

Getting Started with Cloudera Flow Management

Date published: 2019-06-26

Date modified: 2021-04-28

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with a stylized 'E' that has three horizontal bars.

Legal Notice

© Cloudera Inc. 2024. 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

What is Cloudera Flow Management?	4
Who is this guide for?	4
Flow management terminology	5
Set up Cloudera Flow Management	7
Download NiFi and NiFi Registry	7
Install NiFi and NiFi Registry	7
Start NiFi and NiFi Registry	7
Build your first data flow	8
Learn about NiFi processors	12
Work with data flow templates	14
Create a template	14
Download a template	15
Upload a template	15
Use a template	15
Manage templates	16
Connect NiFi to NiFi Registry	18
Create a bucket in NiFi Registry	18
Version data flows	19
Start version control	19
Commit local changes	21
Change the version of your data flow	22
Manage versioned data flows	23
Monitor data flows	24
Trace Data Lifecycle	25

What is Cloudera Flow Management?

Cloudera Flow Management (CFM) powered by Apache NiFi and NiFi Registry is a key part of the Cloudera DataFlow (CDF) platform. You can use CFM to ingest and manage data without writing a single line of code. It helps you to capture any type of data from any source and move it into the enterprise securely, while you transform and dynamically route this data in between.

You can build complex and scalable data flows, and you can manage, edit, and monitor these flows in real time with CFM. You can also adopt a DevOps style flow development lifecycle to deliver your flow applications faster by using templates, version control, collaboration, and various deployment options.

Cloudera Flow Management integrates seamlessly with CDF components such as Kafka, MiNiFi, Flink, and Spark Streaming, or platform components such as HBase, Hive, Impala, and Kudu.

CFM has two main components:

- Apache NiFi – the core data ingestion engine with a no-code graphical user interface. It supports processors for connectivity, transformation, and content routing.
- Apache NiFi Registry – the companion to NiFi that enables DevOps style development of flows. It supports versioning flows, promoting flows, and deploying flow applications across environments.

What is NiFi?

Apache NiFi is a powerful, flexible, extensible, and reliable system to distribute and process data. It is data source agnostic and supports sources of different formats, schemas, protocols, speeds, and sizes. You can design and automate complex data flows on a configurable platform using 300+ processors. You can build your data flow by simply dragging and dropping a series of processors using the intuitive graphical user interface of Apache NiFi. When the data flow is ready, you can monitor and control the pipeline in real time. For more information on Apache NiFi, see the *Apache NiFi User Guide*.

What is NiFi Registry?

Registry is a complimentary application that provides a central location for storing and managing shared and versioned data flow resources. It helps you with the full flow life cycle development. It integrates with NiFi to allow storing, retrieving, and upgrading versioned flows. When designing a data flow in NiFi, you may need to define and test different approaches and work incrementally towards your final goal. NiFi Registry helps you to keep track of several flow versions and navigate between them easily. For more information, see the *Apache NiFi Registry User Guide*.

Related Information

[Apache NiFi User Guide](#)

[Apache NiFi Registry User Guide](#)

Who is this guide for?

This guide is intended for Cloudera Flow Management (CFM) users who have never used, have had limited exposure to, or only accomplished specific tasks within NiFi and NiFi Registry.

Are you new to Cloudera Flow Management? This Getting Started guide gives you an introductory tour of CFM. It provides an overview of how to work with data flows, from defining the basic data flow concepts to introducing some of the most used features. It also gives you a step-by-step guide to building and handling simple data flows.

This guide is not an exhaustive instruction manual or a reference guide. The How To documentation is a much more comprehensive resource that provides you with detailed instructions on how to use CFM components. This content

is intended to show you CFM basics – how the different parts of Flow Management work together and how you can build simple data flows.


What you can learn in this Getting Started guide:








- Role of Apache NiFi and NiFi Registry in Cloudera Flow Management
- Key NiFi components that you can use to define a data flow
- How to create and configure data flows
- How to work with templates
- How to manage data flow versions using NiFi Registry
- How you can monitor data flows
- How you can trace the life of data using data provenance

Flow management terminology

Learn the most essential flow management terms and concepts used in connection with Cloudera Flow Management (CFM) components and operations.

There are a few key terms and concepts that you should be familiar with to understand the details of Cloudera Flow Management. For more terms and detailed definitions, see the *Apache NiFi User Guide* and the *Apache NiFi Registry User Guide*.

