

SMB How to Configure

On this page

- [Step 1. Launch Aspire and Open Content Source Management](#)
- [Step 2. Add an SMB Content Source](#)
 - [Step 2a. Specify basic information](#)
 - [Step 2b. Specify the connector information](#)
 - [Step 2c. Specify workflow information](#)
- [Step 3: Initiate a Full Crawl](#)
 - [During the Crawl](#)
- [Step 4: Initiate an Incremental Crawl](#)

Step 1. Launch Aspire and Open Content Source Management

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

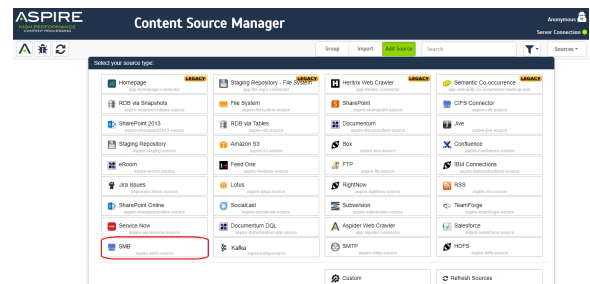
See:

- [Launch Control](#)
- Browse to: <http://localhost:50505>. For details on using the Aspire Content Source Management page, please refer to [Admin UI](#)

Step 2. Add an SMB Content Source

To specify exactly which shared folder to crawl, create a new "Content Source".

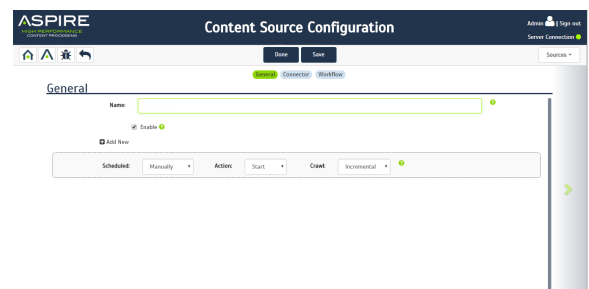
1. From the **Content Source** options, select **Add Source**.
2. Select **SMB**.



Step 2a. Specify basic information

In the **General** tab in the **Content Source Configuration** window, specify basic information for the content source.


1. Enter a content source name in the **Name** field.
The name you specify for the source will be displayed in the **Content Source** page and in error messages, etc.
2. From the **Scheduled** pull-down list, select one of the following: *Manually*, *Periodically*, *Daily*, *Weekly* or *Advanced*.
Aspire can automatically schedule content sources to be crawled on a set schedule, such as once a day, several times a week, or periodically (every N minutes or hours).
For the purposes of this tutorial, you may want to select **Manually** and then set up a regular crawling schedule later.





- From the **Action** pull-down list, select one of the following: *Start*, *Stop*, *Pause*, or *Resume*.
This is the action that will be performed for the specified schedule.
- From the **Crawl** pull-down list, select one of the following: *Incremental*, *Full*, *Real Time*, or *Cache Groups*.
This will be the type of crawl to execute for the specified schedule.

After selecting **Scheduled**, specify the details, if applicable:

- Manually:** No additional options.
- Periodically:** Specify the **Run every** options by entering the number of "hours" and "minutes."
- Daily:** Specify the **Start time** by clicking on the hours and minutes drop-down lists and selecting options.
- Weekly:** Specify the **Start time** by clicking on the hours and minutes drop-down lists and selecting options. Select the day check boxes to specify days of the week to run the crawl.
- Advanced:** Enter a custom CRON Expression (for example: 0 0 ? * *)

 You can add more schedules by selecting the **Add New** option, and rearranging the order of the schedules.

 If you want to disable the content source, clear the **Enable** checkbox. This is useful if the folder will be under maintenance and no crawls are wanted during that period of time.

 Real Time and Cache Groups crawls will be available depending of the connector.


Step 2b. Specify the connector information


In the **Connector** tab, specify the connection information to crawl the SMB2 Protocol.


