Cloudera Data Flow

Tutorial: Build a New Flow

Date published: 2021-04-06
Date modified: 2025-09-30

https://docs.cloudera.com/

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: Building a new flow from scratch

If you are new to flow design and have never used NiFi before, this tutorial is for you. Learn how to build a draft adding and configuring components, connecting them, creating Controller Services, and testing your flow while creating it.

About this task

This tutorial walks you through the creation of 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:

- Creating a draft
- Creating a Controller Service
- Adding processors to your draft
- Configuring processors
- Adding a user-defined property to a processor configuration
- Connecting processors to create relationships between them
- Running a Test Session
- Publishing a draft to the Catalog as a flow definition

Before you begin

The flow you are about to build 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:

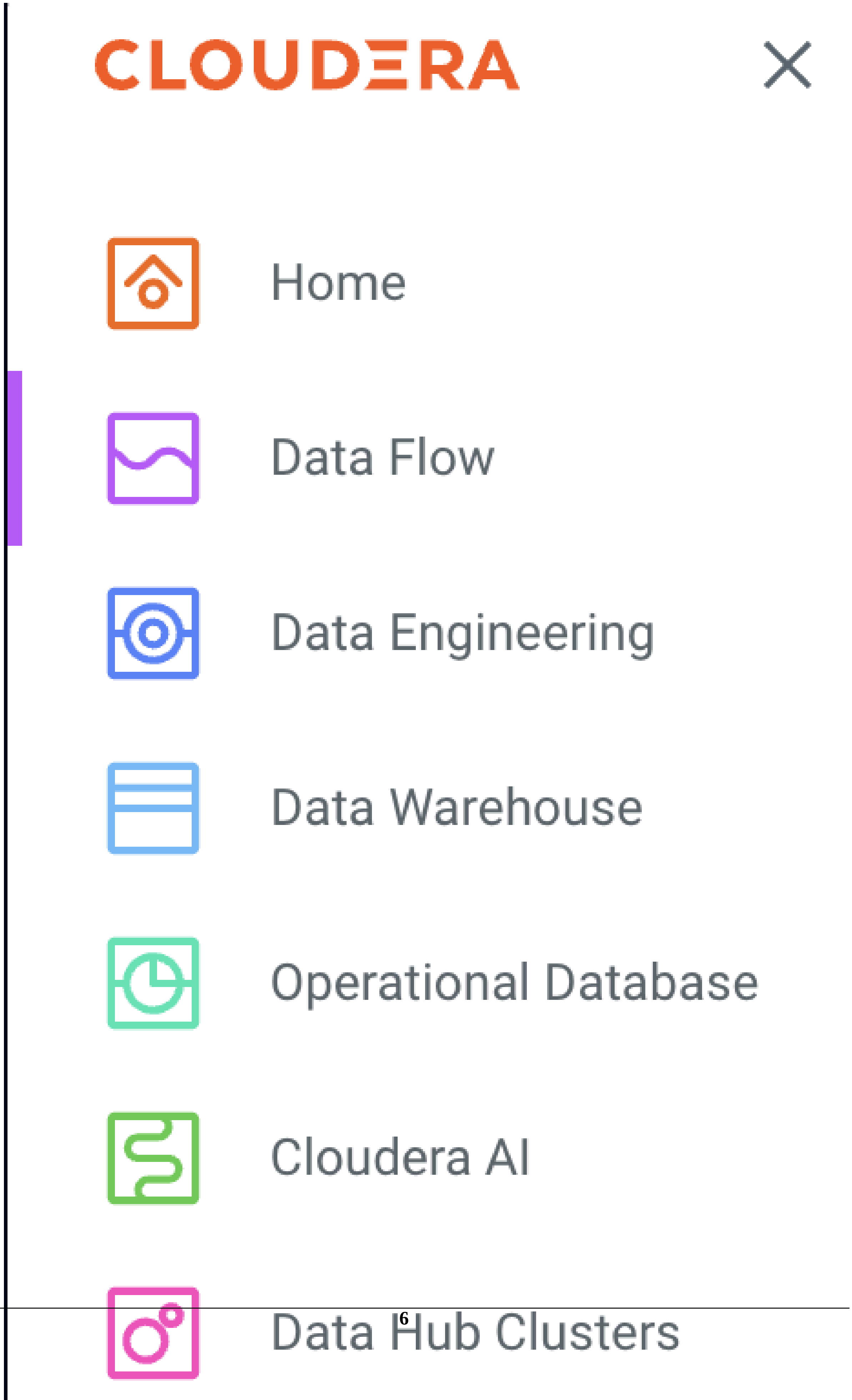
- You must have an enabled and healthy Cloudera Data Flow environment.
- You must be assigned the DFDeveloper role granting you access to the Flow Designer.
- You must be assigned the DFCatalogAdmin or DFCatalogViewer role granting you access to the Catalog. You must have this authorization to publish 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.

Create a new flow

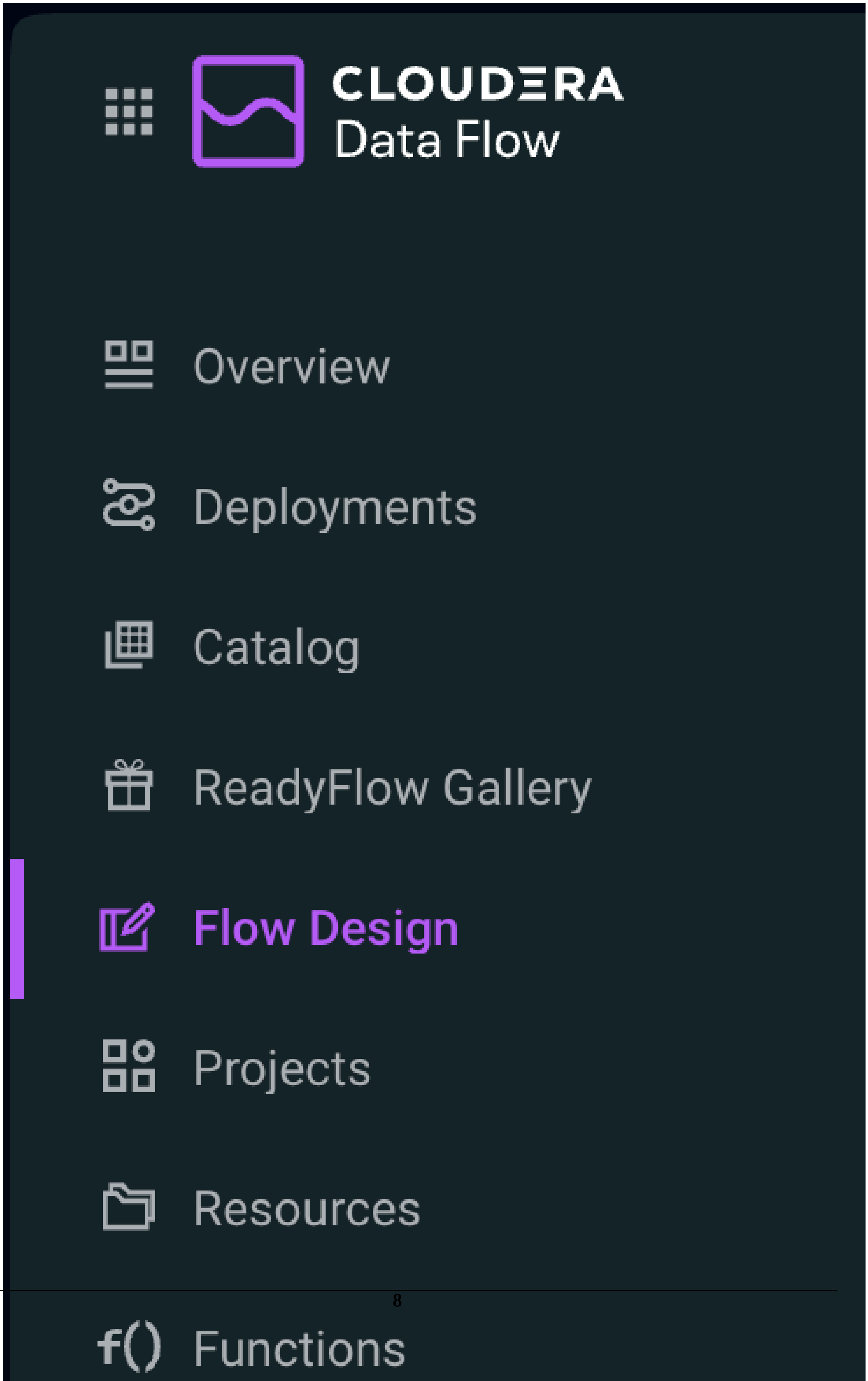
Create and name a new flow in a Flow Designer Workspace.

Procedure

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

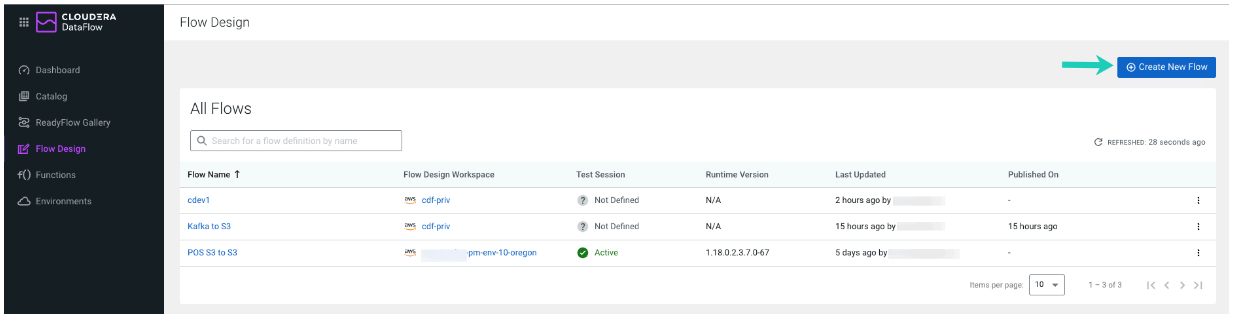

CLOUDERA
Home
Data Flow
Data Engineering
Data Warehouse
Operational Database
Cloudera AI
Data Hub Clusters

2. Click Flow Design in the left navigation pane.


CLOUDERA
Data Flow
Overview
Deployments
Catalog
ReadyFlow Gallery
Flow Design
Projects
Resources
Functions

