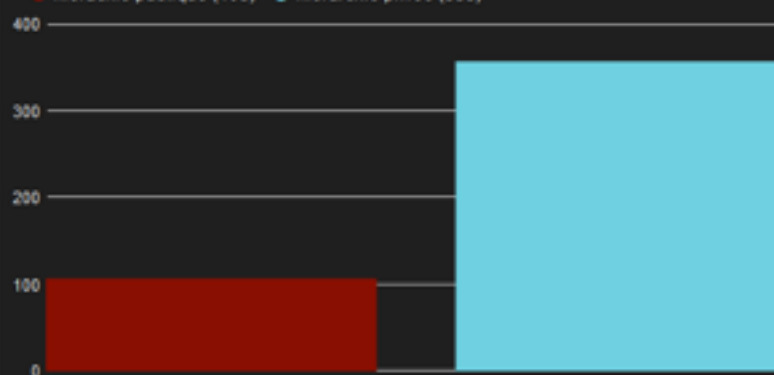
REPARTITION PAR TYPE DE CONTENU AYANT UNE HIERARCHIE

● article (350) ● article_pnp (55)



NOMBRE DE CONTENU DANS CHAQUE HIERARCHIE

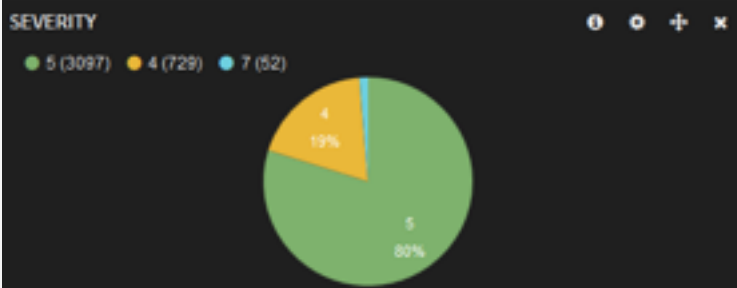
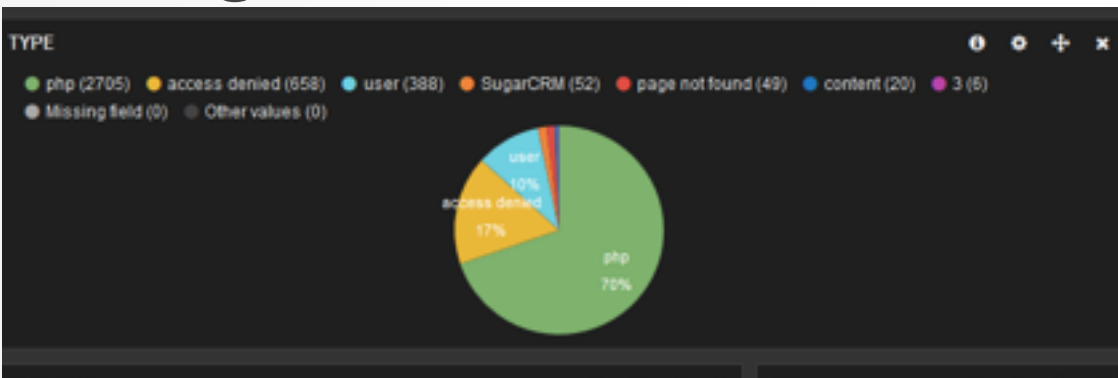
● hierarchie publique (105) ● hierarchie privée (355)



TYPE D'ARTICLE

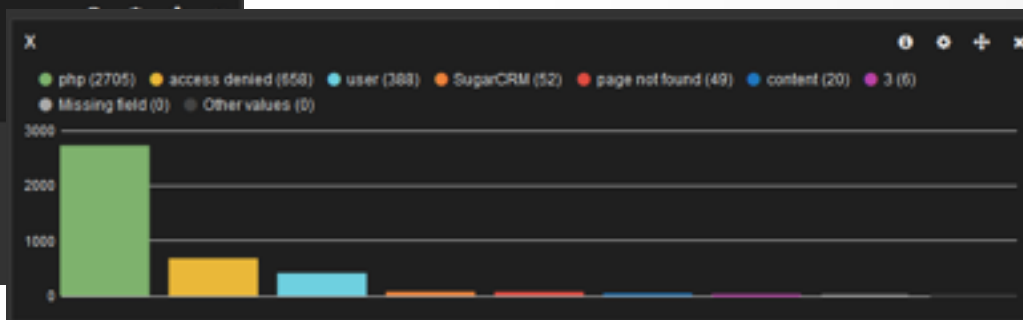
Term	Count	Action
article	350	🔍 🗑️
article_pnp	55	🔍 🗑️
Missing field	0	🔍 🗑️
Other values	0	

Logs



404

49



HITS

3878

+ / -

+ / -

- Moins de module que sur SOLR (mais en croissance rapide)
- Plus facile de se connecter avec des outils externes (envoi json)
- Plus de réactivité coté hébergement car même outil
- Dashboard de contenu très facile à réaliser (à faire avec son client en direct)
- Module supplémentaire que nous avons réalisé :
 - Statistiques de recherche (log spécifique indexé et analysé via kibana)
 - Module Base64 pour indexation documents

Performance

- Oui mais pas besoin de performance très poussé sur ce projet
- Surcharge drupal via la search-api en terme de performance
- Elasticsearch c'est pour le small data aussi

DRUPAL 8 ?

Drupal 8 ?

- Attente de la stabilisation de Search API
 - <https://www.drupal.org/sandbox/daeron/2091893>
- Les modules search_api_elasticsearch et elasticsearch_connector ont pour ambition d'être prêts pour drupal 8
 - <https://www.drupal.org/node/2274757>





thanks

@dadoonet / @gastons78