- Enter the smb folder url to crawl.
- Enter the user's domain.
- Enter the username.
- Enter the user's password.
- Select the **Security** options as required:
 - Retrieve Security Information:** Adds ACLs to the connector output.
 - Resolve Security Information SID:** Translate the retrieved SIDs to a human readable form.
 - Add SID value to retrieved ACLs:** Add the corresponding SID value to each ACL.
 - Add ACL flags to retrieved ACLs:** Add the corresponding ACL flags retrieved by the connector to each ACL.
 - Add ACL type to retrieved ACLs:** Add the corresponding ACL type retrieved by the connector to each ACL.
 - Add ACL access mask to retrieved ACLs:** Add the corresponding ACL mask retrieved by the connector to each ACL.
- Select the **Connection** options as required:
 - Distributed File System (DFS) Resolution:** Enable DFS support.
 - Override Last Accessed date of documents:** Access files without modifying the "last accessed" value. *In SMB every time you access a file the field "last_accessed" is updated. This means that every time the connector goes through the files it will massively modify this date; to avoid this, mark this option to preserve the original date. Note that this feature will require write access to be granted to the connector's account*
- Select options as needed:
 - Index Containers:** Index subfolders as items. If the check boxes are cleared, only files will be indexed.


Connector


Credentials


Multiple Starting Points: 

URL: 
Field required


Partial Scan: 






Domain: 

Username: 
Field required



Password: 
Field required

Security

Retrieve Security Information: 

- ☒ Resolve Security Information SIDs 
- ☒ Add SID value to retrieved ACLs 
- ☐ Add ACL flags to retrieved ACLs 
- ☐ Add ACL type to retrieved ACLs 
- ☐ Add ACL access mask to retrieved ACLs 

Connection

- ☐ Distributed File System (DFS) Resolution 
- ☐ Override Last Accessed date of documents 

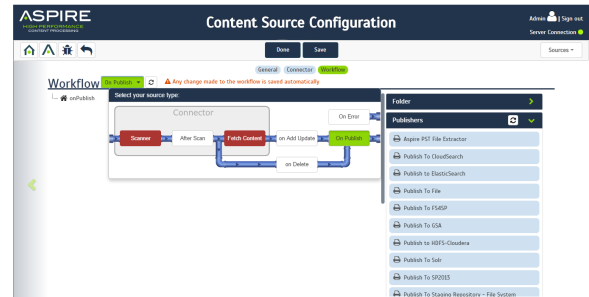
Scan recursively: Scan through the subfolder's child nodes.
Include/Exclude patterns: Enter regex patterns to include or exclude files/folders based on URL matches.

Step 2c. Specify workflow information

In the **Workflow** tab, specify the workflow steps for the jobs that come out of the crawl. Drag and drop rules to determine which steps an item should follow after being crawled.

These rules could be where to publish the document or transformations needed on the data before sending it to a search engine. See [Workflow](#) for more information.

1. For the purpose of this tutorial, drag and drop the *Publish To File* rule found under the **Publishers** tab to the **onPublish** Workflow tree.
2. Specify a **Name** and **Description** for the Publisher.
3. Click **Add**.
4. Click **Save** and **Done**. You'll be sent back to the **Home** page.



Step 3: Initiate a Full Crawl

Now that the content source is set up, the crawl can be initiated.

1. Click on the crawl type option to set it as **Full**.
2. After the first crawl, set it to **Incremental** to crawl for any changes done in the repository).
3. Click **Start**.

During the Crawl

You can do the following:

1. Click **Refresh** on the **Content Sources** page to view the latest status of the crawl. The status will show **RUNNING** while the crawl is going, and **CRAWLED** when it is finished.
2. Click **Complete** to view the number of documents crawled so far, the number of documents submitted, and the number of documents with errors.

If there are errors, you will get a clickable **Error** flag that will take you to a detailed error message page.

Step 4: Initiate an Incremental Crawl

If you only want to process content updates from the SMB (documents which are added, modified, or removed):

1. Click on **Incremental** instead of **Full**. The SMB connector will automatically identify only changes which have occurred since the last crawl.

If this is the first time that the connector has crawled, the action of "Incremental" depends on the exact method of *change* discovery. It may perform the same action as a **Full** crawl crawling everything, or it may not crawl anything. Thereafter, **Incremental** will only crawl updates.

 Statistics are reset for every crawl.