Content Management

In this section, we will review new content management features and how they can help you to manage your set of content sources.

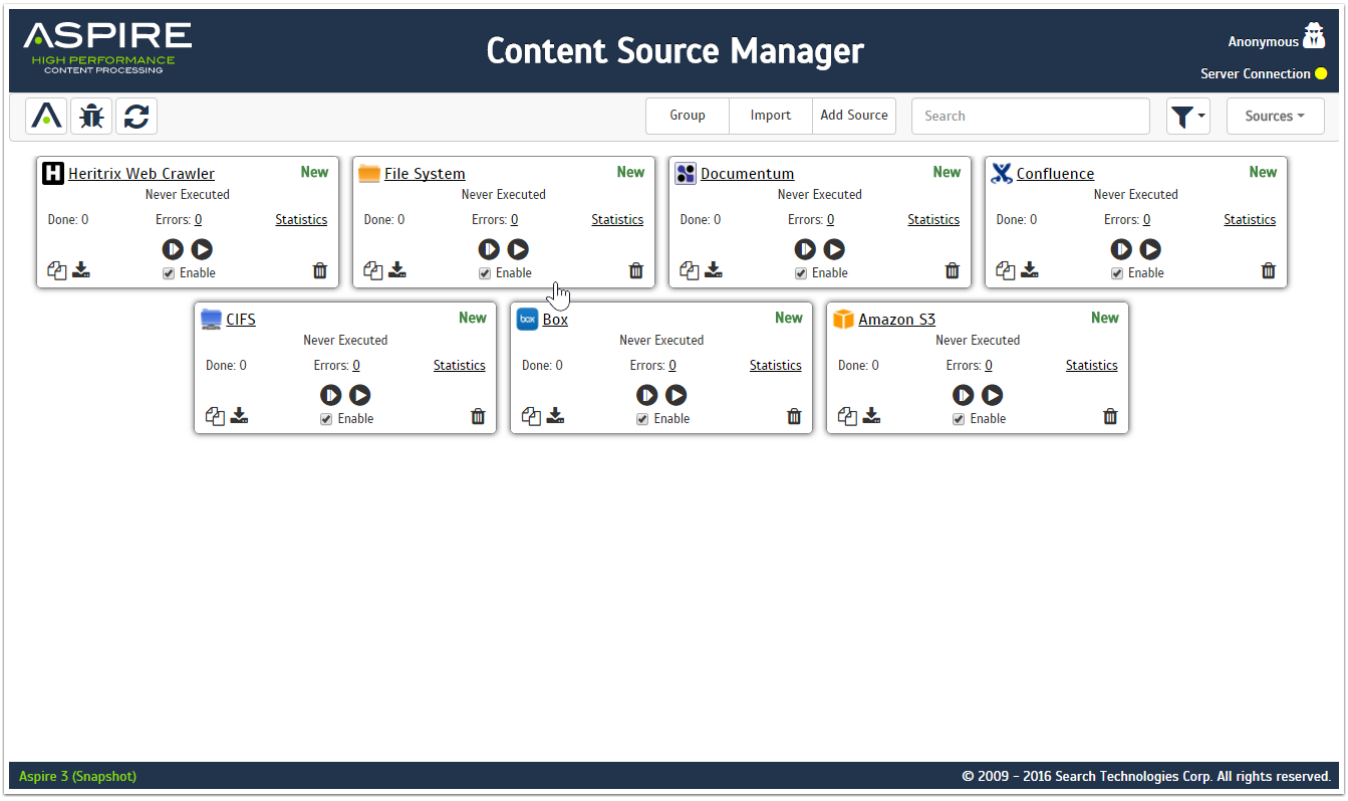
On this page:

- Content Sources
- Controls
- States
- Grouping
- Filtering
- Import
- Add Source

Content Sources

All of the management of Content Sources can be done in the Admin UI Content Sources, where you can access all of the management features.

- You can see all the content sources in their new presentation.
- Each content source card displays further detailed information.
- You can identify which is the content source you want including its current state, how many jobs it has crawled, and how many errors were produced.
- You can access the controls for copy, delete, crawl control, activate, and deactivate.



