

## Tutorial: Customize a ReadyFlow

Date published: 2021-04-06

Date modified: 2025-09-30



# Legal Notice

© Cloudera Inc. 2025. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

# Contents

**Tutorial: customizing a ReadyFlow.....4**

- Open the Hello World ReadyFlow as a template for your draft..... 4
- Create new services..... 5
- Swap the controller services in your draft..... 7
- Start a test session..... 9
- Publish your flow definition to the Catalog..... 12

## Tutorial: customizing a ReadyFlow

Learn how to create a draft using a ReadyFlow as a template.

### About this task

This tutorial shows you how you can open a ReadyFlow from the ReadyFlow Gallery and create a customized flow design using the ReadyFlow as a template. This tutorial uses the Hello World ReadyFlow, a simple flow design that retrieves the latest changes from Wikipedia through invoking the Wikipedia API. The flow converts JSON events to Avro, then filters and routes the events to two different processors which merge events together, and finally a file is written to local disk.

You will learn about the following actions:

- Opening a ReadyFlow as a draft on the Flow Design Canvas
- Creating a Controller Service
- Changing the configuration of processors
- Running a Test Session
- Publishing a draft to the Catalog as a flow definition

### Before you begin

The Hello World ReadyFlow that you are about to customize can be deployed without any external dependencies and does not require any parameter values during deployment. Still, you must meet the following prerequisites before you can start building your first draft:

- You must have an enabled and healthy Cloudera Data Flow environment.
- You must be assigned the DFDeveloper role granting you access to the Flow Designer.
- You must be assigned the DFCatalogAdmin or DFCatalogViewer role granting you access to the Catalog. You will need this authorization to publish your draft as a flow definition to the Catalog.
- You must be assigned the DFFlowAdmin role for the environment where you want to deploy the flow definition.

## Open the Hello World ReadyFlow as a template for your draft

Learn how to create the Hello World ReadyFlow as a starting point for your custom flow in the Flow Designer.

### About this task


ReadyFlows are read-only, therefore you cannot publish a new version of them. When you are done with creating your customized version of the given ReadyFlow, you can only publish your work to the Catalog as a new flow definition.

### Before you begin

- You must have an enabled and healthy Cloudera Data Flow environment.
- You must be assigned the DFDeveloper role granting you access to the Flow Designer.
- You must be assigned the DFCatalogAdmin or DFCatalogViewer role granting you access to the Catalog. You will need this authorization to publish your draft flow as a flow definition to the Catalog.
- You must be assigned the DFFlowAdmin role for the environment to which you want to deploy the flow definition.

### Procedure

1. Open Cloudera Data Flow by clicking the DataFlow tile in the Cloudera sidebar.

2. Select the  ReadyFlow Gallery item in the left navigation pane.
3. Select the Hello World ReadyFlow.
4. Click the Create New Draft button.
5. Select the target Workspace where you want to create the draft flow.
6. Provide a valid name in the Draft Name field for the draft flow.  
Flow names must be unique within a workspace. If a draft with the provided name already exists, you must provide a different name.
7. Click the Create button.  
Hello World opens as a draft in the designated **Flow Designer** workspace with the Draft Name you provided on the **Flow Design** canvas.

### What to do next

Proceed with creating the necessary Controller Services.

### Related Tasks

[Create new services](#)

## Create new services

Learn about creating Controller Services in Cloudera Data Flow **Flow Designer**.