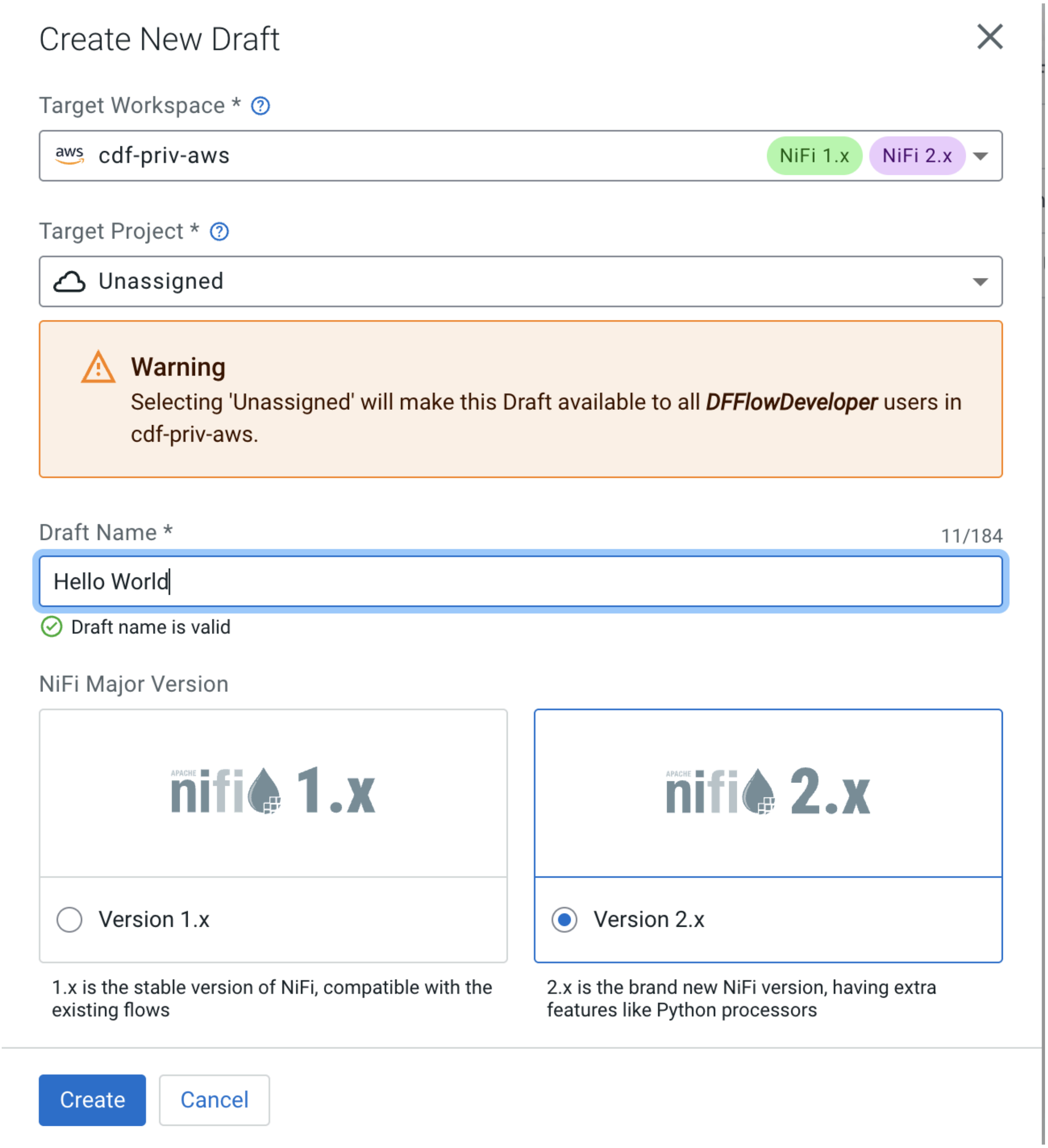
You are redirected to the **Flow Design** page, where previously created draft flows are displayed, one flow per row.

3. Click the ⊕ Create Draft button.



Flow Name ↑	Flow Design Workspace	Test Session	Runtime Version	Last Updated	Published On
cdev1	aws cdf-priv	? Not Defined	N/A	2 hours ago by [redacted]	-
Kafka to S3	aws cdf-priv	? Not Defined	N/A	15 hours ago by [redacted]	15 hours ago
POS S3 to S3	aws [redacted]-pm-env-10-oregon	✓ Active	1.18.0.2.3.7.0-67	5 days ago by [redacted]	-

4. Select a Target Workspace where you want to create the draft.



Create New Draft

Target Workspace * ⓘ

cdf-priv-aws NiFi 1.x NiFi 2.x

Target Project * ⓘ

☁ Unassigned

⚠ **Warning**

Selecting 'Unassigned' will make this Draft available to all ***DFFlowDeveloper*** users in cdf-priv-aws.

Draft Name * 11/184

Hello World

✓ Draft name is valid

NiFi Major Version

○ Version 1.x

1.x is the stable version of NiFi, compatible with the existing flows

● Version 2.x

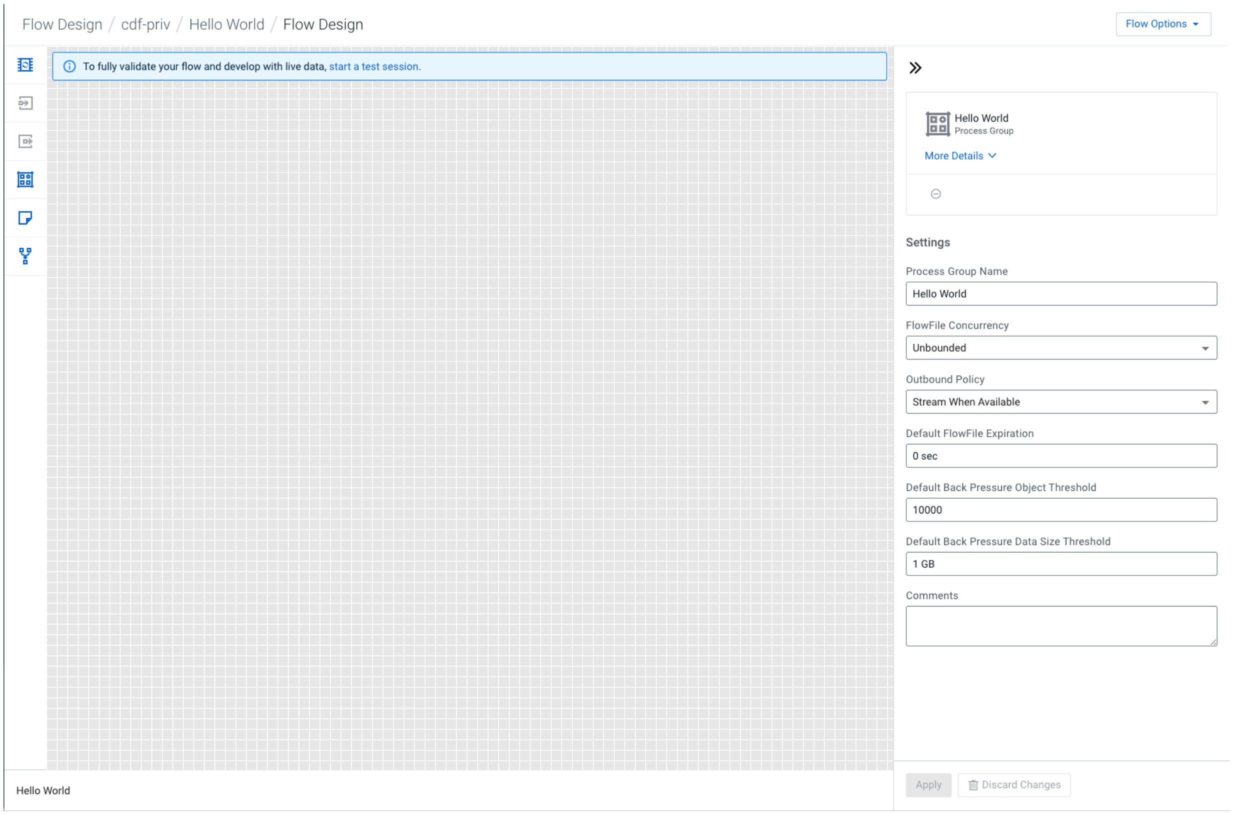
2.x is the brand new NiFi version, having extra features like Python processors

5. In the Target Project field, select the ☁ Unassigned option.
6. Provide a Draft Name.

   For example, provide Hello World.
7. Select the Version 2.x radio card for the NiFi Major Version field.

8. Click the Create button.

Flow Designer creates a default Process Group with the Draft Name you provided, Hello World in this case, and you are redirected to the **Flow Design** canvas. The **Configuration** pane on the right displays configuration options for the default Process Group.



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

Hello World
Process Group

Settings

Process Group Name: Hello World

FlowFile Concurrency: Unbounded

Outbound Policy: Stream When Available

Default FlowFile Expiration: 0 sec

Default Back Pressure Object Threshold: 10000

Default Back Pressure Data Size Threshold: 1 GB

Comments:

Hello World

What to do next

Proceed to creating controller services.

Related Tasks

Create controller services

Create controller services

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

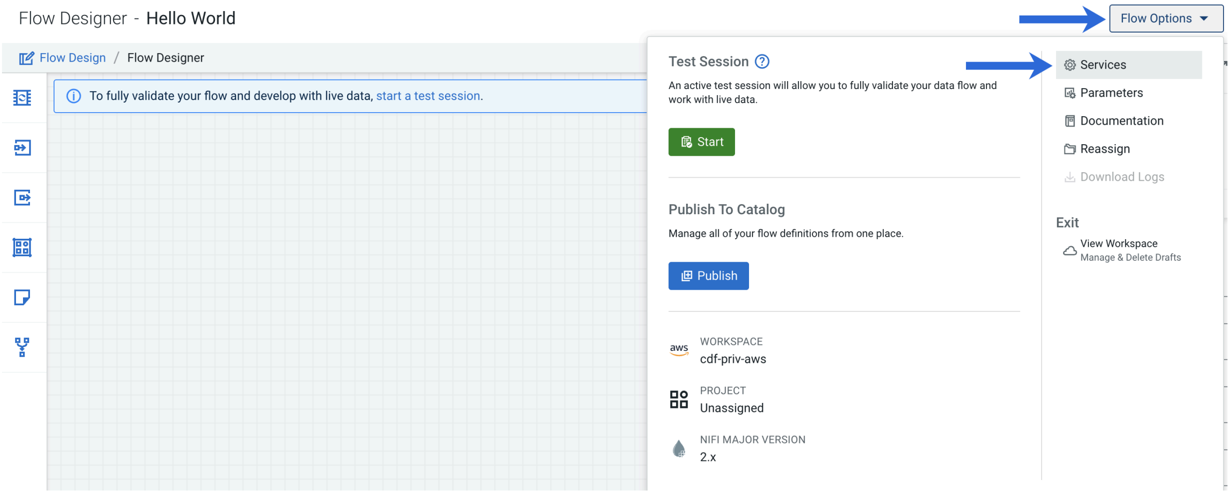
About this task

Controller 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 controller service provides the information for any processor to use as needed.

