SMTP Connector How to Configure

Before Beginning, Create User account and Set Access Right sections should be in prerequisites.

? Unknown Attachment

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

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, refer to Admin UI

Step 2. Add a new SMTP Content Source

? Unknown Attachment

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

1. From the **Content Source** , click **Add Source**.
2. Click **SMTP Connecto**r.

Step 2a. Specify Basic Information

? Unknown Attachment

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.

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 selecting the day check boxes to specify which 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 in the **Add New** option, and rearranging the order of the schedules.

ⓘ If you want to disable the content source just clear the the **Enable** check box. This is useful if the folder will be under maintenance and no crawls are desired during that period of time.

⚠

ⓘ

Step 2b. Specify the Connector Information

? Unknown Attachment

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

1. Server Name
    a. Host name to be used by the SMTP server to get the e-mails.
    b. Example: localhost
2. Server Port
    a. Port where the SMTP server will be listening for incoming e-mails.
    b. Example: 26000
3. Queue size:
    a. Amount of e-mails to be processed before filling the queue again.
    b. Example: 1000
4. Sleep time:
    a. Time in milliseconds to wait before checking for new e-mails.

Step 2c. Specify Workflow Information

? Unknown Attachment

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.

- Drag and drop the *Publish To File* rule found under the **Publishers** tab to the **onPublish** Workflow tree.
    1. Specify a **Name** and **Description** for the Publisher.
    2. Click **Add**.

After completing this step, click **Save** and **Done** to return to the Home Page.

Step 3: Initiate a 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. 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. It will continue until it is stopped by a user, and will show as **COMPLETE** when it
  finish the on-going processing jobs.
- 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.