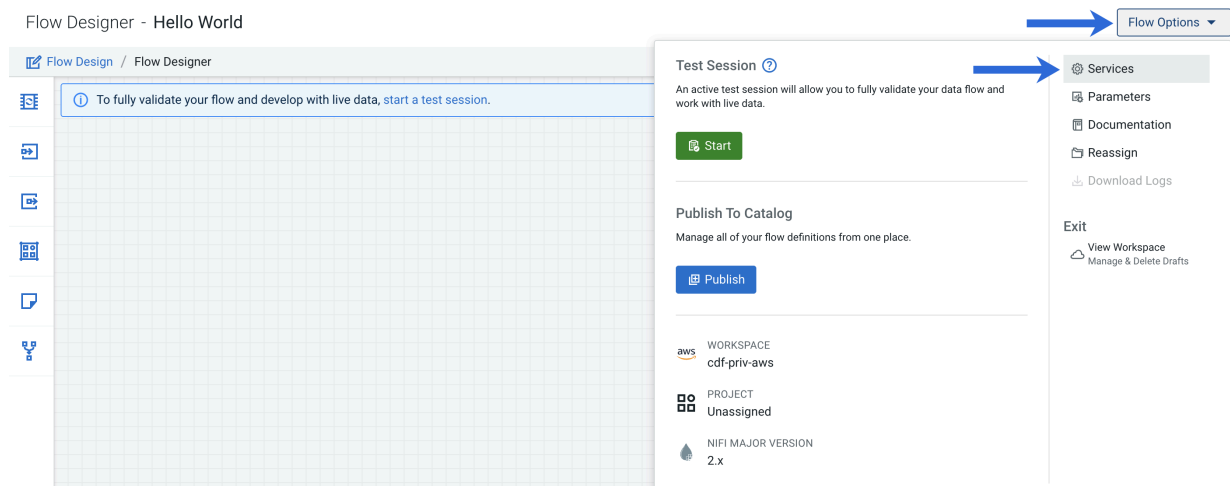
### About this task

Services are extension points that provide information for use by other components, such as processors or other controller services. The idea is that, rather than configuring this information in every processor that might need it, the service provides it for any processor to use as needed.

You will use the services you create now to configure the behavior of several processors you will add to your flow as you are building it.

### Procedure

1. Go to Flow Options Services .



- Click the Add Service button.  
The Add Service page opens.

### Add Service ✕

- AccumuloService >
- ActionHandlerLookup
- ADLSCredentialsControllerService
- ADLSIDBrokerCloudCredentialsProviderCont...
- AlertHandler
- AvroReader
- AvroRecordSetWriter
- AvroSchemaRegistry
- AWSCredentialsProviderControllerService
- AWSIDBrokerCloudCredentialsProviderContr...
- AzureBlobIDBrokerCloudCredentialsProvider...
- AzureCosmosDBClientService
- AzureEventHubRecordSink
- AzureStorageCredentialsControllerService
- AzureStorageCredentialsControllerService\_v...

Service Name

**Type**

**AccumuloService**

IMPLEMENTS SERVICE

BaseAccumuloService 1.18.0.2.3.7.0-64 from nifi-accumulo-services-api-nar

VERSION	GROUP
1.18.0.2.3.7.0-64	org.apache.nifi

BUNDLE

nifi-accumulo-services-nar

DESCRIPTION

A controller service for accessing an Accumulo Client.

TAGS

accumulo, service, client

- In the Search field, filter for CSVReader.
- Provide the Service Name as CSVReader\_Recent\_Changes.
- Click the Add button.  
You do not need to configure the CSVReader\_Recent\_Changes service. You can leave all properties with their default values.
- Click the Add Service button again to create another service.
- In the Search field, filter for CSVRecordSetWriter.
- Provide the Service Name as CSVRecordSetWriter\_Recent\_Changes.
- Click the Add button.  
You do not need to configure the CSVRecordSetWriter\_Recent\_Changes service. You can leave all properties with their default values.
- Select the AvroReader\_Recent\_Changes service and check the list of Referencing Components in the Service Details pane on the right.  
Make a note of the listed components, because that is where you will need to replace AvroReader\_Recent\_Changes with the CSVReader\_Recent\_Changes service.

11. Select the `AvroWriter_Recent_Changes` service and check the list of Referencing Components in the Service Details pane on the right.

Make a note of the listed components, because that is where you will need to replace `AvroWriter_Recent_Changes` with the `CSVRecordSetWriter_Recent_Changes` service.