You will use the controller 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 .



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

An active test session will allow you to fully validate your data flow and work with live data.

Manage all of your flow definitions from one place.

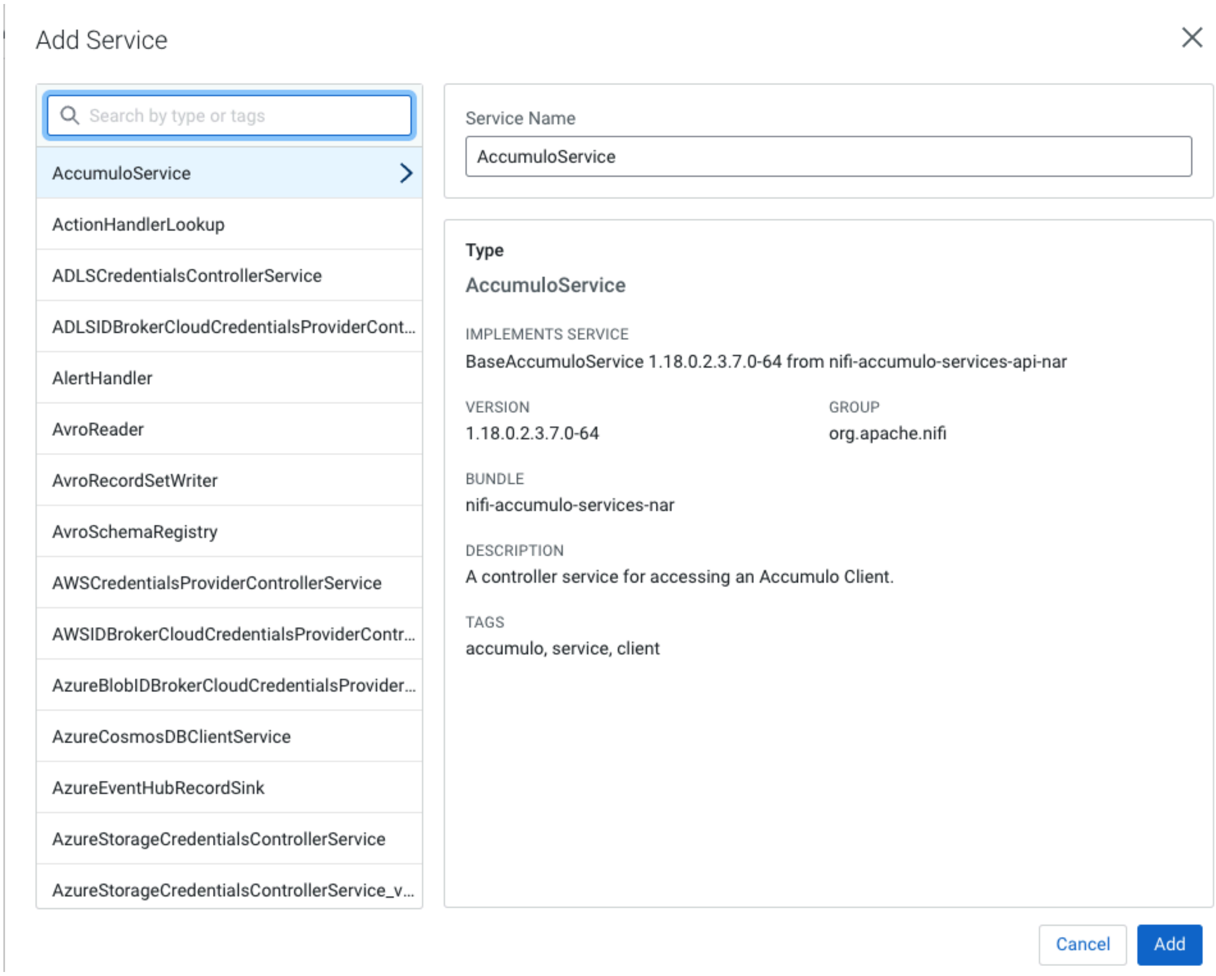
WORKSPACE cdf-priv-aws

PROJECT Unassigned

NIFI MAJOR VERSION 2.x

2. Click the ⊕ Add Service button.

   The **Add Service** page displays.


Add Service
Search by type or tags
AccumuloService
ActionHandlerLookup
ADLSCredentialsControllerService
ADLSIDBrokerCloudCredentialsProviderCont...
AlertHandler
AvroReader
AvroRecordSetWriter
AvroSchemaRegistry
AWSCredentialsProviderControllerService
AWSIDBrokerCloudCredentialsProviderContr...
AzureBlobIDBrokerCloudCredentialsProvider...
AzureCosmosDBClientService
AzureEventHubRecordSink
AzureStorageCredentialsControllerService
AzureStorageCredentialsControllerService_v...
Cancel
Add

Service Name: AccumuloService

Type: AccumuloService

IMPLEMENTS SERVICE: BaseAccumuloService 1.18.0.2.3.7.0-64 from nifi-accumulo-services-api-nar

VERSION: 1.18.0.2.3.7.0-64

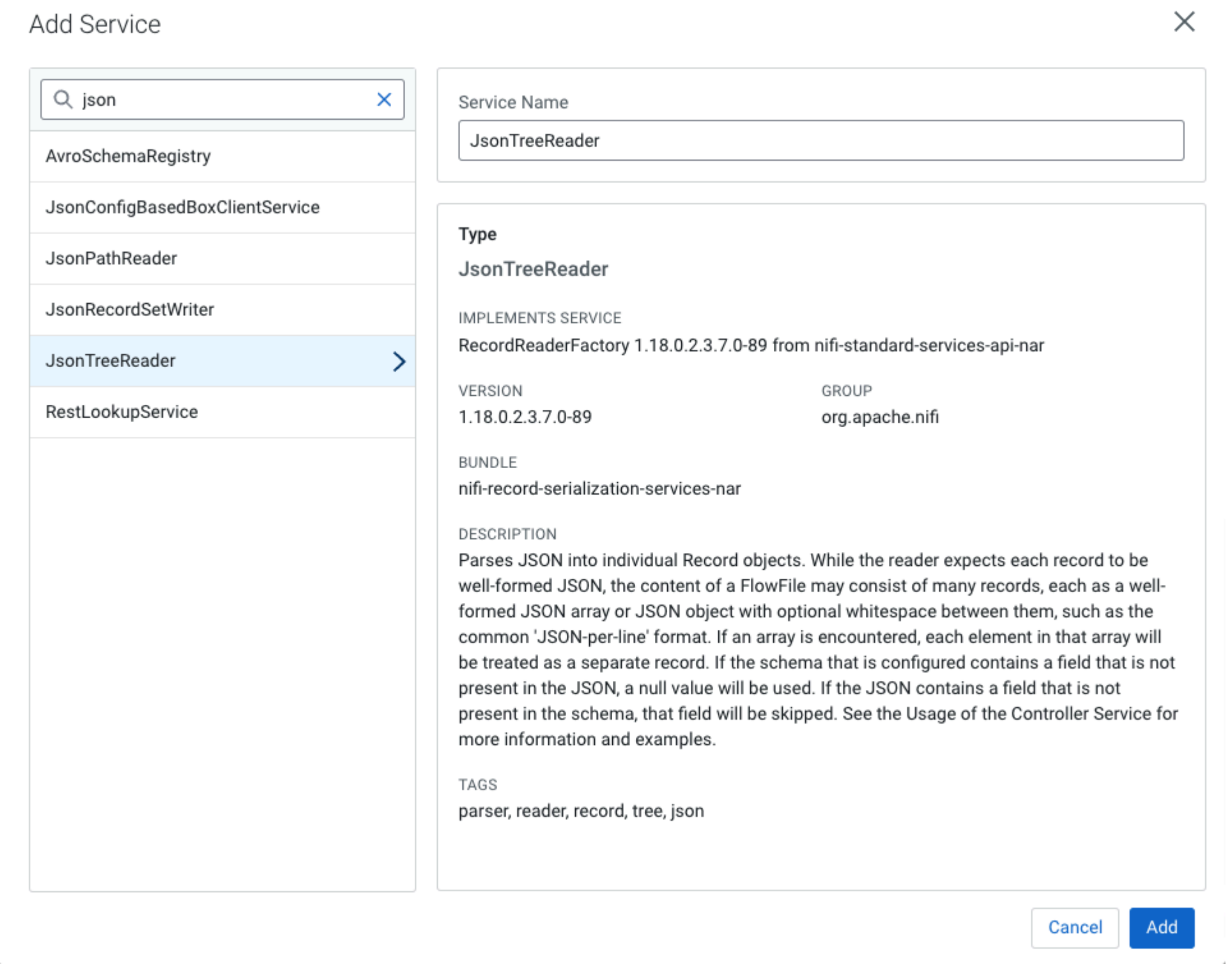
GROUP: org.apache.nifi

BUNDLE: nifi-accumulo-services-nar

DESCRIPTION: A controller service for accessing an Accumulo Client.

TAGS: accumulo, service, client

3. In the Search field, filter for JsonTreeReader.



Service Name

JsonTreeReader

Type

JsonTreeReader

IMPLEMENTS SERVICE

RecordReaderFactory 1.18.0.2.3.7.0-89 from nifi-standard-services-api-nar

VERSION

1.18.0.2.3.7.0-89

GROUP

org.apache.nifi

BUNDLE

nifi-record-serialization-services-nar

DESCRIPTION

Parses JSON into individual Record objects. While the reader expects each record to be well-formed JSON, the content of a FlowFile may consist of many records, each as a well-formed JSON array or JSON object with optional whitespace between them, such as the common 'JSON-per-line' format. If an array is encountered, each element in that array will be treated as a separate record. If the schema that is configured contains a field that is not present in the JSON, a null value will be used. If the JSON contains a field that is not present in the schema, that field will be skipped. See the Usage of the Controller Service for more information and examples.