Controls

For the display of content sources Aspire adds new features and information into card like objects that represents each content source.

? Unknown Attachment

The cards contain several functions to control the content source.

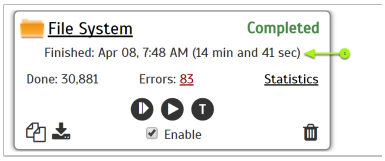
1. **Content Source Name & Icon**
    - The display name of the content source, if the name is to long the name will appear follow by an ellipsis (e.g. ABC...), and if the mouse is over the name, a tooltip will appear with the full name. The Content Source Name is also a link to the configuration of the respective content source.
2. **Content Source Status**
    - Indicates the current status of the content source. When the status changes not only does the label change, the color of the content source card changes.
3. **Time of Crawl**
    - Indicates the length of time the crawl has been crawling, or how long it took the crawl to finish. In case the content source has never being started the label will be **Never Executed**. If you put the mouse over the Time of Crawl control, it will show a tool-tip in your local timezone, of the exact date and time the content source was started.
4. **Jobs Done**
    - Completed jobs shows the total number of successful jobs, you can also click on Complete and open the statistics.
5. **Errors**
    - Errors will show the number of document errors if any; if there is at least one error, you can click on the number and go to the Error Page.
6. **Statistics**
    - It will show all the information available about the current crawl, including crawl type, documents per second, start time, jobs status, etc.
7. **Start Full Crawl**
    - Starts a full crawl, but before that it will show a warning indicating all incremental indexing data will be deleted.
8. **Start Incremental Crawl**
    - Starts an incremental crawl.
9. **Start Test Crawl**
    - Starts a Test Crawl that will ask how many documents to crawl and how many to skip before it starts to crawl.
10. **Copy**
    - Creates a new content source with the same configuration.
11. **Export**
    - Downloads a zip file with the configuration of the content source that you can use to import it into a different aspire instance.
12. **Enable/Disable**
    - It will enable or disable the content source. *(you can only disable the content source if it's status is new, completed, error, failed or canceled)*
13. **Delete**
    - If you click the delete button a confirmation will appear, if you click OK the delete will proceed and the content source will be deleted.

States

The content sources can be in one of several states. In each state, some controls change and some are disabled. In the section above, we saw the controls of a content source. In this section we will see which are the controls for each state.

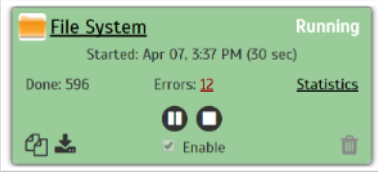
New / Completed / Cancelled

The *Completed / Cancelled* status indicates that the crawl or stopping the content source was successful, for these states the Time of Crawl ***(1)*** changes to the exact time in your local time zone when the crawl was started and how long it took to reach this state.



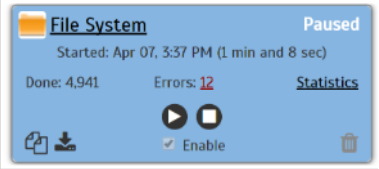
Running

The *Running* status indicates that a crawl is currently in progress, for this state the content source will change color to green, the Time of Crawl changes to the exact time in your local time zone when the crawl was started and with each refresh it will increase the total time that it's been crawling. The Jobs Done number will start to increase and the Errors will show the number of errors at that moment, if any. The Pause and the Stop button will replace the start crawls buttons.



Paused

The *Paused* status indicates that a crawl is currently paused, for this state the content source will change color to blue, the Time of Crawl will still increase the total time that it's been crawling. The Resume button will replace the Pause button.



Error / Failed / Aborted

The *Error* status indicates that a crawl finished in an unsuccessful crawl, for these states the content source will change color to red *(The Content Source Status will also be a link to see the cause of the unsuccessful crawl)*, the Time of Crawl will stop updating the total crawl time. The start crawls Buttons will be set again.