State	Name ↑	Type	Notifications
🔍	AvroReader_Recent_Changes	AvroReader V.1.20.0.2.3.8.0-21	>
🔍	AvroWriter_Recent_Changes	AvroRecordSetWriter V.1.20.0.2.3.8.0-21	>
🔍	CSVReader_Recent_Changes	CSVReader V.1.20.0.2.3.8.0-21	>
🔍	CSVRecordSetWriter_Recent_Changes	CSVRecordSetWriter V.1.20.0.2.3.8.0-21	>
🔍	JSON_Reader_Recent_Changes	JsonTreeReader V.1.20.0.2.3.8.0-21	>

Properties

Property	Value
Schema Access Strategy ⓘ	Use Embedded Avro Schema ⋮
Cache Size ⓘ	1000 ⋮

Referencing Components

State	Referencing Processor	Path
⊗	<a href="#">Filter Edits</a> QueryRecord V.1.20.0.2.3.8.0-21	<a href="#">🔗</a>
⊗	<a href="#">Merge Edit Events</a> MergeRecord V.1.20.0.2.3.8.0-21	<a href="#">🔗</a>
⊗	<a href="#">Merge Edit Events</a> MergeRecord V.1.20.0.2.3.8.0-21	<a href="#">🔗</a>
⊗	<a href="#">Route On Content Size</a> QueryRecord V.1.20.0.2.3.8.0-21	<a href="#">🔗</a>

12. Click the Back To Flow Designer to return to the flow design Canvas.

What to do next

After creating the necessary services, you can start customizing your flow.

Related Tasks

[Swap the controller services in your draft](#)


Swap the controller services in your draft

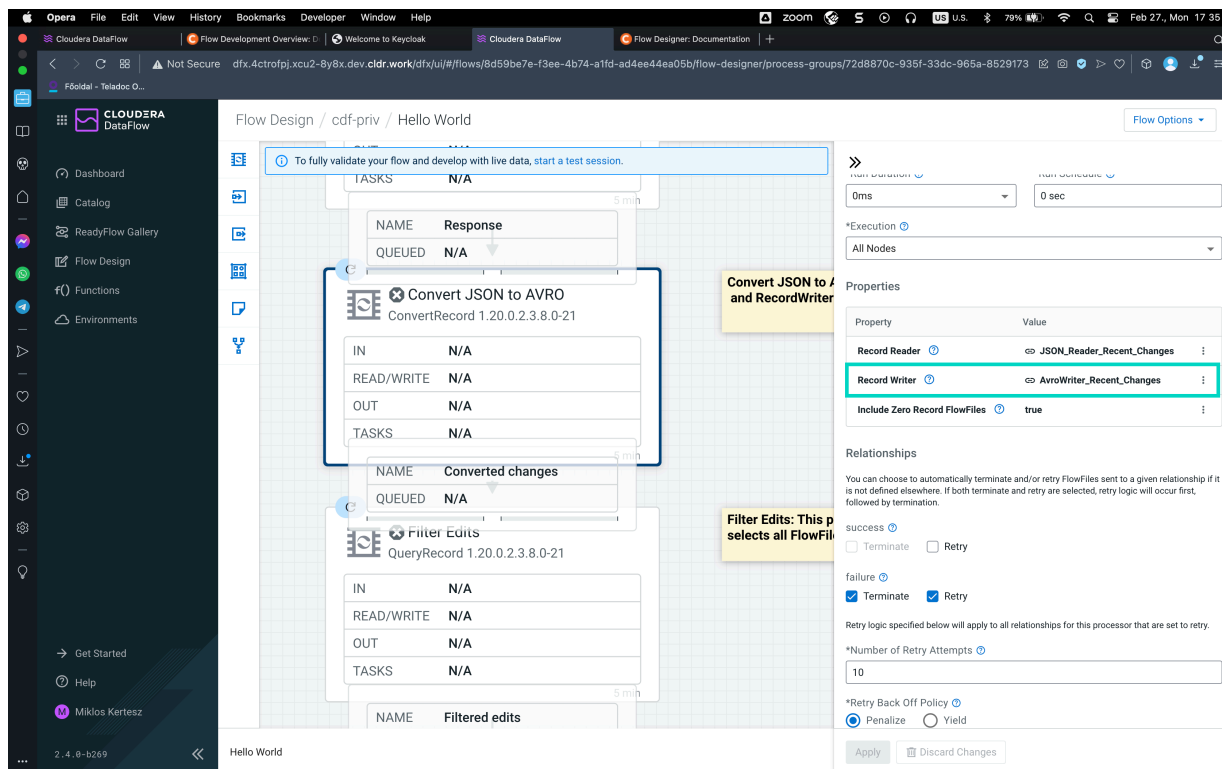
Customize your draft by replacing services to change the destination file format.