TAGS

parser, reader, record, tree, json

4. Provide the Service Name as JSON_Reader_Recent_Changes.
5. Click the Add button.

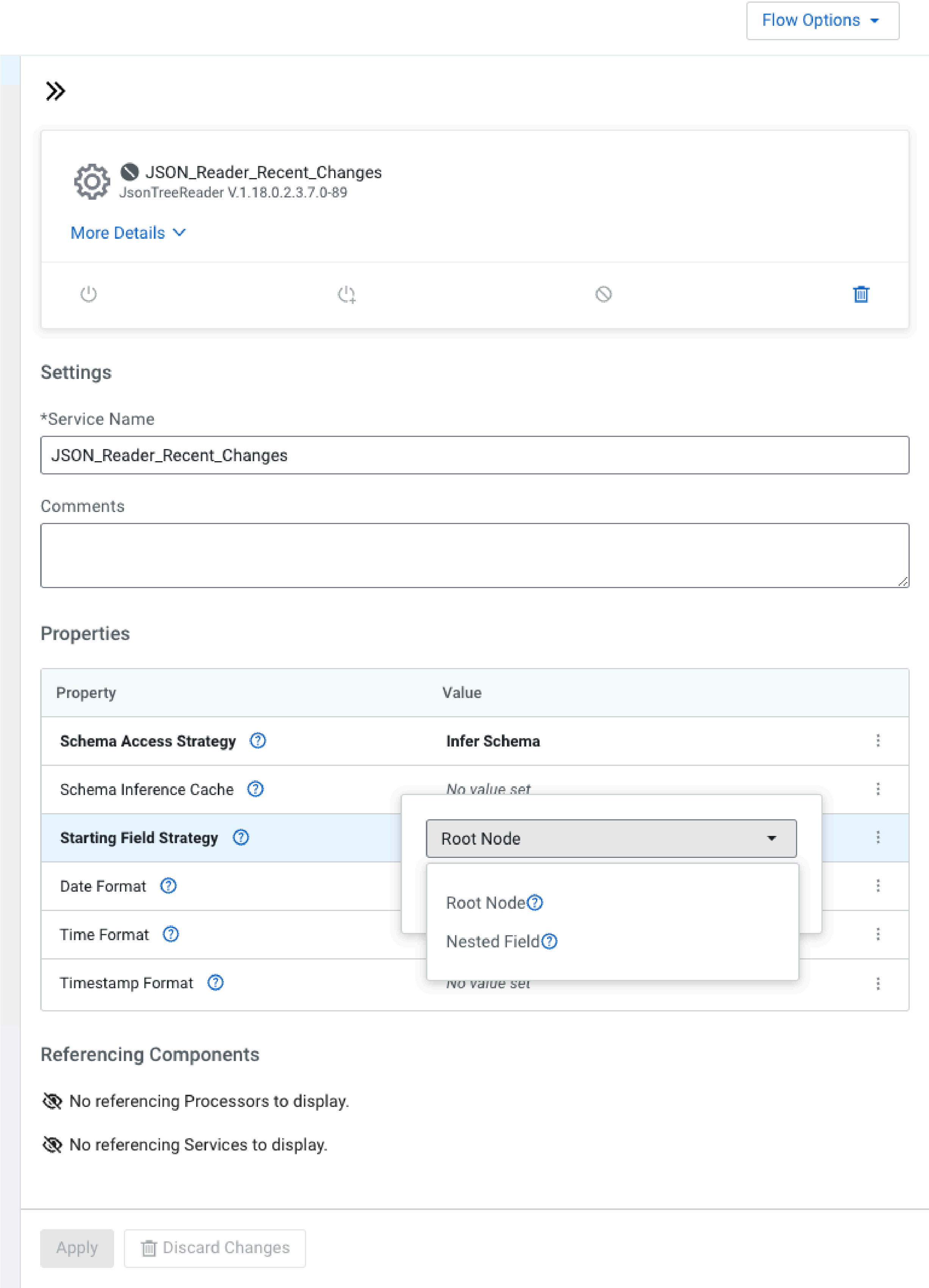
6. Configure the JSON_Reader_Recent_Changes service by setting the following Properties:

   Starting Field Strategy

   Set to Nested Field.

   Starting Field Name

   Set to recentchanges.



Flow Options
JSON_Reader_Recent_Changes
JsonTreeReader V.1.18.0.2.3.7.0-89
More Details

Settings

*Service Name

JSON_Reader_Recent_Changes

Comments

Properties

Property	Value
Schema Access Strategy	**Infer Schema**
Schema Inference Cache	*No value set*
Starting Field Strategy	Root Node
Date Format	
Time Format	
Timestamp Format	*No value set*

Referencing Components

No referencing Processors to display.

No referencing Services to display.

7. Click the Apply button.

8. Click the ⊕ Add Service button to create another service.
9. In the Search field, filter for AvroRecordSetWriter.
10. Provide the Service Name as AvroWriter_Recent_Changes.
11. Click the Add button.

    You do not need to configure the AvroWriter_Recent_Changes service. You can leave all properties with their default values.
12. Click the ⊕ Add Service button to create a third service.
13. In the Search field, filter for AvroReader.
14. Provide the Service Name as AvroReader_Recent_Changes.
15. Click the Add button.

    You do not need to configure the AvroReader_Recent_Changes service. You can leave all properties with their default values.
16. In the breadcrumbs on top, click Flow Designer to return to the flow design **Canvas**.

What to do next

After creating the necessary Controller Services, you can start building and configuring your flow.

Related Tasks

Build your draft flow

Build your draft flow

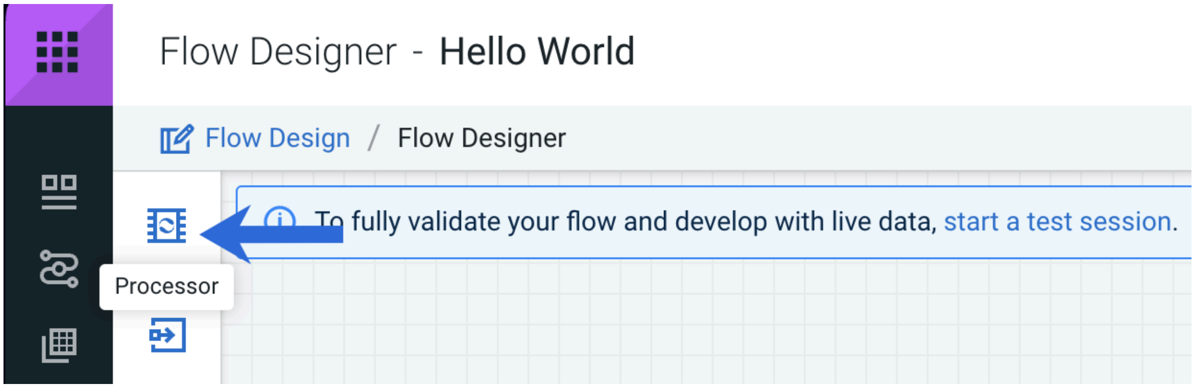
Begin creating your draft flow by adding components to the canvas and setting them up.

Procedure

1. Add an InvokeHTTP processor to the canvas.

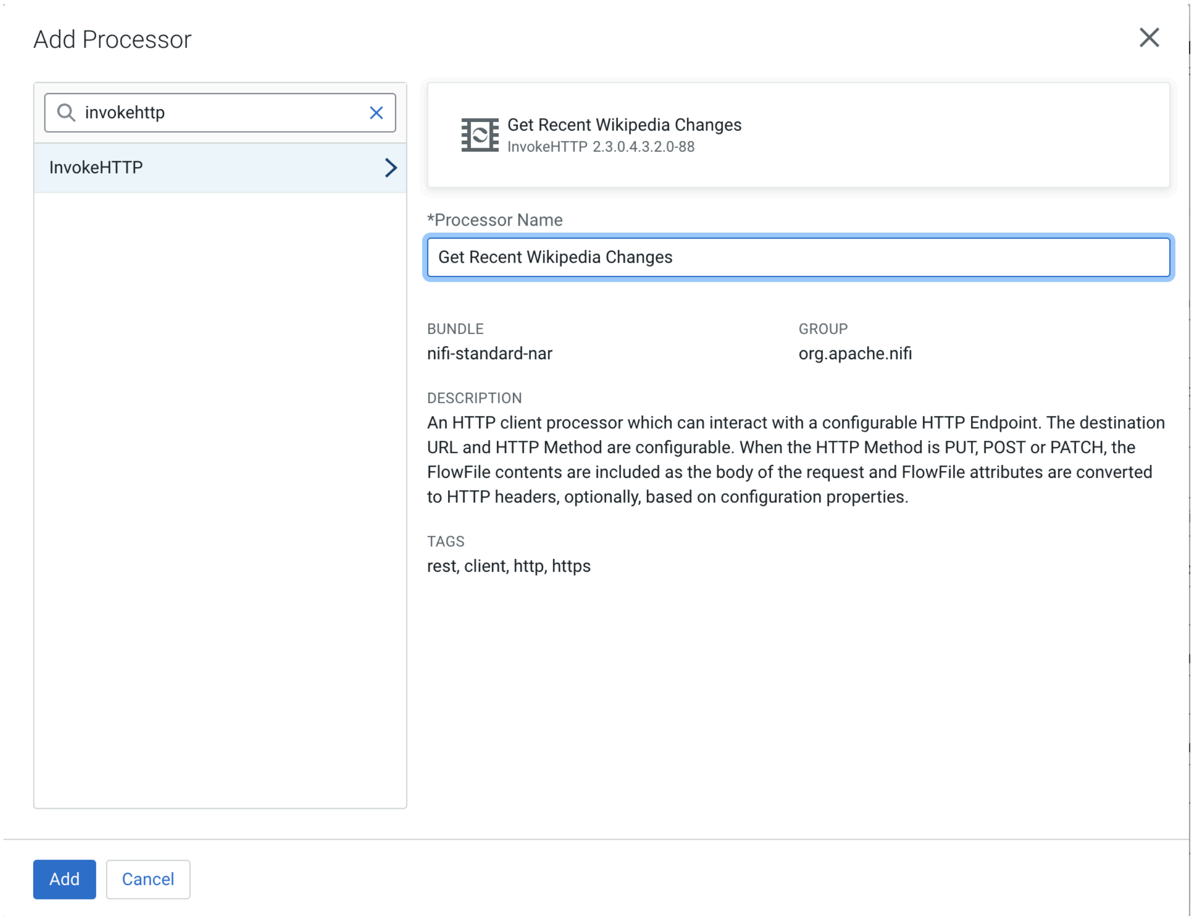
   Once configured, this processor will call the Wikipedia API to fetch the latest changes.

   a) Drag a Processor from the **Components** sidebar to the canvas.

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

   b) In the Search field, filter for InvokeHTTP.

   Search: invokehttp

   Get Recent Wikipedia Changes
   InvokeHTTP 2.3.0.4.3.2.0-88

   *Processor Name: Get Recent Wikipedia Changes

   BUNDLE: nifi-standard-nar

   GROUP: org.apache.nifi

   DESCRIPTION
   An HTTP client processor which can interact with a configurable HTTP Endpoint. The destination URL and HTTP Method are configurable. When the HTTP Method is PUT, POST or PATCH, the FlowFile contents are included as the body of the request and FlowFile attributes are converted to HTTP headers, optionally, based on configuration properties.

   TAGS
   rest, client, http, https

   c) Rename the InvokeHTTP processor by changing the Processor Name to Get Recent Wikipedia Changes.

   d) Click the Add button.