? Unknown Attachment

> The Failed status appears when the content source fails in the initialize phase. The Content Source Status will also be a link to see the reason of the failure

> The Aborted status appears when the content source was aborted by the user

Pausing / Stopping / Resuming

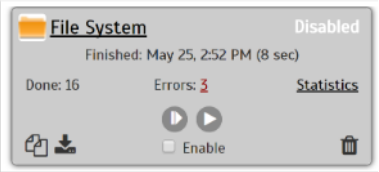
This state indicates a change from one state to another, and they are the only ones in which the user can do an abort.

For these states the content source will change color to yellow, The Abort button will be set

- The Pausing status indicates that a crawl is currently trying to pause the content source
- The Stopping status indicates that a crawl is currently trying to stop the content source
- The Resuming status indicates that a crawl is currently trying to start the crawl again

Disabled

The *Disabled* status indicates that a crawl is currently disabled and it will not perform a crawl, for this state the content source will change color to gray, The start crawls buttons will be disabled.



Grouping

A group has the same shape as a content source (card-like), but its content is different. In the image below, we see all of the controls that a group will have.

? Unknown Attachment

1. **Group Name**

- The display name of the group, if the name is to long the name will appear follow by an ellipsis (e.g. ABC...), and if the mouse is over the name, a tool-tip will appear with the full name. The Group Name is also a link that will expand the content source, so we can see the content source inside it.

2. **Number of Content Sources**
    - The group will display the number of content sources that has contained.
3. **Expand**
    - If clicked, it will expand the content source, so we can see the content source inside it.
4. **Add to Group**
    - If clicked, Add to Group will open the group menu, and put the name of the group so we can only choose the content sources and click on Add Group.
5. **Content Sources Status**
    - Has well as the number of content sources the group will have the status of the content sources and how many content sources inside him has that specific status
        a. Green: Stands for the *Running* status.
        b. Blue: Stands for the *Paused* status.
        c. Red: Stands for the three unsuccessful status *Error*, *Failed* and *Aborted*.
        d. Orange: Stands for the three transitory status *Pausing*, *Resuming* and *Stopping*.
        e. White: Stands for the idle status *New* and *Completed* and *Cancelled*.
        f. Gray: Stands for the *Disabled* status.
6. **Ungroup**
    - If clicked, Ungroup will take all the content sources inside the group and put them in the first level (root), and it will delete the group.

Manage a Group

This section walks through the steps necessary to create, use and dispose of a group, and how to interact with the group itself.

Step 1: Select the content sources

Click on the *Group* button in the Action Bar, a text field will appear, and the bottom of all the content sources will change into a check box that says **Group**. Select the content sources you want to group together by checking the **Group** check box, then put the name of the group in the text field next to the *Group* button.

? Unknown Attachment

- Once the group is created, if you want to add another content source to the group, you can do it by click the **Add One** button and repeating steps 1 and 2
- If you want to cancel the creation of a group by clicking the *X* in the text field

