

Configuring the StageR Connector

Step 1. Launch Aspire and Open the Content Source Management page.

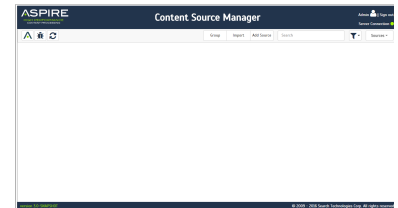
1. Launch Aspire (if it's not already running).
See [Launch Control](#)

2. Browse to: <http://localhost:50505>.

For details on using the Aspire Content Source Management page, please refer to [Admin UI](#).

On this page

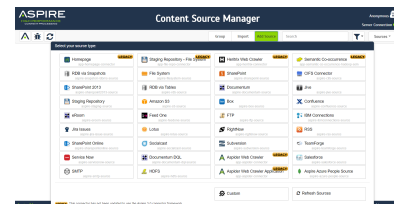
- [Launch Aspire, open Content Source Management page](#)
- [Add STageR content source](#)
- [Initiate full crawl](#)
- [Initiate incremental crawl](#)



Step 2. Add a New StageR Content Source.

To specify exactly what shared folder to crawl, we will need to create a new "Content Source".

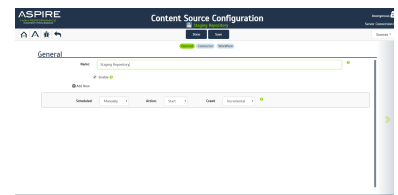
1. From the **Content Source**, click **Add Source**.
2. Click **StageR Connector**.





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.
 - a. This is any useful name which you decide is a good name for the source. It will be displayed in the content source page, in error messages, etc.
2. Click on the **Scheduled** pull-down list and select one of the following: *Manually, Periodically, Daily, Weekly or Advanced*.
 - a. 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.
3. Click on the **Action** pull-down list to select one of the following: *Start, Stop, Pause, or Resume*.
 - a. This is the action that will be performed for that specific schedule.
4. Click on the **Crawl** pull-down list and select one of the following: *Incremental, Full, Real Time, or Cache Groups*.
 - a. This will be the type of crawl to execute for that specific schedule.
5. After selecting a 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, then clicking on the day checkboxes to specify days of the week to run the crawl.
 - *Advanced*: Enter a custom CRON Expression (e.g. 0 0 0 ? * *)



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

 If you want to disable the content source just clear the **Enable** check box. 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 crawl will be available depending of the connector.

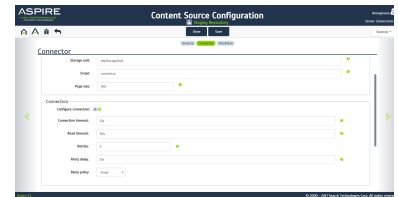
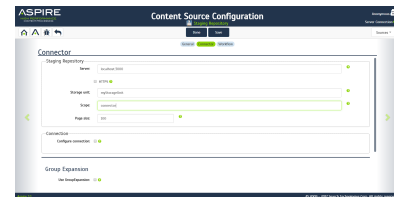
Step 2b. Specify the connector information

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

- a. Server
 - i. The host and port of the server the STageR is running on
 1. The STageR's default port is 3000
- b. Use https
 - i. Check this box if your STageR requires an https connection
- c. Storage Unit
 - i. The name of the STageR storage unit you wish to crawl
- d. Scope
 - i. The scope with in the STageR's storage unit you wish to crawl
- e. Page size
 - i. The number of transactions to retrieve from the STageR in a single call

If required, you can also configure the connection to the STageR. Check the *configure connection* check box to allow you to configure:


- a. Connection timeout
 - i. Period to attempt to connect to the STageR for before timing out
- b. Read timeout
 - i. Period to attempt to read from the STageR for before timing out
- c. Retries
 - i. The number of retries that will be attempted when a connection or read timeout occurs before the STageR reports an error
- d. Retry delay
 - i. The period to wait after a failure before retrying the operation
- e. Retry policy
 - i. Fixed
 1. The same delay will apply on each retry
 - ii. Increasing
 1. The delay will be the attempt number multiplied by a fixed value
 - a. eg for a value of 2, the delay would be 2, 4, 6, 8 etc
 2. A maximum delay can be specified
 - iii. Cumulative
 1. The delay will be increased by a multiple on each attempt
 - a. eg for a delay of 3 and a multiplier of 2, the delay would be 3, 3x2, 3x2x2, 3x2x2x2 (3, 6, 12, 24) etc
 2. A maximum delay can be specified



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 should an item 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.

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

 Unknown Attachment

Step 3: Initiate a full crawl.

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

1. Click on the full crawl icon to start a crawl (if you start an incremental crawl the first time, it'll work like a full crawl as all content will be seen as new)
2. Click **Start**.

During the crawl

During the crawl, you can do the following:

- 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.
- 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 STageR (documents which are added, modified, or removed), then click **Incremental** instead of **Full**. The STageR connector will automatically identify only changes which have occurred since the last crawl, using the transactions returned from the *STageR*.

If this is the first time that the connector has crawled, the action of the "Incremental" crawl will be the same as a "full" crawl. Thereafter, the Incremental option will only crawl updates.



Statistics are reset for every crawl.