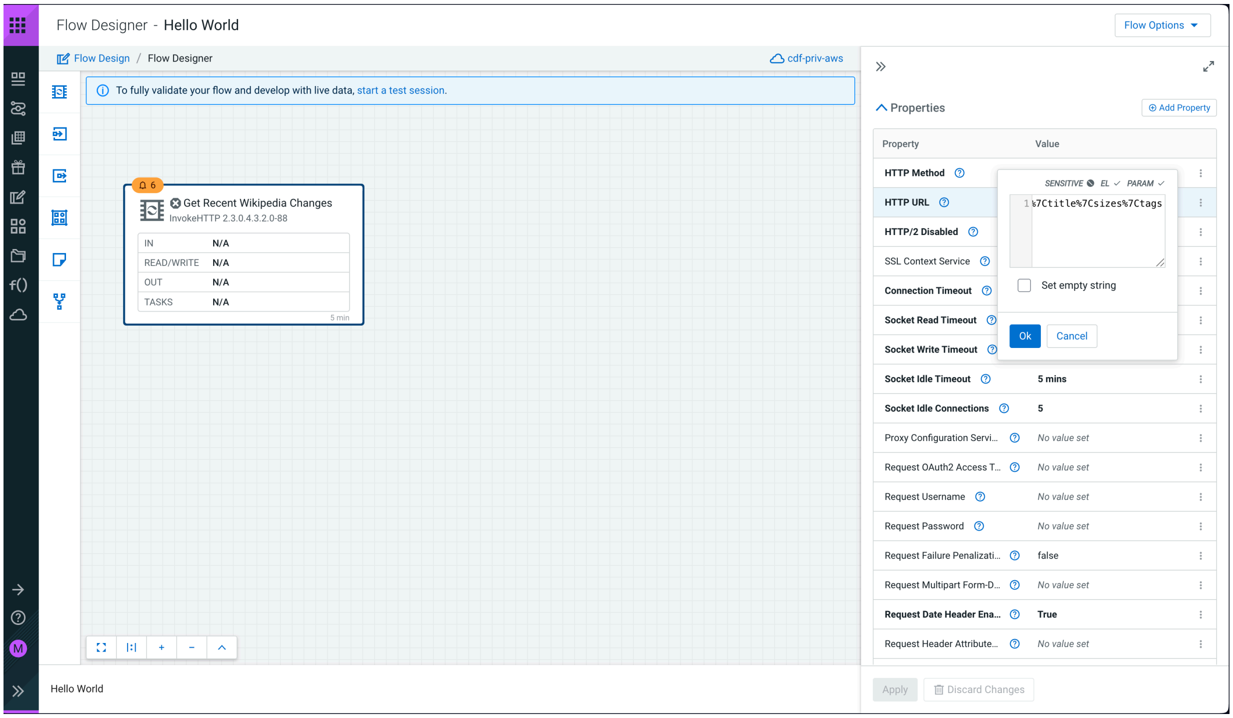
2. Configure the Get Recent Wikipedia Changes processor.

   Properties

   HTTP URL

   Enter the following URL:

   ```
   https://en.wikipedia.org/w/api.php?action=query&list=recentchang
   es&format=json&rcprop=user%7Ccomment%7Cparsedcomment%7Ctimestamp
   %7Ctitle%7Csizes%7Ctags
   ```



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

Get Recent Wikipedia Changes
InvokeHTTP 2.3.0.4.3.2.0-88

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

Properties

Property	Value
HTTP Method	
HTTP URL	%7Ctitle%7Csizes%7Ctags
HTTP/2 Disabled	
SSL Context Service	
Connection Timeout	
Socket Read Timeout	
Socket Write Timeout	
Socket Idle Timeout	**5 mins**
Socket Idle Connections	**5**
Proxy Configuration Servi...	*No value set*
Request OAuth2 Access T...	*No value set*
Request Username	*No value set*
Request Password	*No value set*
Request Failure Penalizati...	false
Request Multipart Form-D...	*No value set*
Request Date Header Ena...	**True**
Request Header Attribute...	*No value set*

□ Set empty string

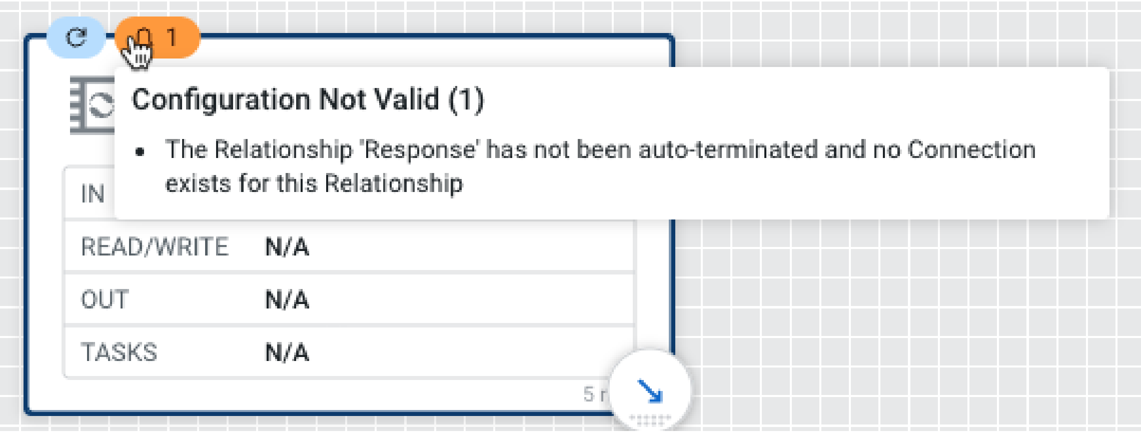
Relationships

Select the following relationships:

- Original –Terminate
- Failure – Terminate, Retry
- Retry – Terminate
- No Retry – Terminate

3. Click the Apply button.

   Notice: Notice the orange 🔔 Notification Pill icon in the upper left corner of your processor. The icon warns you about potential issues with a component.

   Configuration Not Valid (1)

   - The Relationship 'Response' has not been auto-terminated and no Connection exists for this Relationship

   In this example, the Notification Pill icon alerts you about an unaddressed Relationships configuration. The notification will disappear once you connect the processor to another one for the 'Response' relationship.

4. Add a ConvertRecord processor to the canvas.

   a) Drag a Processor from the **Components** sidebar to the canvas.
   b) In the Search field, filter for ConvertRecord.
   c) Change the Processor Name to Convert JSON to AVRO.
   d) Click the Add button.

   This processor converts the JSON response to AVRO format by using RecordReaders and RecordWriters. It infers the JSON schema starting from the recent changes field.

5. Configure the Convert JSON to AVRO processor.

   Properties

   Record Reader

   Select the JSON_Reader_Recent_Changes controller service you created earlier from the dropdown list.

   Record Writer

   Select the AvroWriter_Recent_Changes controller service you created earlier from the dropdown list.

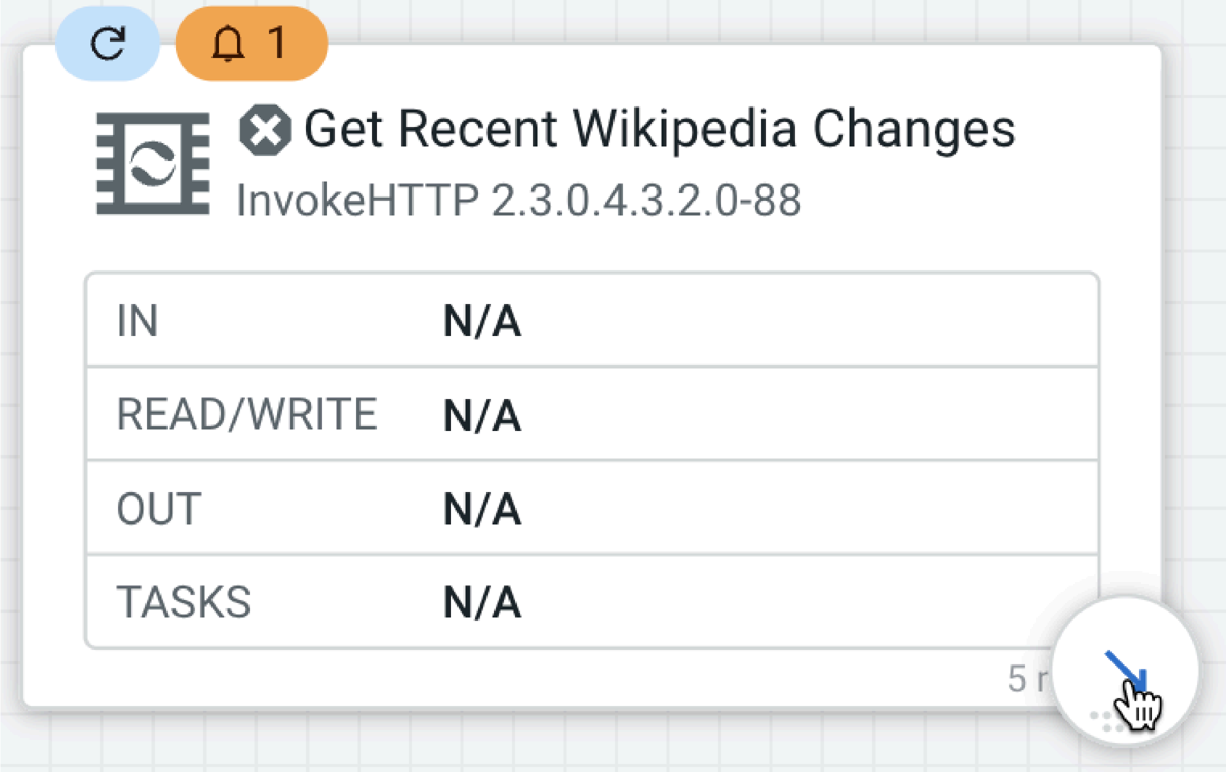
   Relationships

   Select the following relationships:

   Failure - Terminate, Retry

6. Click the Apply button.

7. Connect the Get Recent Wikipedia Changes and Convert JSON to AVRO processors by hovering over the lower-right corner of the Get Recent Wikipedia Changes processor, clicking the arrow that appears and dragging it to the Convert JSON to AVRO processor.

