



APACHE SOLR



Raphaël MANGER – INF05

DESCRIPTION GÉNÉRALE

Serveur/Moteur de recherche

Développé en Java

Maintenu par la fondation Apache

Nombreuses fonctionnalités :

- Recherche de texte
- Indexation en temps réel
- Clustering
- Intégration à une DB
- ...

POUR LA PETITE HISTOIRE

2004 : CNET Networks

→ société de média et d'information

→ besoins internes de l'entreprise

→ Yonik Seeley

2006 : Apache Software Foundation

→ Open Source

2009 : LucidWorks

→ 1ère entreprise de support commercial

APACHE

LUCENE

Bibliothèque Java qui permet d'indexer et de rechercher du texte

Sortie : 2000

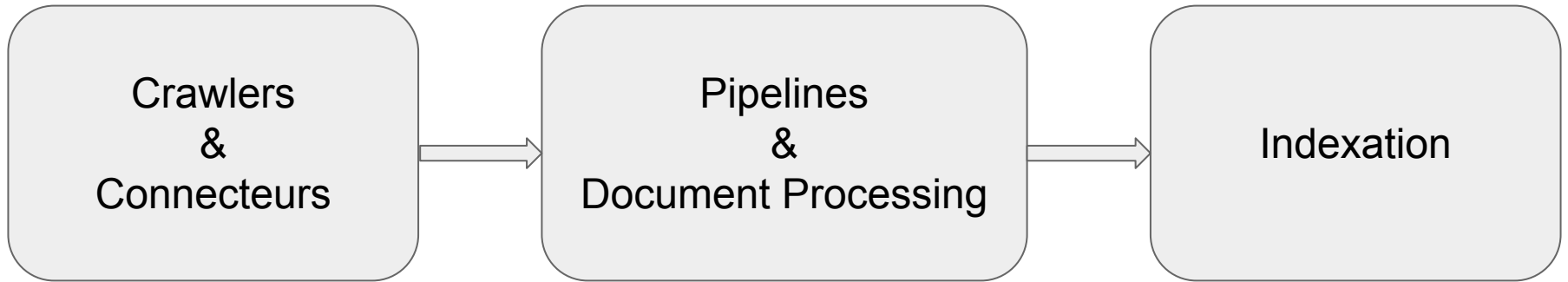
Dérivés : Solr, Hadoop, Tika, Nutch, ...

SOLR

Serveur de recherche qui utilise les fonctionnalités de Lucene

Sortie : 2004/2006

"ACQUISITION DE DONNÉES"



- Nutch
- Aperture
- ...

- Aspire
- Tesseract
- ...

- Lucene
- Tika
- ...

FONCTIONNALITÉS PRINCIPALES

Interface d'administration en HTML et/ou ligne de commandes

API REST-like : HTTP, JSON

Support de données :

- structurées (.csv .json .xml .html ...)
- libres (.txt .doc .pdf .ppt ...)

Recherche à facettes / Analyse de pertinence

Possibilité d'extension par plugins

COMMANDES ESSENTIELLES

```
bin/solr start [-p ...]
```

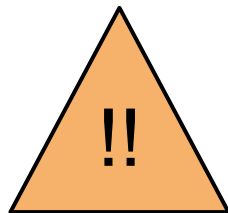
```
bin/solr stop [-p ...]
```

```
bin/solr status
```

```
bin/solr create -c <name>
```

```
bin/post -c <name> [ http://... | ~/... ]
```

```
bin/solr delete -c <name>
```



REQUÊTES ESSENTIELLES

`http://<adresse>:<port>/<server>/<collection>/...`

`http://localhost:8983/solr/Genius/`

→ Rechercher des données : `select`

→ Ajouter des données : `update`

COMPARAISON AVEC ELASTICSEARCH

API

Feature	Solr 7.2.1	Elasticsearch 6.2.4
Format	XML, CSV, JSON	JSON
HTTP REST API	✓	✓
Binary API 🔗	✓ SolrJ	✓ TransportClient, Thrift (through a plugin)
JMX support	✓	✗ ES specific stats are exposed through the REST API
Official client libraries 🔗	Java	Java, Groovy, PHP, Ruby, Perl, Python, .NET, Javascript Official list of clients
Community client libraries 🔗	PHP, Ruby, Perl, Scala, Python, .NET, Javascript, Go, Erlang, Clojure	Clojure, Cold Fusion, Erlang, Go, Groovy, Haskell, Java, JavaScript, .NET, OCaml, Perl, PHP, Python, R, Ruby, Scala, Smalltalk, Vert.x Complete list
3rd-party product integration (open-source) 🔗	Drupal, Magento, Django, ColdFusion, Wordpress, OpenCMS, Plone, Typo3, ez Publish, Symfony2, Riak (via Yokozuna)	Drupal, Django, Symfony2, Wordpress, CouchBase
3rd-party product integration (commercial) 🔗	DataStax Enterprise Search, Cloudera Search, Hortonworks Data Platform, MapR	SearchBlox, Hortonworks Data Platform, MapR etc Complete list
Output 🔗	JSON, XML, PHP, Python, Ruby, CSV, Velocity, XSLT, native Java	JSON, XML/HTML (via plugin)

Liste complète sur www.solr-vs-elasticsearch.com

POINTS FORTS

Installation **rapide** et **facile** (.zip)

Prise en main **rapide** et **facile** :

- Documentation claire
- Exemples déjà fournis
- Interface graphique HTML

Indexation **rapide** et **facile** (pour les débutants) :

- Script déjà fourni
- Fichiers de configuration par défaut

POINTS FAIBLES

(...)

- Interface graphique HTML incomplète
- Scripts et configurations par défaut peu fiables
 - Web Crawler, Lecteur PDF, Encodage

CONCLUSION

MERCI POUR VOTRE
ATTENTION

LA DÉMO !