

Publish to SolrCloud CDH App Bundle

On this page:

- [Configuration](#)
- [Example Configuration](#)
- [Edit Xsl](#)
 - [Add metadata field](#)
 - [Change the document ID](#)
 - [Advanced Edit](#)

The Publish to SolrCloud CDH application sends document feeds to the SolrCloud index through SolrJ. SolrJ has a CloudSolrClient (or HTTPSolrClient) class to communicate with SolrCloud.

Instances of this class communicate with Zookeeper to discover Solr endpoints for SolrCloud collections (or directly to a Solr endpoint), and then use the LBHttpSolrClient to issue requests. The feed to Solr can be customized by editing the XSL transformation file provided by the user.

Publish to SolrCloud	
Factory Name	com.searchtechnologies.aspire:app-publish-to-solrj
subType	default
Inputs	AspireObject from a connector's subjob with metadata and content extracted from a specific file/folder.
Outputs	An XML transformation of the AspireObject sent to the Solr's xmlfeed URL.
Versions	3.1
Type Flags	job-input

Configuration

This section lists all configuration parameters available to configure the Publish to SolrCloud component.

Property	Type	Default	Description
connectMethod	string	zk	Connection method to communicate with Solr
zkHosts	string	none	Zookeeper hosts and port comma separated or Solr host and port (depends of connection method selected)
chroot	string	/solr	Zookeeper route for Solr znode tree.
collection	string	none	Solr collection name to submit post documents.
postXsl	string	\${appbundle.home}/config/xsl/aspireToSolr.xsl	Location of the XSL to transform the job data to a Solr feed. See Edit Xsl .
maxTries	int	3	Maximum number of retry attempt to index.
retryWait	int	3	Seconds to wait before re-trying to index.
fieldListPath	string	/add/doc/fieldList	XPath expression to process field list.
idField	string	url	Field used to identify the jobs on Solr Cloud
commit	boolean	false	Force commit before send a batch
commitMS	int	30000	Milliseconds to wait before send the batch to index
zkClientTimeout	int	10000	Zookeeper Solr Cloud Client timeout
zkConnectTime out	int	10000	Zookeeper Solr Cloud Connect timeout
useKerberos	boolean	false	To use kerberos authentication
coreSite	string	none	the hadoop core-site.xml path
jaasOptions	string	none	The configuration file properties path
debug	boolean	false	Enable debug messages.

Example Configuration

```

<!-- Post to Solr stage. This will post to a running Solr instance a new document -->
<component name="SolrJPublisher" subType="default" factoryName="aspire-solrj-publisher-cdh">
  <debug>${debug}</debug>
  <connectMethod>${connectMethod}</connectMethod>
  <zkHosts>${zkHosts}</zkHosts>
  <chroot>${chroot}</chroot>
  <collection>${SolrCollection}</collection>
  <postXsl>${aspireToSolrXsl}</postXsl>
  <maxTries>${maxTries}</maxTries>
  <retryWait>${retryWait}</retryWait>
  <fieldListPath>${fieldListPath}</fieldListPath>
  <idField>${idField}</idField>
  <useKerberos>${useKerberos}</useKerberos>
  <commit>${commit}</commit>
  <commitMS>${commitMS}</commitMS>
  <zkClientTimeout>${zkClientTimeout}</zkClientTimeout>
  <zkConnectTimeout>${zkConnectTimeout}</zkConnectTimeout>
  <jaasOptions>${jaasOptions}</jaasOptions>
  <coreSite>${coreSite}</coreSite>
  <trustAllCertificates>${trustAllCertificates}</trustAllCertificates>
</component>

```

Edit Xsl

The default XSL transformation file can be found in [AspireToSolr.xsl](#).

The default transformation XSL file provided by the publisher expects metadata as described in [Connector AspireObject Metadata](#).

Add metadata field

To add a new metadata field extracted by an Aspire Connector add an XSL element under the **<doc>** tag.

```

<field name="metafieldNameInSolr_t">
  <xsl:value-of select="metafieldNameFromAspireObject" />
</field>

```

Notice that the dynamic field `_t` is being used by default. If you have a Solr schema that supports your field, then just enter the field name as defined in the schema.

Change the document ID

The **id** of a Solr document is used to uniquely identify a file in the index. By default, Publish To SolrCloud will use the following fields from the Aspire document in order of precedence (if one is missing, then the next will be used):

- `fetchUrl`
- `url`
- `displayUrl`
- `id`

If you want to change this behavior, edit or create a new XSL file which has the following element:

```

<field name="id">
  <xsl:value-of select="idFieldNameFromAspireObject" />
</field>

```

Advanced Edit

More advanced changes can be accomplished reading the [Solr Update XML Messages](#) wiki.