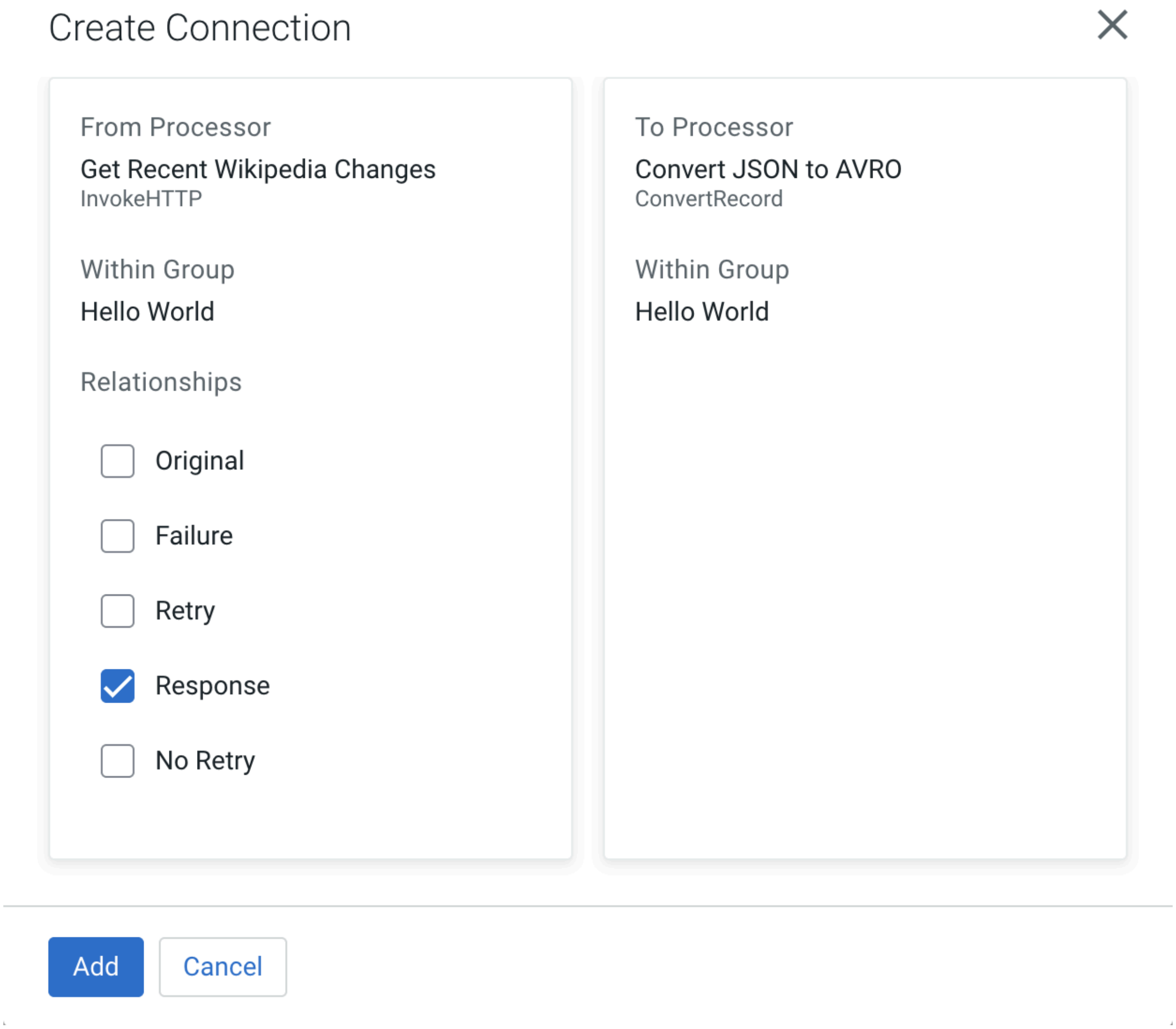


Get Recent Wikipedia Changes

InvokeHTTP 2.3.0.4.3.2.0-88

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

8. In the configuration popup, select the Response relationship and click the Add button.



Create Connection

From Processor

Get Recent Wikipedia Changes
InvokeHTTP

Within Group

Hello World

Relationships

□ Original

□ Failure

□ Retry

☑ Response

□ No Retry

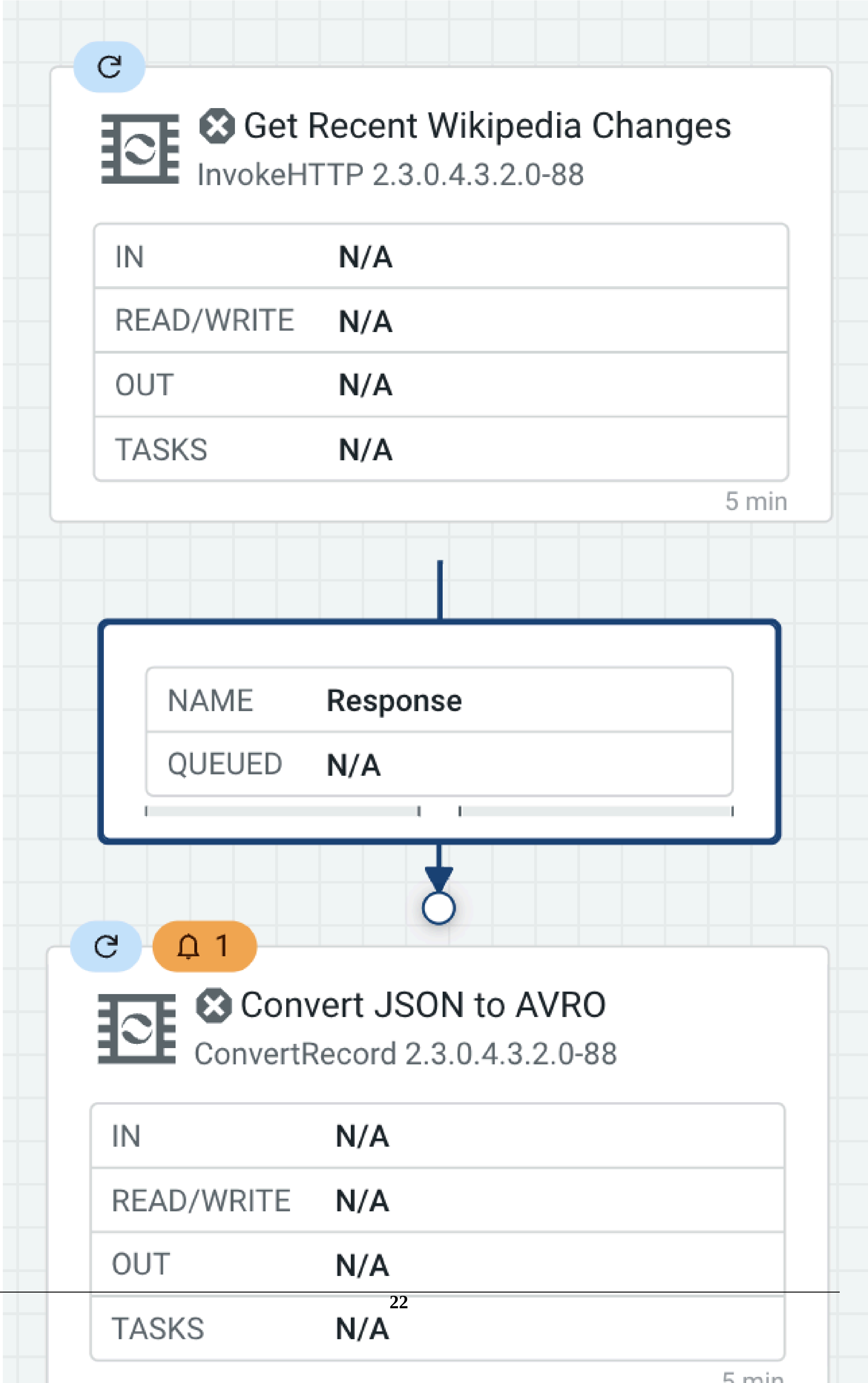
To Processor

Convert JSON to AVRO
ConvertRecord

Within Group

Hello World

Notice: Notice, that the 🔔 Notification Pill warning about the unconfigured 'Response' relationship disappeared from your Get Recent Wikipedia Changes processor.



Get Recent Wikipedia Changes

InvokeHTTP 2.3.0.4.3.2.0-88

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

5 min

NAME	Response
QUEUED	N/A

Convert JSON to AVRO

ConvertRecord 2.3.0.4.3.2.0-88

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

5 min

9. Add a QueryRecord processor.

a) Drag a Processor from the **Components** sidebar to the canvas.
b) In the Search field, search for QueryRecord.
c) Change the processor Name to Filter Edits.
d) Click the Add button.

This processor filters out anything except actual page edits. To achieve this, the processor runs a query that selects all FlowFiles (events) of the edit type.

10. Configure the Filter Edits processor.

Properties

Record Reader

Select the AvroReader_Recent_Changes controller service you have created from the dropdown list.

Record Writer

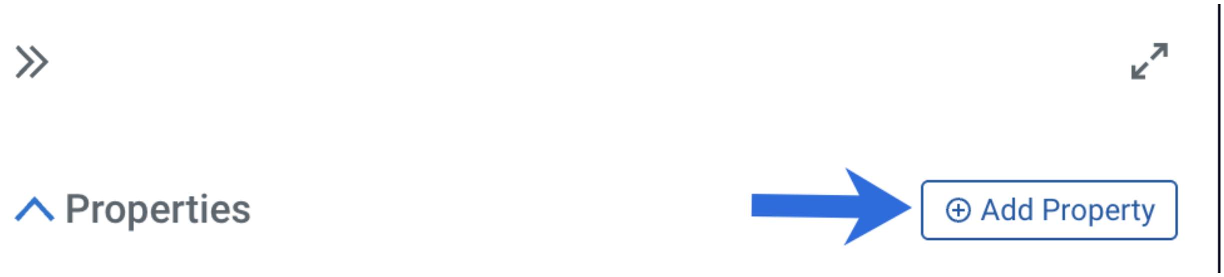
Select the AvroWriter_Recent_Changes controller service you have created from the dropdown list.

Relationships

Select the following relationships:

- Failure - Terminate
- Original - Terminate

11. For the Filter Edits processor you also must add a user-defined property. Click the ⊕ Add Property button.



a) Provide the Name as Filtered edits.
b) Provide the Value as Select * from FLOWFILE where type='edit'.
c) Click the Apply button.

12. Connect the Convert JSON to AVRO and Filter Edits processors by hovering over the lower-right corner of the Convert JSON to AVRO processor, clicking the arrow that appears and dragging it to the Filter Edits processor.

13. In the configuration pane, select the Success and Failure relationships and click the Add button.

14. Add a second QueryRecord processor.

a) Drag a Processor from the Components sidebar to the canvas.
b) In the Search field, filter for QueryRecord.
c) Change the processor Name to Route on Content Size.
d) Click the Add button.

This processor uses two SQL statements to separate edit events that resulted in a longer article from edit events that resulted in a shorter article.

15. Configure the Route on Content Size processor.

Properties

Record Reader

Select the AvroReader_Recent_Changes controller service you have created from the dropdown list.