Procedure

- 1. On the **Flow Designer** canvas select the Convert JSON TO AVRO processor by clicking on it.

2.

In the Configuration pane scroll down to the Properties section, click the  icon next to the Record Writer property and select the CSVRecordSetWriter\_Recent\_Changes service from the dropdown list.



The screenshot shows the Cloudera DataFlow Flow Designer interface. The canvas displays a flow with several processors. The 'Convert JSON to AVRO' processor is highlighted with a blue box. The 'Filter Edits' processor is also visible. The configuration pane on the right is open to the 'Properties' section for the 'Record Writer' property. The 'Record Writer' dropdown is highlighted, showing 'AvroWriter\_Recent\_Changes' selected. The 'Include Zero Record FlowFiles' property is set to 'true'. The 'Filter Edits' processor is also visible on the canvas.


3. Click the OK button.

4. In the Settings section, rename the processor from Convert JSON to AVRO to Convert JSON to CSV.

5. Click the Apply button.


6. On the **Flow Designer** canvas select the Filter Edits processor by clicking on it.

7.

In the Configuration pane scroll down to the Properties option, click the  icon next to the Record Reader property and select the CSVReader\_Recent\_Changes service from the dropdown list.

8. Click the OK button.

9.

Click the  icon next to Record Writer property and select the CSVRecordSetWriter\_Recent\_Changes service from the dropdown list.

10. Click the OK button.

11. Click the Apply button.

12. Proceed with updating the remaining Record Reader and Record Writer references.

You must modify the configurations of the Route on Content Size and the two Merge Edit Events processors.



13. Check that you have updated all the processors.

- a) Go to Flow Options Services .
- b) Select the CSVReader\_Recent\_Changes service and check the Referencing Components section.

Check if you see the following components:

- Filter Edits
- Merge Edit Events
- Merge Edit Events
- Route On Content Size

- c) Select the CSVRecordSetWriter\_Recent\_Changes service and check the Referencing Components section.

Check if you see the following components:

- Convert JSON to CSV
- Filter Edits
- Merge Edit Events
- Merge Edit Events
- Route On Content Size

- d) If you check the AvroReader\_Recent\_Changes and AvroWriter\_Recent\_Changes services, you are expected to see the No referencing Processors to display. text.

14. Delete the AvroReader\_Recent\_Changes and AvroWriter\_Recent\_Changes services by selecting the service and

clicking the  Delete icon.

In the confirmation pop-up click the Delete button.

15. Click Back To Flow Designer to return to the flow design Canvas.

### What to do next

Start a Test Session to validate your draft.

## Start a test session

To validate your draft, start a test session. This provisions an Apache NiFi cluster where you can test your draft.

### About this task

Starting a Test Session provisions NiFi resources, acting like a development sandbox for a particular draft. It allows you to work with live data to validate your data flow logic while updating your draft. You can suspend a test session any time and change the configuration of the NiFi cluster then resume testing with the updated configuration.

### Procedure

1. Click the start a test session link in the banner on top of the Canvas.

 To fully validate your flow and develop with live data, start a test session.


- Review the suggested NiFi configuration. You do not need to change anything.