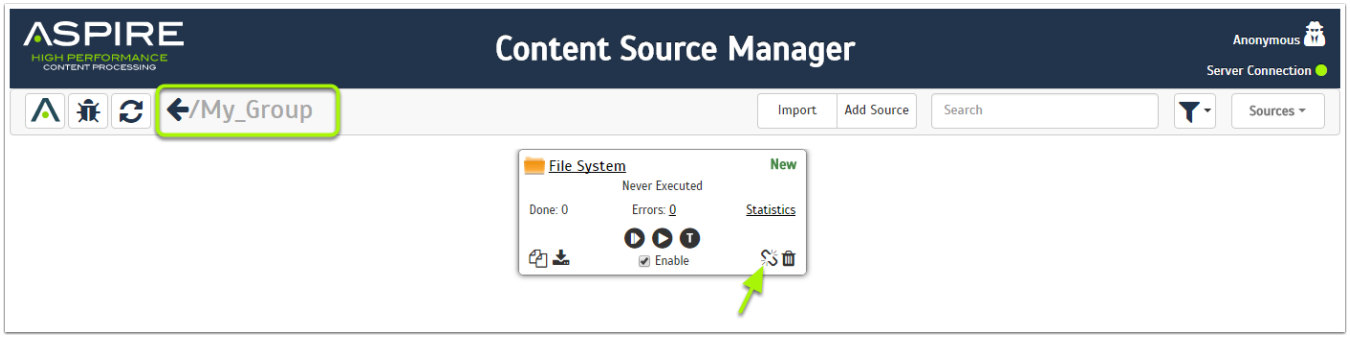
Step 2: Create the group

Once you have selected the content sources and filled the name, you can click on *Add Group*, this will fade out and in the content source, and a group card will appear at the end of the content sources. This will be the group you just create with all the selected content sources inside it.

Step 3: Expand the group

Once we have the group created we can expand it by clicking on the *Group Name* or by clicking on the expand button , this will fade out and in the content sources and display only the content sources inside the group. Also a legend will appear in the *Action Bar* indicating in which group we are in now, right next to the legend is the *Return* turn button , if we click on it, it will return us to the first level (root).

Also while we are in the expanded group we can see that the content source has another button right before the copy button, this button is the *Ungroup One* button , it looks exactly as the *Return* button , this button will remove the content source from the group and put it on the first level (root)



Step 4: Remove the group

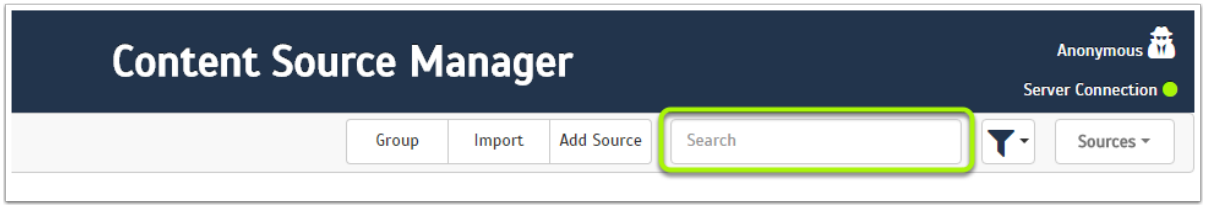
If you want to remove a group, the only way is to ungroup the entire group by clicking the *Ungroup* button , this will move all the content sources from our group to the first level (root) and delete the group.

Filtering

In Admin UI, you can filter the content sources, and we can apply several filters for a more accurate result.

Text Filter (Search)

You can do regex searches base on the *Content Source Name* and the *Group Name*. To make a search we just put the regex we want in the text field on the top right corner of the *Action Bar* and press the *Enter* key. Only the content sources that match the regex will be displays



Cookie Filters

The cookie filters are regular filters that will be saved in a cookie so once we apply a cookie filter, this will be active until we remove it. We can access the cookie filters by clicking on the *Filter* button ,

We have 3 categories of cookie filters, all of then explained below.

? Unknown Attachment

General filters

We have 3 general filters:

1. **Active:** If checked shows all the content sources that are active *(Checked by default)*
2. **Groups:** If checked shows all the groups. *(Checked by default)*
3. **Inactive:** If checked shows all the content sources that are inactive *(Checked by default)*

Status filters

The status filters has all the possible status for the content sources, also it will have the **All** filter *(Checked by default)*, if any of the status filters is checked, it will unchecked the **All** filter and only the content sources matching the checked status will be displayed. If the **All** filter is checked again, it will unchecked all the status filters.

? Unknown Attachment

Connector filters

The connector filters will be build according to the types of content sources available for the aspire account, but it will always have the **All** filter *(Checked by default)*, if any of the connector filters is checked, it will unchecked the **All** filter and only the content sources matching the checked connector types will be displayed. If the **All** filter is checked again, it will unchecked all the connector filters.

? Unknown Attachment

Time filters

The time filters can be applied for start and end time of the crawl.

