

Solr-powered Search

Introduction

Apache Solr is a search engine platform which is relatively mature and has a lot of powerful and flexible configurations. There has been extensive work to implement it into the HUBzero CMS and is currently a work-in-progress.

Solr is an open-source, mature, and stable searching service that is built upon the Apache Lucene search engine. The service provides features which lend itself to scaling and has a rich open source community. It is a Java-based service which provides search results through HTTP. Many companies such as Instagram, eBay, and StubHub rely on Solr to provide advanced searching capabilities.

The integration with Solr is currently under heavy development. It is **strongly recommended** to test on a QA / Stage host before using in a production environment.

Installation & First Time Configuration

Step 1: Install the hubzero-solr package

A system administrator must install the hubzero-solr RedHat or Debian Package using a package manager such as yum or aptitude. The package contains a version of Apache Solr and the configuration necessary for Solr to integrate with the CMS.

For RedHat / CentOS:

```
$ sudo yum install hubzero-solr
```

For Debian:

```
$ sudo apt-get install hubzero-solr
```

Once installed the service will need to be enabled.

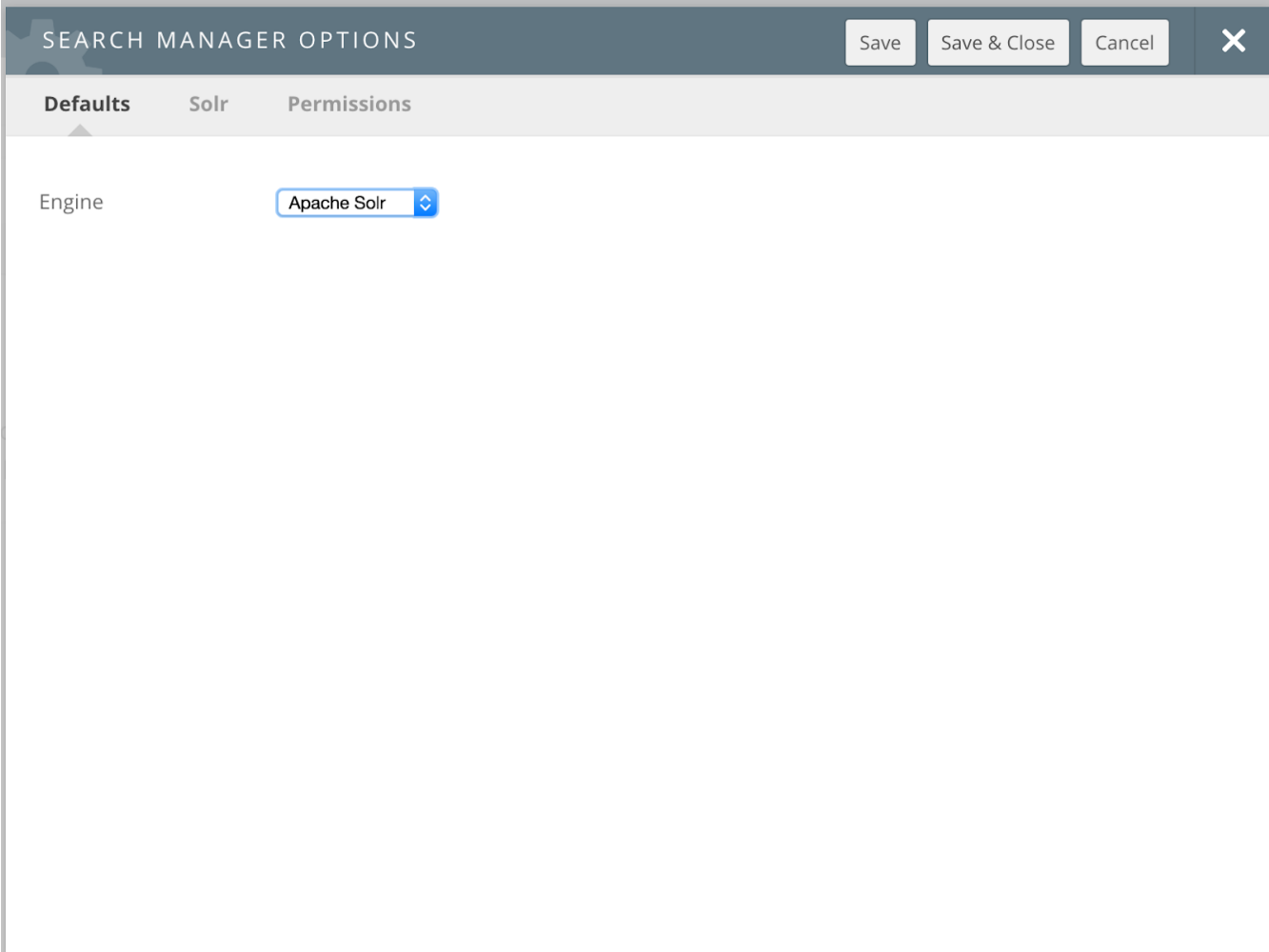
```
$ sudo service hubzero-solr start
```

