

Developer's Intro to Apache Solr

Peter Wolanin, Ph.D.

`pwolanin`

<http://drupal.org/user/49851>

- NYC meetup, April 3, 2013

Overview: Get a Local Install!

- < 5 minute install
- Basic Understanding
- Monitoring
- Keeping it secure

Solr Install in < 5 min

- <http://drupal.org/apachesolr/quick-start-solr-3>
- <http://nickveenhof.be/blog/simple-guide-install-apache-solr-3x-drupal-7>
- Make sure you have java 6 (or java 7 latest) installed.
- Hint: almost everything I know came from: <http://wiki.apache.org/solr/>

Quick start for Solr 3.6

*Last updated April 3, 2013. Created by [pwolanin](#) on December 1, 2011.
Edited by [davidarthur](#). [Log in to edit this page](#).*

The attached .zip contains only the essential files needed to run Solr 3.6.2 locally for development and testing with the Apache Solr Search Integration module.

It is a 17 MB download instead of the approx 83 MB download of the full release. Download and unzip it in any convenient location This saves you the first several steps.

the solr/conf directory has the Drupal module configuration files from the apachesolr 7.x-1.1 module (schema name ="drupal-3.0-0-solr3").

the multicore directory d7-core has the same conf files, while the d6-core directory has the apachesolr 6.x-1.7 release conf (schema name="drupal-1.7")

To start solr with a single core for use with Drupal 7.x-1.x, just run:

```
java -jar start.jar
```

To run multicore with the 2 configured cores run:

```
java -Dsolr.solr.home=multicore -jar start.jar
```

In either case, then navigate to: <http://localhost:8983/solr/>

For more details see: <http://www.nickveenhof.be/blog/simple-guide-install-apache-solr-3x-drupal-7>

Attachment	Size
apache-solr-3.6.2-quickstart.zip	17.18 MB

Getting it running

```
java -jar start.jar
```

```
java -Dsolr.solr.home=multicore \  
-jar start.jar
```

Caveats:

- ☹ No HA
- ☹ No restart on reboot
- ☹ No security



Solr Interface/API is HTTP

- Drupal sends data to Solr as XML documents
- POST XML to /update to add or delete.
- Search via GET requests.
- If something is not working as expected, you can try searching directly in Solr via URL
- Solr also includes admin and analysis interfaces (you need to lock this down for production).

Solr Admin (drupal-3.0-0-solr3)



twvna-ip-f-92.princeton.org:8983

cwd=/Users/Shared/www/apache-solr-3.5.0-quickstart SolrHome=multicore/d7-core/

HTTP caching is ON

Solr	[SCHEMA] [CONFIG] [ANALYSIS] [SCHEMA BROWSER] [STATISTICS] [INFO] [DISTRIBUTION] [PING] [LOGGING]
Cores:	[D6-CORE] [D6.X-3.X-CORE] [d7-core] [EXT]
App server:	[JAVA PROPERTIES] [THREAD DUMP]

Make a Query [\[FULL INTERFACE\]](#)

Query String:

Search

Assistance [\[DOCUMENTATION\]](#) [\[ISSUE TRACKER\]](#) [\[SEND EMAIL\]](#) [\[SOLR QUERY SYNTAX\]](#)

Current Time: Sat Feb 02 10:52:41 EST 2013

Server Start At: Sat Feb 02 10:52:24 EST 2013

Enable the Modules

▼ SEARCH TOOLKIT

ENABLED	NAME	VERSION	DESCRIPTION
<input type="checkbox"/>	Apache Solr Access	7.x-1.1	Integrates node access and other permissions with Solr Requires: Apache Solr framework (enabled)
<input checked="" type="checkbox"/>	Apache Solr framework	7.x-1.1	Framework for searching with Solr Required by: Apache Solr search (enabled), Acquia search (disabled), Apache Solr external index (disabled)
<input checked="" type="checkbox"/>	Apache Solr search	7.x-1.1	Search with Solr Requires: Search (enabled), Apache Solr framework (enabled) Required by: Acquia search (disabled), Apache Solr external index (disabled)
<input checked="" type="checkbox"/>	Current Search Blocks	7.x-1.2	Provides an interface for creating blocks containing search results Requires: Facet API (enabled), Chaos tools (enabled)
<input checked="" type="checkbox"/>	Facet API	7.x-1.2	An abstracted facet API that can be used by various search engines Requires: Chaos tools (enabled) Required by: Apache Solr external index (disabled), C...



