

OneDrive How to Configure

On this page

- [Step 1. Launch Aspire and open the Content Source Management page](#)
- [Step 2. Add a new OneDrive 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 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 OneDrive content source

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

1. From the Content Source, select "Add Source".
2. Select "OneDrive Connector".

? Unknown Attachment

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** pulldown 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** pulldown 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** pulldown 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 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 ? * *)

? Unknown Attachment



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



If you want to disable the content source just unselect 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 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 OneDrive.



1. Enter the "Tenant Domain" of the OneDrive you want to crawl.
 - a. Example: "<yourdomain>.onmicrosoft.com"
2. Enter the "Application ID" that you generate as [prerequisite \(Create the Azure AD Application\)](#). It will be automatically encrypted by Aspire.
3. Enter the "Application Secret" that you generate as [prerequisite \(Create Application Key\)](#). It will be automatically encrypted by Aspire.
4. Enter the "Graph URL"
 - a. Example "https://graph.microsoft.com/v1.0"
5. Enter the "OneDrive Users", this could be a list of usernames (emails), or a file with the file with the username list, one by line.
6. Check on the other options as needed:
 - a. Index containers?: index subfolders as items. If unchecked, only files will be indexed.
 - b. Scan recursively?: Scan through subfolder's child nodes.
 - c. Include/Exclude patterns: Enter regex patterns to include or exclude files/folders based on URL matches.

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. This 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.
 - a. Specify a *Name* and *Description* for the Publisher.
 - b. Click *Add*.

After completing this steps click on the **Save** then **Done** and 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" (is set as "Incremental" by default and the first time it'll work like a full crawl. After the first crawl, set it to "Incremental" to crawl for any changes done in the repository).
2. Click on the Start button.

During the Crawl

During the crawl, you can do the following:

- Click on the "Refresh" button 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 on "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 OneDrive (documents which are added, modified, or removed), then click on the "Incremental" button instead of the "Full" button. The OneDrive 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 the "Incremental" button 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, the Incremental button will only crawl updates.



Statistics are reset for every crawl.