Publish to SolrCloud CDH How to configure

On this page:

Step 1. Launch Aspire and open the Content Source Management Page

Before launching Aspire, you need to change the felix.properties file and add these lines:

```
# To append packages to the default set of exported system packages,
# set this value.
org.osgi.framework.system.packages.extra=\
 ...
 sun.security.krb5, \
 com.sun.security.auth.callback

# The following property makes specified packages from the class path
# available to all bundles. You should avoid using this property.
org.osgi.framework.bootdelegation=\
 ...
 javax.security.sasl, \
 sun.security.krb5
```

Launch Aspire (if it's not already running). See:

- Launch Control
- Browse to: http://localhost:50505.

  For details on using the Aspire Content Source Management page, please refer to Admin UI

Unknown Attachment

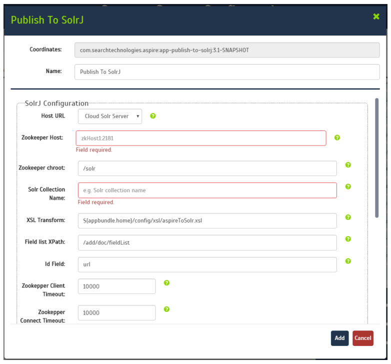
Step 2. Add a new Content Source

For this step, please follow the step from the Configuration Tutorial of the connector of you choice, please refer to Connector list

Step 3. Add a new Publish to SolrCloud to the Workflow

To add a Publish to SolrCloud, drag from the **Publish to SolrCloud CDH** (or a custom component using app-publish-to-sorlj-cdh coordinates) rule from the *Workflow Library* and drop to the *Workflow Tree* where you want to add it.

This will automatically open the *Publish to SolrCloud* window for the configuration of the publisher.



Coordinates: com.searchtechnologies.aspire:app-publish-to-solrj:3.1-SNAPSHOT

Name: Publish To SolrJ

SolrJ Configuration

Host URL: Cloud Solr Server ▾

Zookeeper Host: zkHost1:2181

Field required.

Zookeeper chroot: /solr

Solr Collection Name: e.g. Solr collection name

Field required.

XSL Transform: ${appbundle.home}/config/xsl/aspireToSolr.xsl

Field list XPath: /add/doc/fieldList

Id Field: url

Zookepper Client Timeout: 10000

Zookepper Connect Timeout: 10000

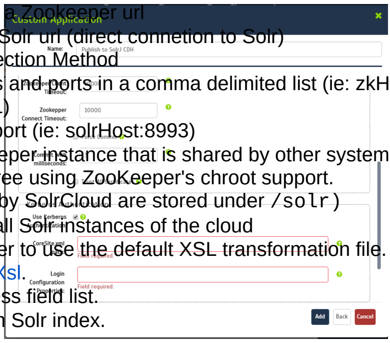
Step 3a. Specify Publisher Information

In the *Publish to SolrCloud* window, specify the following connection information to publish to SolrCloud.

Name: Enter a unique name for the publisher.

SolrJ Configuration

1. Choose a Connection Method:
   - **CloudSolrServer**: Connects using a Zookeeper url
   - **HttpSolrClient**: Connects using a Solr url (direct connetion to Solr)
2. Enter a Host URL, depending on the Connection Method
   - **Zookeeper Host**: Zookeeper hosts and ports in a comma delimited list (ie: zkHost1:2181, zkHost2:2181, zkHost3:2181)
   - **Solr Host**: A single Solr host and port (ie: solrHost:8993)
3. **Zookeeper chroot**: If you're using a ZooKeeper instance that is shared by other systems, we recommend isolating the SolrCloud znode tree using ZooKeeper's chroot support. (For example, to ensure all znodes created by SolrCloud are stored under `/solr`)
4. **Solr Collection Name**: Same collection of all Solr instances of the cloud
5. **XSL Transform**: Set to the default parameter to use the default XSL transformation file. To use a custom file, follow the instructions in Edit Xsl.
6. **Field list XPath**: XPath expression to process field list.
7. **Id Field**: The Id used for identification the on Solr index.



Name: Publish to SolrJ CDH

Zookepper Connect Timeout: 10000

Use Kerberos

CoreSite.xml

Login Configuration Properties

Field required.

Custom Application

1. **Zookeeper Client Timeout**: Timeout for the client.
2. **Zookeeper Connect Timeout**: Timeout for the client.
3. **Force Commit**: If true, force a commit every time a batch is sent to the index.
4. **Commit with milliseconds**: Milliseconds to wait before commit a batch to the index.

Kerberos Authentication

1. **Use Kerberos Authentication**: Enable the authentication through Kerberos
   - **CoreSite.xml**: The path for the hadoop core-site.xml file
   - **Login Configuration Properties**: The path of the pseudo jaas properties file, please check Login Config File.
2. **Max Tries**: Maximum number of retry attempt to index.
3. **Retry Wait**: Seconds to wait before re-trying to index.
4. **Debug**: Select the check box to run the publisher in debug mode.

Click ***Add***. It will take a moment for Aspire to download all of the necessary components (Jar files) from the Maven repository and load them into Aspire.

When that's done, the publisher will appear in the *Workflow Tree*.

> ⓘ For details on using the *Workflow* section, please refer to Workflow introduction.

Step 3b. Edit Login Config File

The Login Config file is a pseudo JAAS (Java Authentication and Authorization Service) file. Basically, use only the properties inside the JAAS file.

For example: **Normal JAAS File**

```
Client {
 com.sun.security.auth.module.Krb5LoginModule required
 useKeyTab=true
 keyTab="/home/centos/solr.keytab"
 storeKey=true
 doNotPrompt=true
 useTicketCache=false
 principal="solr/novalocal@DEV.LOCAL";
};
```

The **Login Config file** for the publisher should be something like this, with only the properties **propertyName=propertyValue** (module name and login type are not necessary).

```
useKeyTab=true
keyTab="/home/centos/solr.keytab"
storeKey=true
doNotPrompt=true
useTicketCache=false
principal="solr/novalocal@DEV.LOCAL"
```