Search Environments Reference Different Servers and/or Config

- Most people need only one to start.
- The most important use is to bundle different sets of enabled facets and their configuration - e.g. for different search pages.
- Can also be used to search multiple servers.
- Each has its own ID and config variables.

Apache Solr search

DEFAULT INDEX

PAGES/BLOCKS

SETTINGS

[+ Add search environment](#)

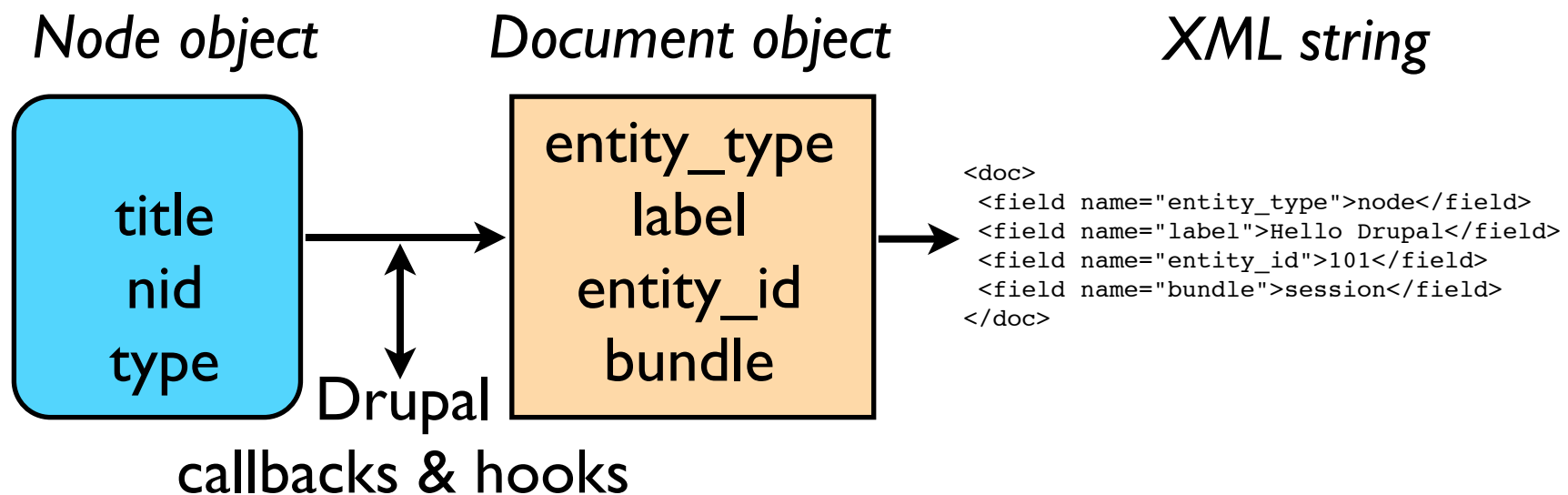
NAME	URL	CONFIGURATION			OPERATIONS	
D7 server <i>(Default)</i>	http://localhost:8983/solr/d7- core	Facets	Bias	Index	Edit	Clone

▶ **ADVANCED CONFIGURATION**

Save configuration

The Module Has a Pipeline for Indexing Drupal Content to Solr

- Drupal entities are processed into one (or more) document objects. Each document object is converted to XML and sent to Solr.