Record Writer

Select the AvroWriter_Recent_Changes controller service you have created from the dropdown list.

Relationships

Select the following relationships:

- Failure - Terminate, Retry
- Original - Terminate

16. For the Route on Content Size processor you also must add two user-defined properties.

a) Click the Add Property button.
b) Provide the Name as Added content.
c) Provide the Value as Select * from FLOWFILE where newlen>=oldlen.
d) Click the Add button.
e) Click the Add Property button, to create the second property.
f) Provide the Name as Removed content.
g) Provide the Value as Select * from FLOWFILE where newlen<oldlen .
h) Click the Add button.
i) Click the Apply button.

17. Connect the Filter Edits and Route on Content Size processors by hovering over the lower-right corner of the Filter Edits processor, clicking the arrow that appears and drawing it to Route on Content Size.

In the Create Connection pop up, select the Filtered edits relation and click Add.

18. Add two MergeRecord processors.

a) Drag a Processor from the **Components** sidebar to the canvas.
b) In the Search field, filter for MergeRecord.
c) Change the processor Name to Merge Edit Events.
d) Click the Add button.
e) Repeat steps a. to d. to add another identical processor.

These processors are configured to merge at least 100 records into one flowfile to avoid writing lots of small files. The MaxBinAge property is set to 2 minutes, which makes the processors merge records after two minutes even if less than 100 records have arrived.

19. Configure the two Merge Edit Events processors.

Properties

Record Reader

Select the AvroReader_Recent_Changes controller service you have created from the dropdown list.

Record Writer

Select the AvroWriter_Recent_Changes controller service you have created from the dropdown list.

Max Bin Age

Set to two minutes by providing a value of 2 min.

Relationships

Select the following relationships:

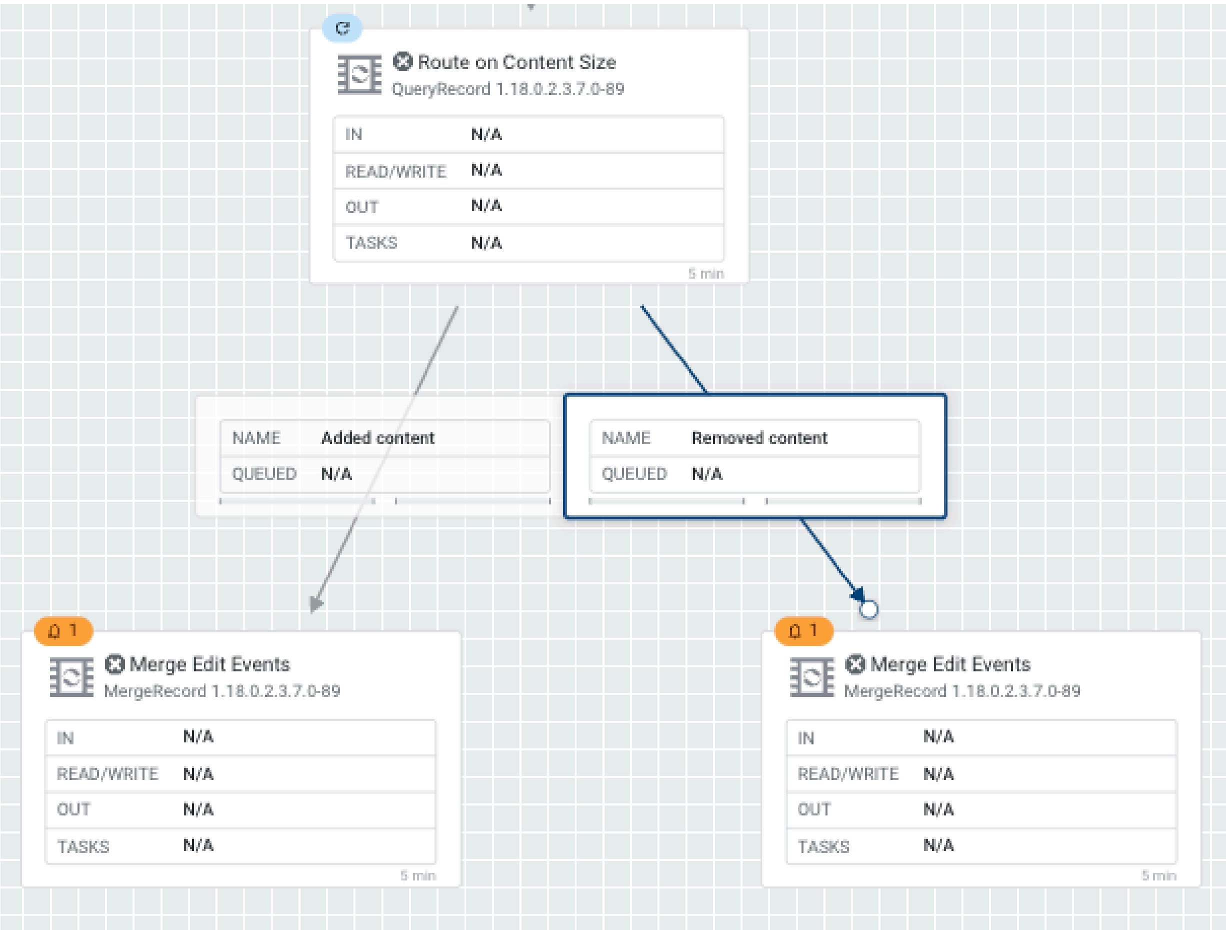
- Failure - Terminate
- Original - Terminate

Note: Do not forget to perform configuration for both Merge Edit Events processors.

20. Connect the Route on Content Size processor to both of the Merge Edit Events processors.

a) For the first Merge Edit Events processor, select Added content value from the Relationships options and click the Add button.

b) For the second Merge Edit Events processor, select the Removed content value from the Relationships option and click the Add button.



Route on Content Size
QueryRecord 1.18.0.2.3.7.0-89

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

5 min

NAME	Added content
QUEUED	N/A

NAME	Removed content
QUEUED	N/A

Merge Edit Events
MergeRecord 1.18.0.2.3.7.0-89

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

5 min

Merge Edit Events
MergeRecord 1.18.0.2.3.7.0-89

IN	N/A
READ/WRITE	N/A
OUT	N/A
TASKS	N/A

5 min

21. Add two PutFile processors to the canvas.

a) Drag a Processor from the Components sidebar to the canvas.

b) In the Search field, filter for PutFile.

c) Change the processor Name to Write "Added Content" Events To File.

d) Click the Add button.

e) Repeat steps a. to d. to add another identical processor, naming this second PutFile processor Write "Removed Content" Events To File.

These processors write the filtered, routed edit events to two different locations on the local disk. In Cloudera Data Flow, you typically do not write to local disk but replace these processors with processors that resemble your destination system, such as Kafka, Database, or Object Store.

22. Configure the Write "Added Content" Events To File processor.

Properties:

Directory

Set to /tmp/larger_edits.

Maximum File Count

Set to 500.

Relationships

Select the following relationships:

- Failure - Terminate
- Success - Terminate

23. Click the Apply button.

24. Configure the Write "Removed Content" Events To File processor.

Properties:

Directory

Set to /tmp/smaller_edits.

Maximum File Count

Set to 500.

Relationships

Select the following relationships:

- Failure - Terminate
- Success - Terminate

25. Click the Apply button.

26. Connect the Merge Edit Events processor with the Added content connection to the Write "Added Content" Events To File processor.

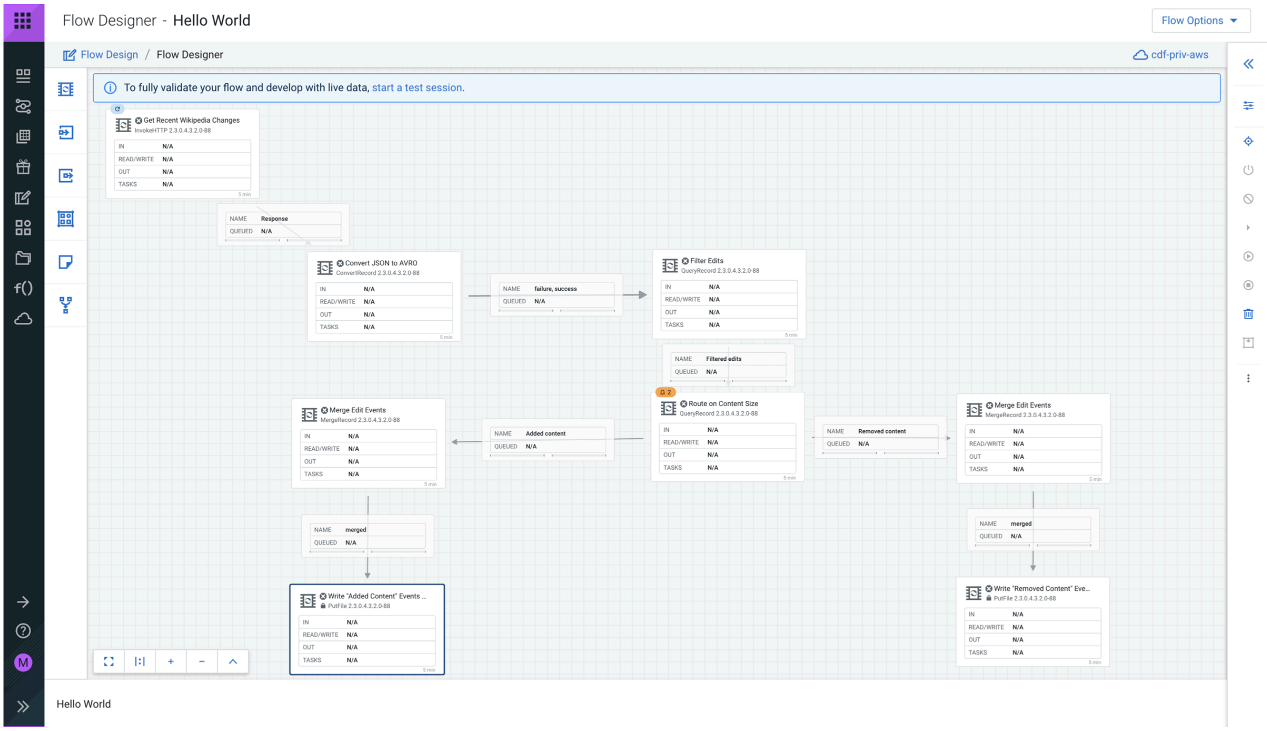
In the Create Connection modal select merged and click the Add button.

27. Connect the Merge Edit Events processor with the Removed content connection to the Write "Removed Content" Events To File processor.

In the Create Connection modal select merged and click the Add button.

Results

Congratulations, you have created your first draft flow. Now proceed to testing it by launching a Test Session.