Test Session - cdev\_test\_2

[Flow Design](#) / [Flow Designer](#) / Test Session

## NiFi Configuration

NiFi Runtime Version [Change Version](#)

 CURRENT VERSION  
 Latest Version (2.3.0.4.3.2.0-89)

 Review the [Cloudera Data Flow and Cloudera Runtime support matrix](#) to ensure the selected NiFi Runtime Version is compatible.

Inbound Connections

☐ Allow NiFi to receive data [?](#)




Custom NAR Configuration


☐ This flow deployment uses custom NARs [?](#)

Custom Python Configuration

☐ This flow deployment uses custom python processors [?](#)

NiFi Node Sizing [?](#)

	vCores Per Node	RAM Per Node
<input checked="" type="radio"/>  Extra Small	2	4 GB
<input type="radio"/>  Small	3	6 GB
<input type="radio"/>  Medium	6	12 GB

 Start Test Session [Cancel](#)

- Click the  Start Test Session button.

Test Session status  Initializing Test Session... appears on top of the page.

- Wait for the status to change to

Active Test Session 

This can take several minutes.

- Go to [Flow Options Services](#) to enable Controller Services.

6.




Select a service you want to enable, then click the  Enable Service and Referencing Components icon.

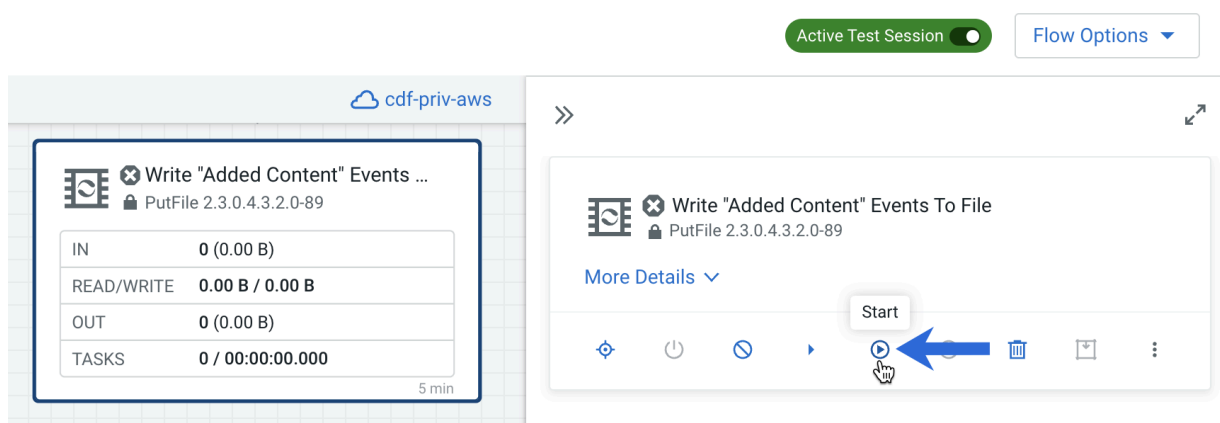
This option does not only enable the controller service, but also any component that references it. This way, you do not need to enable the component separately to run the test session. In the context of this tutorial, enabling the AvroReader\_Recent\_Changes controller service will also enable the Filter Edits, Route on Content Size, and Merge Edit Events processors as well.


Repeat this step for all Controller Services.