Term	Definition	UI icon (if applicable)
Processor	Apache NiFi processors are the basic blocks of creating a data flow. Every processor has a different functionality, which contributes to the creation of the output FlowFile. NiFi processors are responsible for creating, sending, receiving, transforming, routing, splitting, merging, and processing FlowFiles. There is a large number of processors available by default in NiFi, with the ability to write your own custom processors as well. Processors can be placed on the NiFi UI canvas and connected creating a data flow graph.	
FlowFile	Each piece of user data (data that you bring to NiFi for processing and transfer) is wrapped in entities called FlowFiles. They have the original data as content, and some attributes. Attributes are key-value pairs that are associated with the user data. Some of these attributes are set by default, but processors can add, remove, or edit them.	
Connection	In a NiFi data flow, FlowFiles move from processor one to another through a connection that gets validated using a relationship between processors. A connection is where FlowFiles are temporarily held between two connected NiFi processors. Whenever you create a connection, you have to select one or more relationships between those processors. The overall size of data in a connection is controlled by the configured Back Pressure Object Threshold and Back Pressure Data Size Threshold settings you can define per connection.	
Data flow	CFM data flows are built from processors. Each processor has output relationships such as success or failure. You can connect these to the appropriate processing elements to create your custom data flow.	

Data flow template	It is a NiFi component that helps you to combine basic NiFi data flow building blocks (processor, funnel, input/output port, process group, and remote process group) into larger building blocks. A template can be dragged to the canvas and used when creating a new data flow, or can be exported as an XML file and shared with others. Templates received from other users can be imported into NiFi.	
Process group	It is a NiFi component that enables you to group together different data flows into a logical construct. Process group can make data flows more understandable from a higher level. A process group can be based on different projects, teams, or organizations.	
Remote process group	It is a NiFi component similar to process groups. However, the remote process group references a remote instance of NiFi.	
Input port	It is a NiFi component that is used to get data from a processor, which is not present in that process group.	
Output port	It is a NiFi component that is used to transfer data to a processor, which is not present in that process group.	
Funnel	It is a NiFi component that is used to combine the data from many connections into a single connection.	
Label	It is a NiFi component that is used to add text about any component on the NiFi canvas. You can choose from a range of colors to add aesthetic sense.	
Flow controller	It maintains the knowledge of how processes connect and manages the threads and allocations that all processes use. The flow controller is the brains of the operation. It acts as the broker facilitating the exchange of FlowFiles between processors. It provides threads for extensions to run on, and manages the schedule of when extensions receive resources to run.	
Bucket	A logical container that stores and organizes versioned items/resources (for example, flows) in NiFi Registry.	
Versioned flow	It is a process group placed under version control with the help of NiFi Registry. Each versioned flow has a name, a description, and one or more 'snapshots' (versions).	
Versioned flow snapshot	It means a single version of a versioned data flow. Each snapshot (or version) has metadata and content. The metadata contains a version number, a commit message, an author name, and a commit date. The	

	content is the representation of the workflow itself when it has been committed.	
Event	Events represent the change in FlowFiles while traversing through a NiFi data flow. These events are tracked in data provenance.	
Data provenance	NiFi keeps granular information about each piece of data it handles. Each point in a data flow where a FlowFile is processed in some way is considered a 'provenance event'. Various types of provenance events can occur, depending on the data flow design. As the data is processed through the system, all historical information on what happened to a particular data object (FlowFile) is stored in NiFi's Provenance Repository.	

Related Information[Apache NiFi User Guide](#)[Apache NiFi Registry User Guide](#)

Set up Cloudera Flow Management

If you want to create and manage your flows with Cloudera Flow Management (CFM), download and install NiFi and NiFi Registry. Following the step-by-step instructions will help you get started with CFM.

Download NiFi and NiFi Registry

You can download NiFi and NiFi Registry from the CFM repository. Review the repository Download Locations to identify the correct download link suitable for your operating system and operational objectives.

Related Information[Download from the CFM Repository](#)

Install NiFi and NiFi Registry

Check the Cloudera Flow Management deployment workflow with the different deployment scenarios, and installation instructions in the CFM Deployment documentation.

Related Information[Understand your deployment scenario](#)

Start NiFi and NiFi Registry

You can access NiFi and NiFi Registry using their web user interfaces (UIs). The UIs are very intuitive and it is easy to start creating data flows with CFM.

The NiFi UI provides a wide variety of options for creating dataflows, as well as visualizing, editing, monitoring, and administering those dataflows. For information on how to use it, see the *Apache NiFi User Guide*.

The NiFi Registry UI displays the available shared resources and provides mechanisms for creating and administering users/groups, buckets, and policies. For information on how to use it, see the *Apache NiFi Registry User Guide*.

Related Information

[Apache NiFi User Guide](#)

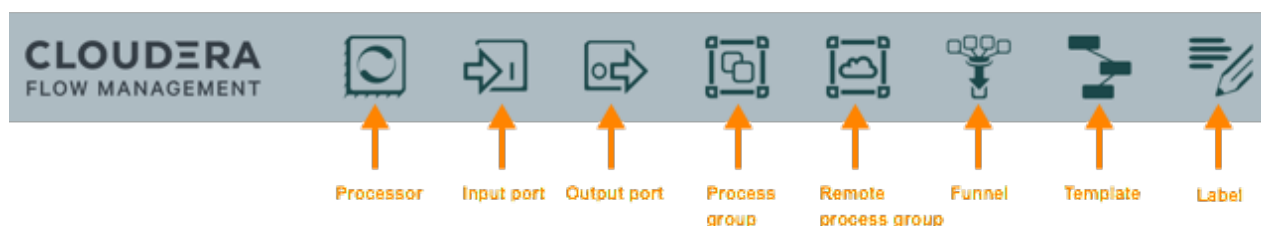
[Apache NiFi Registry User Guide](#)

Build your first data flow

You can use Apache NiFi to create data flows for any protocols or data sources. Start your flow with a listener component, which gets the data from a source system. On the other end of the flow, add a transmitter component that writes the data to a target system. Additionally, you can add other components in between that process the data.

