

Publish to Web HDFS How to Configure

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 Publish to HDFS to the Workflow](#)
 - [Step 3a. Specify publisher information](#)
- [3.a. Publish using Web HDFS](#)

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

? Unknown Attachment

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

Step 3. Add a new Publish to HDFS to the Workflow

? Unknown Attachment

To add a Publish to HDFS, drag from the **Publish to HDFS** rule from the *Workflow Library* and drop to the *Workflow Tree* where you want to add it. This will automatically open the Publish to HDFS window for the configuration of the publisher.

Step 3a. Specify publisher information

In the Publish to HDFS window, specify the connection information to publish to the HTTP.

- Enter the name of the publisher. (*This name must be unique*).
- Enter the description of the publisher that will be shown in the Workflow Tree.



Not all HDFS clusters have WebHDFS enabled.

3.a. Publish using Web HDFS

In the **HDFS** section of the *Publish to Web HDFS* window specify the connection information to publish to HDFS.

- Enter the *HDFS URL*. Use **hdfs://** protocol and the port (by default between 50070-50075). I.e. <http://name-node-server:50070/webhdfs/v1/>
- Specify the absolute *HDFS Folder Path* where the files will be published to. I.e. `/user/jsmith/my_aspire_output`. (The user which runs Aspire must have write access to the HDFS folder).
- Specify the Delegation Token it will contain the credentials to publish into HDFS
- Debug: Check if you want to run the publisher in debug mode.
- Click on the *Add* button.

Custom Application

Name:

WebHDFS URL:

e.g. http://name-node-server:50070/webhdfs/v1/

Field required.

HDFS Folder Path:

e.g. /user/myuser/folder

Field required.

Delegation Token:

please insert the delegation token

Field required.

☐ Debug

Add

Back

Cancel

Once you've clicked **Add**, 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 publisher will appear in the Workflow Tree.

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