Step 2: Configure Search Service in the CMS

SOLR-POWERED SEARCH

The HUBzero CMS needs to know to use Solr Search instead of Basic Search. To do this, a Hub administrator will need to log into the Administrative Backend and Configure the Search Component.

You will need to set **Engine** to *Apache Solr*. Then click the "Solr tab".



The screenshot shows a dialog box titled "SEARCH MANAGER OPTIONS". At the top right, there are buttons for "Save", "Save & Close", "Cancel", and a close icon (X). Below the title bar, there are three tabs: "Defaults", "Solr", and "Permissions". The "Solr" tab is selected. In the main area, there is a label "Engine" followed by a dropdown menu that currently displays "Apache Solr".

The Solr tab's default settings will work for the open-source distribution.

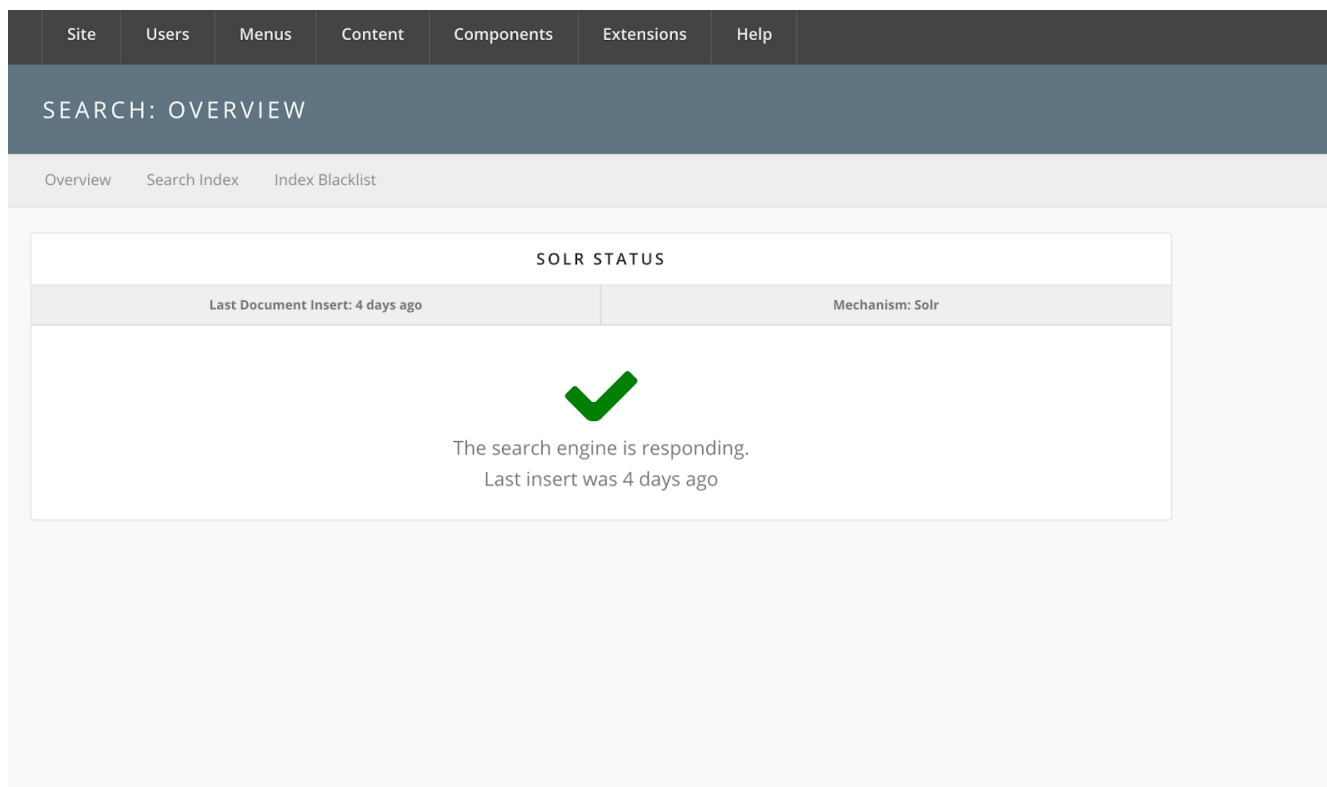
SOLR-POWERED SEARCH

HUBzero-hosted hubs are configured with different ports! The following scheme is used:

```
Development (dev.hub.org): 2090
Stage (stage.hub.org): 2091
Scan / QA (qa.hub.org): 2092
Production (hub.org): 2093
```

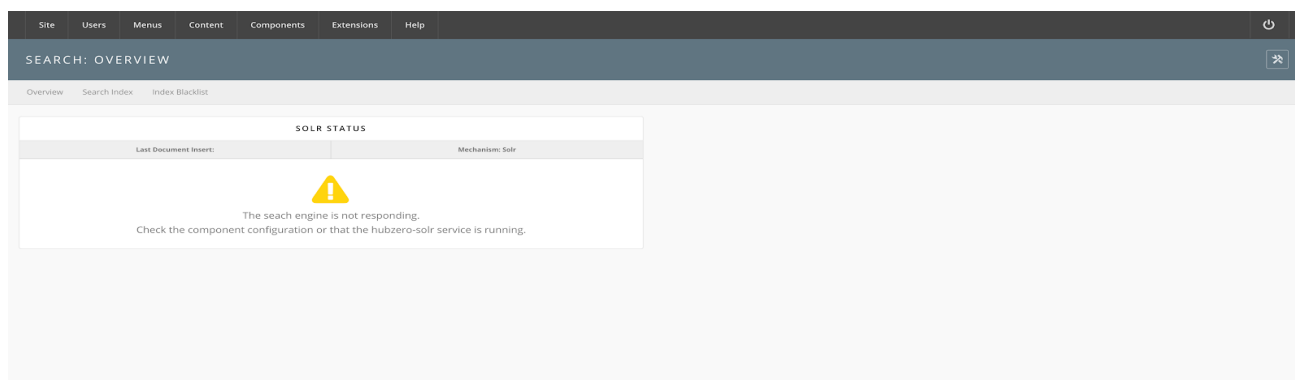
Click "Save and Close" to save the settings. If the hubzero-solr service is started and the correct settings were set in the steps above, the status screen should indicate that the search engine is responding.

