

Publish to Avro File How to Configure

On this page

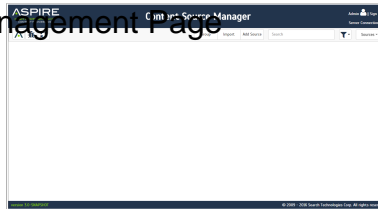
- [Step 1. Launch Aspire](#)
- [Step 2. Open the Content Source Management Page](#)
- [Step 2. Add a new Content Source](#)
- [Step 3. Add a New Publish to Avro File to the Workflow](#)
 - [Step 3a. Specify Publisher Information](#)

Step 1. Launch Aspire

Step 2. 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 Content Source

Follow the step from the Configuration Tutorial of the connector of your choice. Please refer to the [Connector list](#).

Step 3. Add a New Publish to Avro File to the Workflow

Add a Publish to Avro File, drag from the **Publish to Avro File** rule from the *Workflow Library* and drop to the *Workflow Tree* where you want it.

This will automatically open the Publish to Avro File window for the configuration of the publisher.

Step 3a. Specify Publisher Information

A screenshot of the 'Aspire Publish To Avro Application' configuration window. It has a dark blue header with the title and a close button. The form contains several fields: 'Connections' (a dropdown menu), 'System Name' (a text field with 'Aspire_Publish_To_Avro_Application'), 'Name' (a text field with 'Aspire Publish To Avro Application'), 'Specify the output type' (a dropdown menu), 'Output directory path' (a text field with 'S:\aspire\hadoop\log\log\app\name'), 'Batch Size' (a numeric input field with '1000'), 'Directory/folder to move file into' (a text field), 'Time rollover threshold' (a numeric input field with '5'), and a 'Debug' checkbox. At the bottom right are 'Update' and 'Cancel' buttons.

In the **Publish to Avro File** window, specify connection information to publish to the Avro File.

1. Enter the name of the publisher. (*This name must be unique*).
2. Enter the description of the publisher that will be shown in the Workflow Tree.
3. Enter the Avro Schema path
4. Output type - Size limited file
 - a. Enter the output file path
 - b. Enter the maximum size of the input file
5. Output type - Internal batches
 - a. Enter the output directory path for Avro files
 - b. Enter the batch size: the number of Avro records per the batch file
6. Time rollover threshold - Time rollover threshold in seconds
7. Directory/folder to move file into - Enter the directory path where the Avro files will be moved

After you click **Add**, it will take a moment for Aspire to download all of the necessary components (Jar files) from the Maven repository and load them into Aspire.

After that's done, the publisher will appear in the Workflow Tree.