VIEW

EDIT

DEVEL

Apache Solr

Load

Render

... (Array, 36 elements)

id (String, 16 characters) a520t1/node/2203

site (String, 20 characters) http://drupal-7.dev/

hash (String, 6 characters) a520t1

entity_id (String, 4 characters) 2203

entity_type (String, 4 characters) node

bundle (String, 7 characters) article

bundle_name (String, 7 characters) Article

path (String, 9 characters) node/2203

url (String, 37 characters) http://drupal-7.dev/node-2203-article

language (String, 3 characters) und

path_alias (String, 17 characters) node-2203-article

label (String, 21 characters) Consectetuer Pertineo

content (String, 2937 characters) Abluo aliquam eu facilisi...

teaser (String, 296 characters) Abluo aliquam eu facilisi...

ss_name (String, 7 characters) brutrod

tos_name (String, 7 characters) brutrod

How can I debug Solr queries?

- Enable extra debugging info

```
select/
```

```
?q=Robin+Hood&debugQuery=on&debug=on
```

- Indentation and analysis!

```
select/?q=Robin+Hood&indent=true
```

```
admin/analysis.jsp?highlight=on
```

- Tomcat logs, jetty logs!

Basic params: q, start, rows, sort

- Query (q)

```
select/?q=superhero
```

- start, rows, sort

```
select/?q=superhero&start=0&rows=10&  
sort=sort_name+asc
```

More key params: fq, fl

- Filter Query (fq)

```
select/?q=superhero&fq=bundle:person&fq=attribute:cape
```

- Fields (to return) (fl)

```
select/?q=superhero&fl=id,entity_id,name,attribute,score
```


Highlighting and query type

- Highlighting (hl, hl.q, hl.fl)

```
select/?q=superhero&hl=true&hl.q=super&hl.fl=name,content,comments
```

- query parser: defType (or qt)

```
select/?q=superhero+AND+evil&defType=edismax
```

Edismax params: q.alt, qf, pf

- Alternative Query (q.alt), same as `defType=luene`

```
select/?q.alt=bundle:person
```

- Query fields (qf)

```
select/?q=Superhero&qf=teaser^2.0
```

- Phrase Fields (pf)

```
select/?q=Robin Hood&pf=name^10
```

Built-in monitoring and status

- Average Time Per Request & Requests per second

```
solr/core_name/admin/mbeans?  
wt=json&stats=true&  
key=org.apache.solr.handler.component.SearchHa  
ndler& stats=true&cat=QUERYHANDLER
```

Monitoring output

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 1},
  "solr-mbeans": [
    "QUERYHANDLER", {
      "org.apache.solr.handler.component.SearchHandler": {
        ...
        "docs": null,
        "stats": {
          "handlerStart": 1345463690388,
          "requests": 2,
          "errors": 0,
          "timeouts": 0,
          "totalTime": 75,
          "avgTimePerRequest": 37.5,
          "avgRequestsPerSecond": 0.0013287809}}}}]}
```

List Solr cores

- Number and identity of the cores

`/admin/cores?wt=json&action=STATUS`

```
{
  "responseHeader":{
    "status":0,
    "QTime":6},
  "status":{
    "core0":{
      "name":"core0",
      "instanceDir":"multicore/core0/",
      "dataDir":"multicore/core0/data/",
      "startTime":"2012-08-20T11:54:50.275Z",
      "uptime":2015408,
      "index":{
        "numDocs":887,
        "maxDoc":1279,
        "version":1323430446081,
        "segmentCount":5,
        "current":true,
        "hasDeletions":true,
        "lastModified":"2012-08-02T15:43:12Z"}},
    }
```

Keeping it secure (when live)

- NO security by default, not present at all for per-core security. Google this:
"[SCHEMA] [CONFIG] [ANALYSIS] [SCHEMABROWSER]"
- firewall rules
- SSL (behind a load balancer or configure tomcat)
- SSL + basic auth
- Acquia Search uses HMAC authentication & validation that is secure with or without SSL

The take home

- Solr communicates over HTTP - you can use the URL as a debugging “command line”
- Local install takes < 5 minutes - just do it
<http://drupal.org/apachesolr/quick-start-solr-3>
- Learn: <http://wiki.apache.org/solr/>
- Integrate: <http://drupal.org/project/apachesolr>

- Acquia is hiring - talk to me if interested!