SOLR-POWERED SEARCH



The screenshot shows the 'SEARCH: OVERVIEW' page with a navigation menu at the top containing 'Site', 'Users', 'Menus', 'Content', 'Components', 'Extensions', and 'Help'. Below the menu, there are tabs for 'Overview', 'Search Index', and 'Index Blacklist'. The main content area is titled 'SOLR STATUS' and contains a table with two columns: 'Last Document Insert: 4 days ago' and 'Mechanism: Solr'. A large green checkmark is centered on the page, with the text 'The search engine is responding. Last insert was 4 days ago' below it.

If there were any issues with configuration, the following screen will appear.



The screenshot shows the 'SEARCH: OVERVIEW' page with a navigation menu at the top containing 'Site', 'Users', 'Menus', 'Content', 'Components', 'Extensions', and 'Help'. Below the menu, there are tabs for 'Overview', 'Search Index', and 'Index Blacklist'. The main content area is titled 'SOLR STATUS' and contains a table with two columns: 'Last Document Insert:' and 'Mechanism: Solr'. A yellow warning icon is centered on the page, with the text 'The search engine is not responding. Check the component configuration or that the hubzero-solr service is running.' below it.

This would be a point where a support ticket is filed for the system administrator to confirm that the service is running. Please include all configuration parameters contained in Step #3 when filing the ticket.

Step 3: Enable the Search Background Worker

In order to keep the search index fresh, a background worker is implemented to process data from the CMS and push it into the Solr service.

Currently the background worker is implemented as a Cron task that is called once a minute. There is work being done to develop a daemon which listens to CMS events and processes data without relying on Cron.

SOLR-POWERED SEARCH

To setup the Cron-based worker a Hub administrator must go into the Administrative Backend, go to Components, Cron, and add the Task as shown below:

Click "Save and Close".

Step 4: Build the Initial Index

This implementation of Solr has hooks into the CMS which updates the index when a new record is added or marked for deletion. It will be necessary to add items which have been added before Solr was activated.

SOLR-POWERED SEARCH

The screenshot shows the 'SOLR SEARCH: INDEXED HUBTYPES' interface. At the top, there is a navigation menu with 'Site', 'Users', 'Menus', 'Content', 'Components', 'Extensions', and 'Help'. Below this, the title 'SOLR SEARCH: INDEXED HUBTYPES' is displayed. Underneath the title, there are three tabs: 'Overview', 'Search Index', and 'Index Blacklist'. A 'Full Index' button is highlighted with a red box and a red arrow. To the right of the 'Full Index' button, there are 'Filter', 'Go', and 'Clear' buttons. Below these buttons is a table with three columns: 'HUB TYPE', 'INDEXED DOCUMENTS', and an empty column. The table contains the following data:

HUB TYPE	INDEXED DOCUMENTS	
blog-entry	0	
citation	0	
content	19	
event	0	
forum	0	

This operation should only need to be completed once. You will be unable to start this operation until it finishes for the first time.

The screenshot shows the 'SOLR SEARCH: INDEXED HUBTYPES' interface. At the top, there is a navigation menu with 'Site', 'Users', 'Menus', 'Content', 'Components', 'Extensions', and 'Help'. Below this, the title 'SOLR SEARCH: INDEXED HUBTYPES' is displayed. Underneath the title, there are three tabs: 'Overview', 'Search Index', and 'Index Blacklist'. A blue warning message is displayed: 'The index is still building, depending on the amount of Hub content it may take some time.' Below the warning message, there are 'Filter', 'Go', and 'Clear' buttons. Below these buttons is a table with three columns: 'HUB TYPE', 'INDEXED DOCUMENTS', and an empty column. The table contains the following data:

HUB TYPE	INDEXED DOCUMENTS	
blog-entry	0	
citation	0	
content	0	
event	0	
forum	0	
group	1	
kb-article	29	
member	2	
question	0	
resource	0	
wiki	1	
wishlist	1	
course	0	
collection	2	
project	0	
publication	0	

The "Full Index" button populates a Queue which is periodically serviced by a worker. The worker will process the records and format for consumption by the Solr service. **This may take several hours to fully complete if the Hub has a lot of content.**

If an error with the worker occurs, a warning message such as this will appear.

