

Search

Installation & First Time Configuration

1. A system administrator must **install the hubzero-solr** RedHat or Debian Package using a package manager such as yum or aptitude
 1. On RedHat: `sudo yum install hubzero-solr`
 2. On Debian: `sudo apt-get install hubzero-solr`
2. To start the service: **sudo service hubzero-solr start OR sudo /etc/init.d/hubzero-solr start**
3. A Hub administrator may configure the Hubzero CMS to use Solr instead of basic search by going to Components > Search > Options > Engine > Apache Solr
4. A Hub administrator must further configure the service by clicking the “Solr” tab in the Search Options. The default parameters will work out-of-the-box for Open Source Hubs. HUBzero Managed Hubs will use the following port-numbering.
 - Development: 2090
 - Stage: 2091
 - Scan/QA: 2093
 - Production: 2093
 - Note: The default configuration should be acceptable. Running Solr on another host, changing the core, the path, or the log path are at the Hub Administrator's own risk.
5. Click **Save & Close**
6. Return to the **Search** and you should see a status page like the one below:
7. If this is the first time the Hub has used Solr as its search engine, the index will need to be filled with current data
 1. This is a lengthy operation during the first run
 2. Unless there is massive data corruption, this will only need to be performed once
 3. To check the progress of the indexing process, you may click on the **Index Queue** tab in the **Administrative Search** interface
8. Once the full index has been built, searching can be completed from the enabled search interfaces

If there were any issues with configuration, please submit a support ticket and the HUBzero team will assist you further.

Maintaining the Index

There should be very little effort needed to maintain the index. Solr maintains the index and the HUBzero CMS will instruct Solr to add, remove, or update records inside of its index.

Solr saves its index on the filesystem of the server which allows the retention of data if the server needs to reboot or the Solr process crashes. This prevents having to rebuild the index from scratch in such events.

IMPORTANT NOTE: Due to the large amount of processing power needed to convert database content into a searchable document and the need to communicate with a system outside of the CMS, changes to the index **WILL NOT** be reflected immediately. The queue will be worked on a first-in-first-out basis. This means that the oldest item in the work queue will be processed first. The amount of time it takes to perform indexing operations depends on the amount of data contained on the hub. If there is a large amount of content, the time to perform a full index will be greater. Once the full index is built, indexing operations should be noticeably quicker

Plugins

Search plugins allow developers to add support for different component data in the hub. In order to appear in the search index, the plugin must be enabled. This can be accomplished by going to the Administrative backend > Plug-in Manager > [Filter by Type: Search] > and ensuring that the types are enabled. For example Solr will index wiki pages when the “Search Wiki” plugin is enabled.

The plugin provides some necessary information for indexing and other search-related operation. By default, the categories in the search interface correspond with these plugins.

System Search Plugins

In order to keep the search index “fresh”, a couple of new system Events have been made that capture when items using the Relational Class / ORM are created, edited, or deleted. Once the system event fires, it calls an event in the Search - Index plugin which handles placing the newly-updated data into the processing queue. A migration has been written to ensure that the System - Content plugin and the Search - Solr plugin have been enabled. If you notice that the index is not being refreshed with new content, check that both of these plugins are enabled.

Blacklist

The blacklist allows Hub administrators to “strike” things from the search index. This may be necessary to override If Solr indexes something, it will be reindexed unless it is on the blacklist.

To remove an item from the search index, go to the Administrative Backend > Component > Search > Search Index and Click on the name of the type of record you would like to remove. Let's say, for example, you needed to remove a Resource.

1. Click **Resource**
2. You will then see all resources indexed by Solr
3. You may use the search box to locate the record
4. Once you locate the record, click **Add to Blacklist**
5. Once the button is pressed, the request to remove the record will be placed into the queue
6. Once the worker processes the record, it will no longer be searchable by anyone

Administration

Restarting Solr

If Solr needs to be restarted, a system administrator can issues the following commands:

1. `sudo service hubzero-solr restart`
2. `sudo /etc/init.d/hubzero-solr restart`

Solr Index CRON Updater

In order to keep the Solr index up-to-date, the CMS will periodically call a routine to process the queue. Although CRON is not the best tool for the job, it will dutifully process the queue every minute if configured. There are plans to develop a background process which will make this process more efficient.

To configure the CRON Task, follow these steps:

1. Navigate to **/administrator**
2. Hover over **Components** and click **Cron** from the drop-down
3. Click **Add a New Task** and set the "Event:" to **Process Queue**
4. The **New Cron Task** should be configured to run every minute

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 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.

For instance, clicking “Resources” shows the following screen:

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

Boosting

Search boosting allows hub administrators to move categories of matching results higher or lower in Solr search results.

Note: your hub must have Solr installed and enabled as boosting only affects Solr search results.

To create a boost follow the steps below:

1. Access your hub's administrator portal
2. Click on Search under the Components drop-down menu
3. Click on the Boosts tab under the component title bar
4. Click on the plus in the upper right hand corner
5. Select a category from the Type drop-down
6. Enter a number for the boost strength
 1. Positive value moves matching results higher
 2. Negative value moves matching results lower
7. Save the boost

1.

1.

1.