Flow Designer - Hello World
Flow Options
Flow Design / Flow Designer
cdf-priv-aws
Get Recent Wikipedia Changes
InvokeHTTP 2.3.0.4.3.2.0-88
Response
Convert JSON to AVRO
ConvertRecord 2.3.0.4.3.2.0-88
failure, success
Filter Edits
QueryRecord 2.3.0.4.3.2.0-88
Filtered edits
Route on Content Size
QueryRecord 2.3.0.4.3.2.0-88
Added content
Removed content
Merge Edit Events
MergeRecord 2.3.0.4.3.2.0-88
merged
Write "Added Content" Events ...
PutFile 2.3.0.4.3.2.0-88
Write "Removed Content" Eve...
PutFile 2.3.0.4.3.2.0-88
IN N/A
READ/WRITE N/A
OUT N/A
TASKS N/A
NAME
QUEUED N/A
Hello World

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

Start a test session

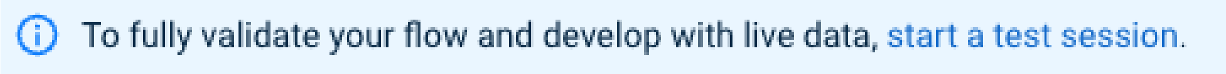
To validate your draft, start a test session. This provisions an Apache NiFi cluster where you can test your draft.

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.

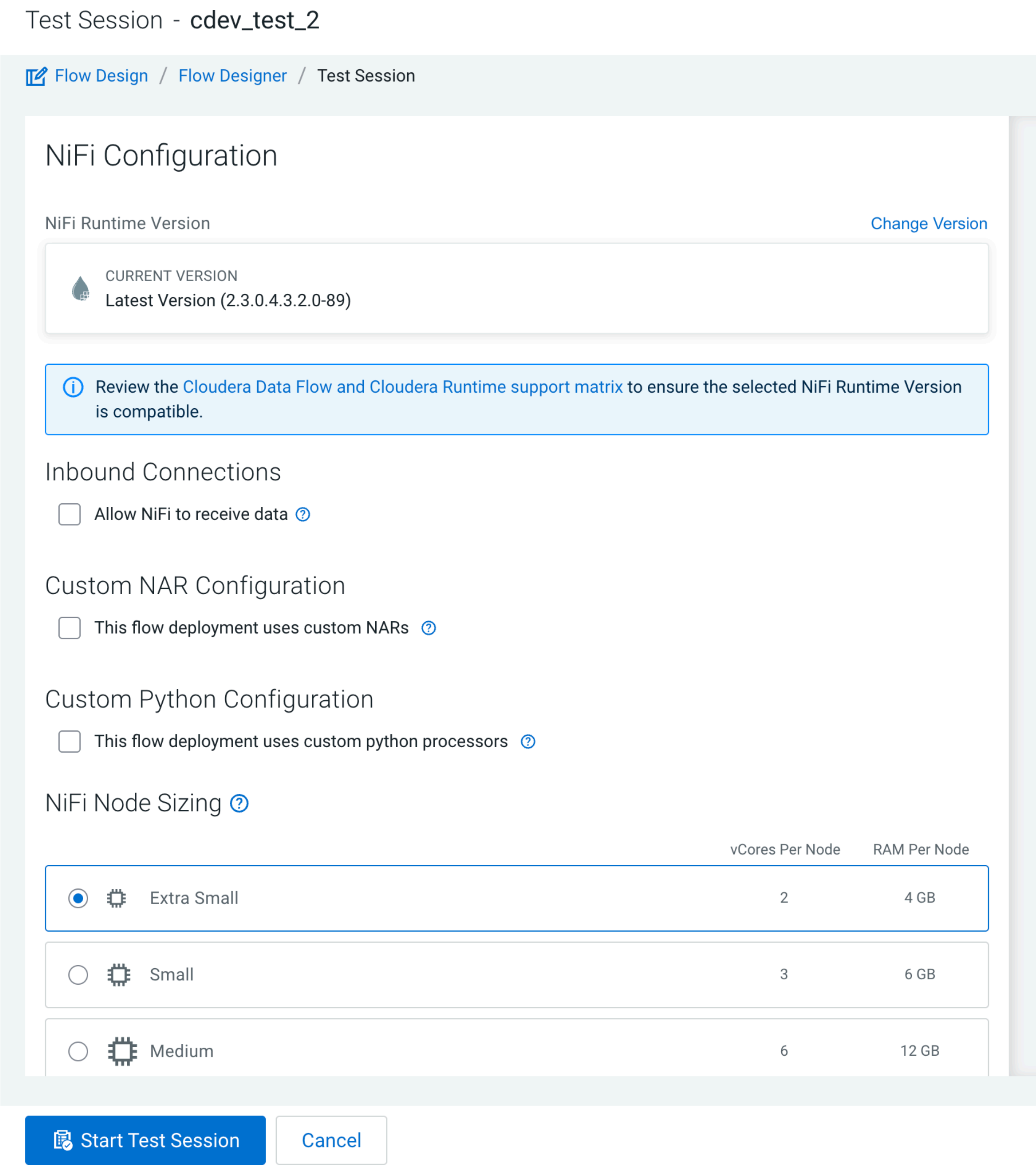
Procedure

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.

2. 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
● Extra Small	2	4 GB
○ Small	3	6 GB
○ Medium	6	12 GB

3. Click the Start Test Session button.

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

4. Wait for the status to change to

   Active Test Session.

   This can take several minutes.

5. 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.

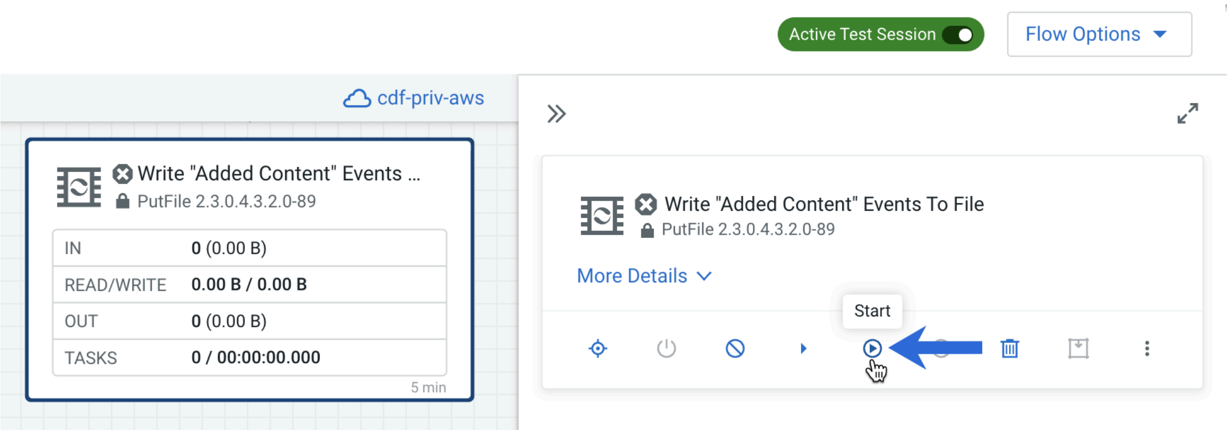
7. Click the Flow Designer link in the breadcrumbs on the top of the page to return to the Flow Design canvas.

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.

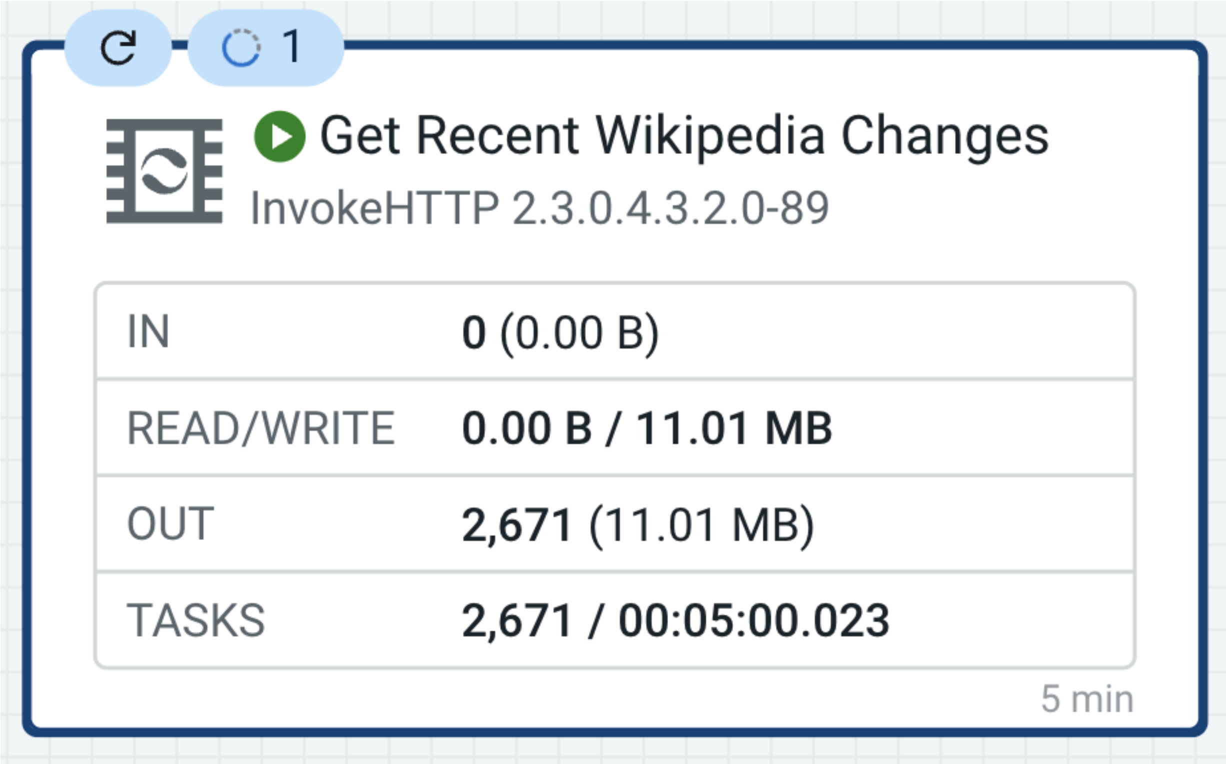


IN	0 (0.00 B)
READ/WRITE	0.00 B / 0.00 B
OUT	0 (0.00 B)
TASKS	0 / 00:00:00.000

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.



Get Recent Wikipedia Changes

InvokeHTTP 2.3.0.4.3.2.0-89

IN	0 (0.00 B)
READ/WRITE	0.00 B / 11.01 MB
OUT	2,671 (11.01 MB)
TASKS	2,671 / 00:05:00.023

5 min

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