SOLR-POWERED SEARCH

Site Users Menus Content Components Extensions Help

SOLR SEARCH: INDEXED HUBTYPES

Overview Search Index Index Blacklist

The index build has stalled. Please check if the Cron task is reporting errors.

Filter Go Clear

<input type="checkbox"/>	HUB TYPE ▾	INDEXED DOCUMENTS ▾	
<input type="checkbox"/>	blog-entry	0	
<input type="checkbox"/>	citation	0	
<input type="checkbox"/>	content	19	
<input type="checkbox"/>	event	0	
<input type="checkbox"/>	forum	0	
<input type="checkbox"/>	group	1	
<input type="checkbox"/>	kb-article	29	
<input type="checkbox"/>	member	2	
<input type="checkbox"/>	question	0	
<input type="checkbox"/>	resource	0	
<input type="checkbox"/>	wiki	1	
<input type="checkbox"/>	wishlist	1	
<input type="checkbox"/>	course	0	
<input type="checkbox"/>	collection	2	
<input type="checkbox"/>	project	0	
<input type="checkbox"/>	publication	0	

Search Breadth

The question is “What can I search for?”. The answer is “anything you have access to contained within the list in Search Categories. To see all content within these categories perform a simple query using the wildcard character “*” as shown below.

A better of what is currently inside Solr's index can be viewed on the administrative backend by going to Components >> Search >> Search Index Tab. The number of index items is located to next to each type. Clicking on the name of the hub type will perform a search on that type, displaying all items that are within the index of that type.

SOLR-POWERED SEARCH

<input type="checkbox"/>	HUB TYPE ▾	INDEXED DOCUMENTS ▾	COM_SEARCH_COL_LAST_INDEXED ▾
<input type="checkbox"/>	blog-entry	3	
<input type="checkbox"/>	citation	2	
<input type="checkbox"/>	content	3	
<input type="checkbox"/>	event	2	
<input type="checkbox"/>	forum	5	
<input type="checkbox"/>	group	3	
<input type="checkbox"/>	member	6	
<input type="checkbox"/>	question	1	
<input type="checkbox"/>	resource	360	
<input type="checkbox"/>	wiki	4	
<input type="checkbox"/>	course	1	
<input type="checkbox"/>	project	3	

For instance clicking “Resources” shows the following screen:

SOLR-POWERED SEARCH

SOLR SEARCH: RESOURCE DOCUMENTS

Overview Search Index Index Blacklist

Filter Go Clear

HUBID	HUBTYPE	TITLE	ACCESS_LEVEL	OWNER_TYPE	OWNER	
-	resource	Effects of Pre-Growth Catalyst Annealing on Carbon Nanotube Synthesis	public	user	-	Add to blacklist
-	resource	Feasibility of Molecular Assemblers	public	user	-	Add to blacklist
-	resource	Hydrodynamic Separation of Micron-sized Particles through Magnetization	public	user	-	Add to blacklist
-	resource	Spectroscopic Studies of Dye-Metal Complexes	public	user	-	Add to blacklist
-	resource	Measurements of Interface Trap Density in MOS Capacitors Using AC Conductance Method	public	user	-	Add to blacklist
-	resource	Visualization of CNT FET Electrical Field Lines	public	user	-	Add to blacklist
-	resource	Self-Assembled Monolayer of DNA Microarrays	public	user	-	Add to blacklist
-	resource	Visualization of and Educational Tool for Quantum Dots	public	user	-	Add to blacklist
-	resource	Quantum Dots Visualization Software using Electron Wave Function	public	user	-	Add to blacklist
-	resource	SURI 2004 Program Highlights	public	user	-	Add to blacklist

< 1 2 3 4 5 6 > Display # 10 Results 1 - 10 of 360

One can perform additional searching using the “Filter” bar on top of the results listing.