7. Click the Flow Designer link in the breadcrumbs on the top of the page to return to the Flow Design canvas.

8. Start the remaining inactive processors.

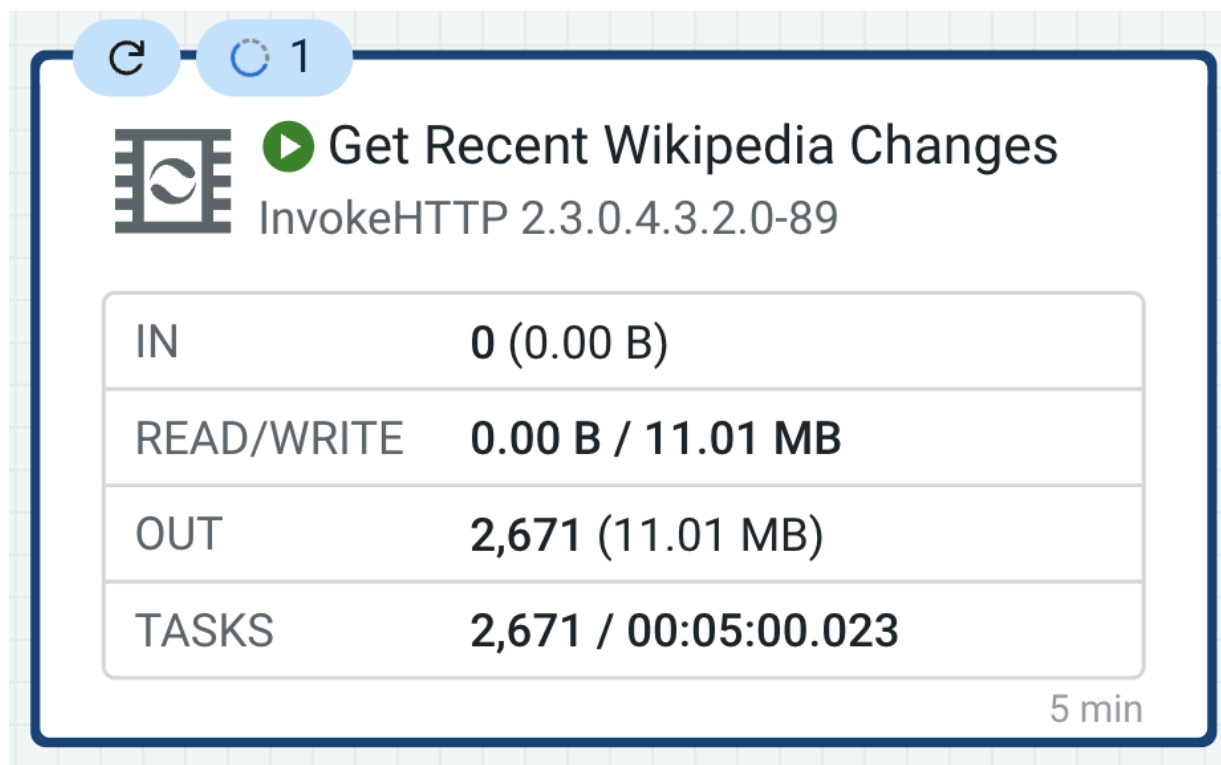
- a) Start the Get Recent Wikipedia Changes processor by selecting it on the canvas, then clicking the  [Start] icon in the component details pane.
- b) Start the Write "Added Content" Events To File processor by selecting it on the canvas, then clicking the  [Start] icon in the component details pane.
- c) Start the Write "Removed Content" Events To File processor by selecting it on the canvas, then clicking the  [Start] icon in the component details pane.



All other components were auto-started when you selected the Enable Service and Referencing Components option in the Flow Options  Services view.

9. Observe your first draft flow processing data.

On the **Flow Designer** canvas you can observe statistics on your processors change as they consume and process data from Wikipedia. You can also observe one or more blue Notification Pills, providing information about the current task.



## Publish your flow definition to the Catalog

Now that you have tested your draft and it works fine, you can go on and publish it to the Catalog as a flow definition so that you can create a Cloudera Data Flow deployment.

### Procedure

1. On the **Flow Designer** canvas, go to Flow Options Publish To Catalog Publish .
2. Fill in the fields in the **Publish A New Flow** modal window.
  - Provide a Flow Name for your flow definition.  
You can only provide a name when you publish your flow for the first time.
  - Optionally, provide a Flow Description.  
You can only provide a description when you publish your flow for the first time.
  - Optionally, provide Custom Tags.  
You can filter flow definition versions by tags in the Catalog.
  - Optionally, provide Version Comments.
3. Click the Publish button.

### Results

Your draft is published to the Catalog as a flow definition.

### Related Information

[Deploying a flow definition](#)