

How to Configure Parquet Extractor

On this page

- [Step 1. Launch Aspire and Open the Content Source Management Page](#)
- [Step 2. Add a new Content Source](#)
- [Step 3. Add a new Parquet Extractor to the Workflow](#)
 - [Step 3a. Specify Application Information](#)
 - [Step 3b. Share rule to a library](#)
 - [Step 3c. Copy the rule from the shared library](#)

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, please refer to [Admin UI](#)

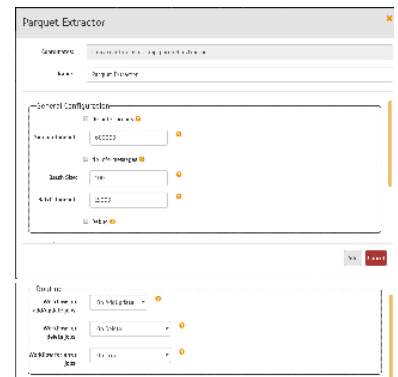


Step 2. Add a new Content Source

- For this step please follow the step from the Configuration Tutorial of the connector of you choice, please refer to [Connector list](#)

Step 3. Add a new Parquet Extractor to the Workflow

To add a Parquet Extractor application drag from the **Parquet Extractor** rule from the *Workflow Library* and drop to the *Workflow Tree* where you want to add it. This will automatically open the Parquet Extractor window for the configuration of the application.



Step 3a. Specify Application Information

In the Parquet Extractor window, specify the configuration information of the application.

1. General configuration
 - a. Do not store ids - Check if you do not want to store ids in NoSql. Warning! Aspire will not be able to delete items in the index in future crawls.
 - b. Sub Job timeout - Time out for secondary jobs in milliseconds..
 - i. **Example:** 60000
 - c. No info messages - Check if you want info messages disabled
 - d. Batch Size - Sub job batch size.
 - e. Batch Timeout - Sub job batch timeout in milliseconds.
 - f. Debug:
 - i. Enable debug messages.
2. Routing
 - a. Workflow for add/update jobs:
 - i. Workflow to send the generated add or update jobs.
 - b. Workflow for delete jobs:
 - i. Workflow to send the generated delete jobs.
 - c. Workflow for error jobs:
 - i. Workflow to send the generated error jobs.

Step 3b. Share rule to a library

Once the application is set, it must be shared to a (new or existing) library.

Note: This is a required step.

Step 3c. Copy the rule from the shared library

Add the shared application from the library to the Delete workflow.

Note: This is a required step.



In order to work, the application requires Extract Text to be disabled in the connector configuration.

Extract Text

Enabled: ☐ ?

Once you've clicked on the *Add* button, it will take a moment for Aspire to download all of the necessary components (the Jar files) from the Maven repository and load them into Aspire. Once that's done, the application will appear in the Workflow Tree.



For details on using the Workflow section, please refer to [Workflow](#) introduction.