- **Start Time:** *(Unchecked by default)* Compares the time given with the start time of the content source, if the time given is after or the exact time of the content source, the content source will be display, if we check the *Start Time* filter but we don't give a start time, the filter won't be applied. Also if the content source doesn't have a start time, the content source won't be displayed.
    - To set the date and time of the filter click on the calendar button **(2)**
    - To enable the filter check the filter **(1)**

- **End Time:** *(Unchecked by default)* Compares the time given with the end time of the content source, if the time given is before or the exact time of the content source, the content source will be display, if we check the *End Time* filter but we don't give a start time, the filter won't be applied. Also if the content source doesn't have a end time, the content source won't be displayed.
    - To set the date and time of the filter click on the calendar button **(2)**
    - To enable the filter check the filter **(1)**

? Unknown Attachment

Import

In Admin UI, we can import a content source zip file and load the content source directly to our management page.

To import a content source do the following steps:

? Unknown Attachment

1. Click on **Import** in the Action Bar
2. Use the browse window to find and select the zip file
3. Click on Open to import

Once you have clicked on open, the content source will appear in the screen as loading with a warning icon **(1)**, which means the source icon is missing.

? Unknown Attachment

After it has loaded the content source will appear as New, and the source icon **(1)** will show up

? Unknown Attachment

The import can only be successful if the zip file contains all 4 necessary files, and this files are correctly formatted.

Add Source

With *Add Source*, select the type of content source you want by choosing the connector. The *Add Source* menu has three main sections.

? Unknown Attachment

1. **Add Source**
    - Access the Source Menu from the *Add Source* button
2. **Legacy Connectors**
    - These connector will be identified with a LEGACY label, which means these connector haven't been updated to use the new connector framework.
3. **New Connectors**
    - These connectors have been updated to use the new connector framework.

4. . **Artifact Id**
    - Indicates the maven Artifact Id of the connector.
5. **Legacy Label**
    - The Legacy label will indicate which connectors are not updated to use the new connector framework.
6. **Custom**
    - It will open a menu to install a custom connector from maven coordinates or a set of config files.
7. **Refresh Sources**:
    - With the *Refresh* button we can update the list of the connectors available for us.

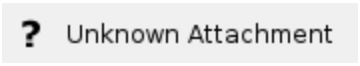
The Official connector may change according to your connector entitlements.

Custom Connector

Click **Custom** to open a window where you can choose between two methods to install a custom connector repository and configuration files. Both show as toggle buttons at the top of the window.

Repository

The repository method is always the default, With this option, you can download the custom connector from a Maven repository. To install the custom connector, fill the following fields.



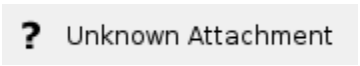
1. **Repository Tab**
    a. Indicates the method currently being use to add a new connector
2. **Group Id**
    - The groupId of the maven artifact
    - e.g. *com.searchtechnologies.aspire*
3. **Artifact Id**
    - The artifactId of the maven artifact representing the connector
    - e.g. *app-custom-connector*
4. **Version***(Optional)*
    - If the version of the artifact isn't specify, Aspire will use the same version as it.
5. **OK Button**
    - Click to load the connector. This may take a few seconds.
6. **Cancel**
    - You can close the window by either clicking on the *X* o r clicking on *Cancel*

All the connectors added using this method will be added to the **Add Source** menu.

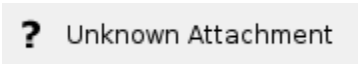
It is not recommended to use an older version of a connector is a new version is available.

Configuration Files

Before accessing the configuration file method, an alert will indicate that the connectors added using this method will not be included in the *Add Source* menu.



The configuration files method requires both the application file and DXF file in the Aspire server. To install a custom connector using this method, specify the direction of the application file.



1. **FileTab**
    a. Indicates the method currently being use to add a new connector
2. **File Path**
    - Path to the xml File
    - e.g. *config/application.xml*
3. **OK Button**

- Click to load the connector. This may take a few seconds.

4. **Cancel**
   - You can close the window by either clicking on the *X* o r clicking on *Cancel*

- The dxf file must be call as the application xml file with the *dxf* suffix *(e.g. application-dxf.xml)*
- And it must be in the same folder as the *application.xml*

If the dxf file doesn't have the new valid format for connectors, it won't be possible to configure the connector.