NiFi components

Apache NiFi components are the building blocks of data flows. The NiFi Component toolbar contains several data flow components that you can use when creating a new flow.



Create a data flow

About this task

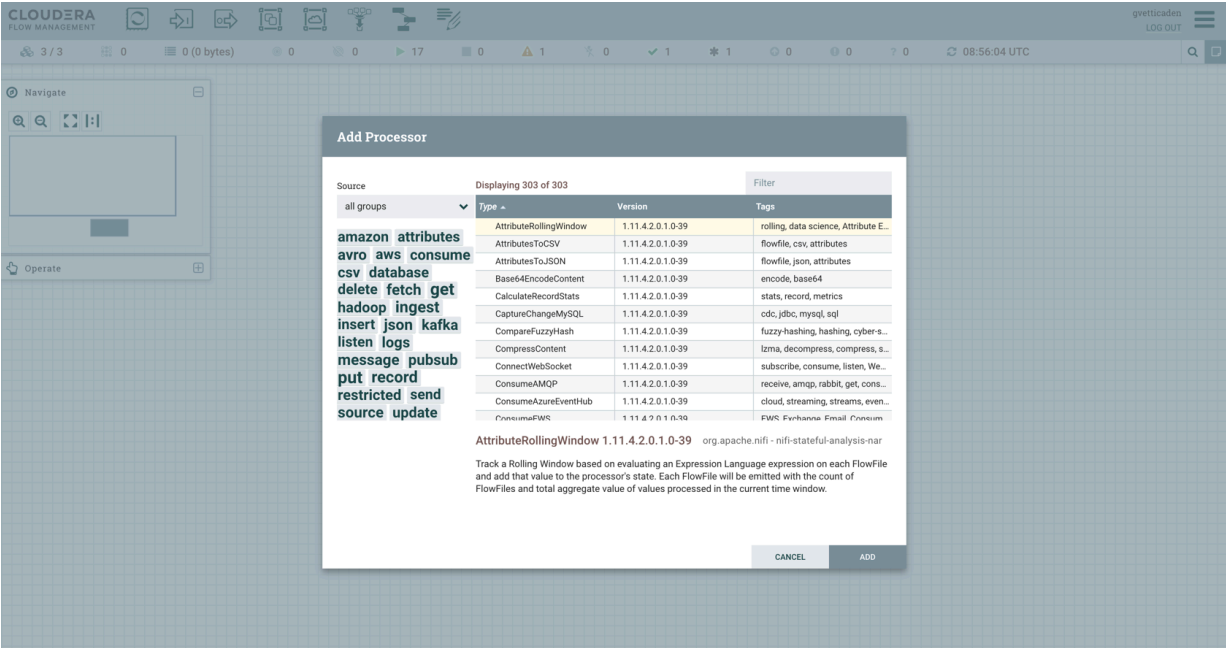
This section walks you through the process of creating NiFi data flows from scratch using basic flow options.

For information about building advanced data flows (using parameters, custom properties, controller services, or reporting tasks), see the *Apache NiFi User Guide*.

Procedure

1. Add data flow components.

You can add processors and other components to build your data flow as you need. To add a processor to your flow, drag the processor icons to the canvas from the component toolbar in the top-left of the screen. For more information, see *Adding Components to the Canvas* in the *Apache NiFi User Guide*.




2. Configure the processors.

- Right-click the processor on the canvas and choose Configure in the context menu. Alternatively, double-click on the processor.

A configuration dialog with the following tabs is displayed: Settings, Scheduling, Properties, and Comments.

- Configure the processor according to the behavior you desire in your data flow.
- When you have finished configuring the options you need, click Apply to save the changes or cancel all changes by clicking Cancel.

Configure Processor

 Invalid

SETTINGS


SCHEDULING

PROPERTIES

COMMENTS

Name
RouteOnAttribute


☒ Enabled


Automatically Terminate Relationships 
☐ unmatched
FlowFiles that do not match any user-defined expression will be routed here



Id
dd219e19-81a7-1394-0000-00007022e5c2

Type
RouteOnAttribute 1.11.4.2.0.1.0-39

Bundle
org.apache.nifi - nifi-standard-nar

Penalty Duration 
30 sec

Yield Duration 
1 sec

Bulletin Level 
WARN 

CANCEL

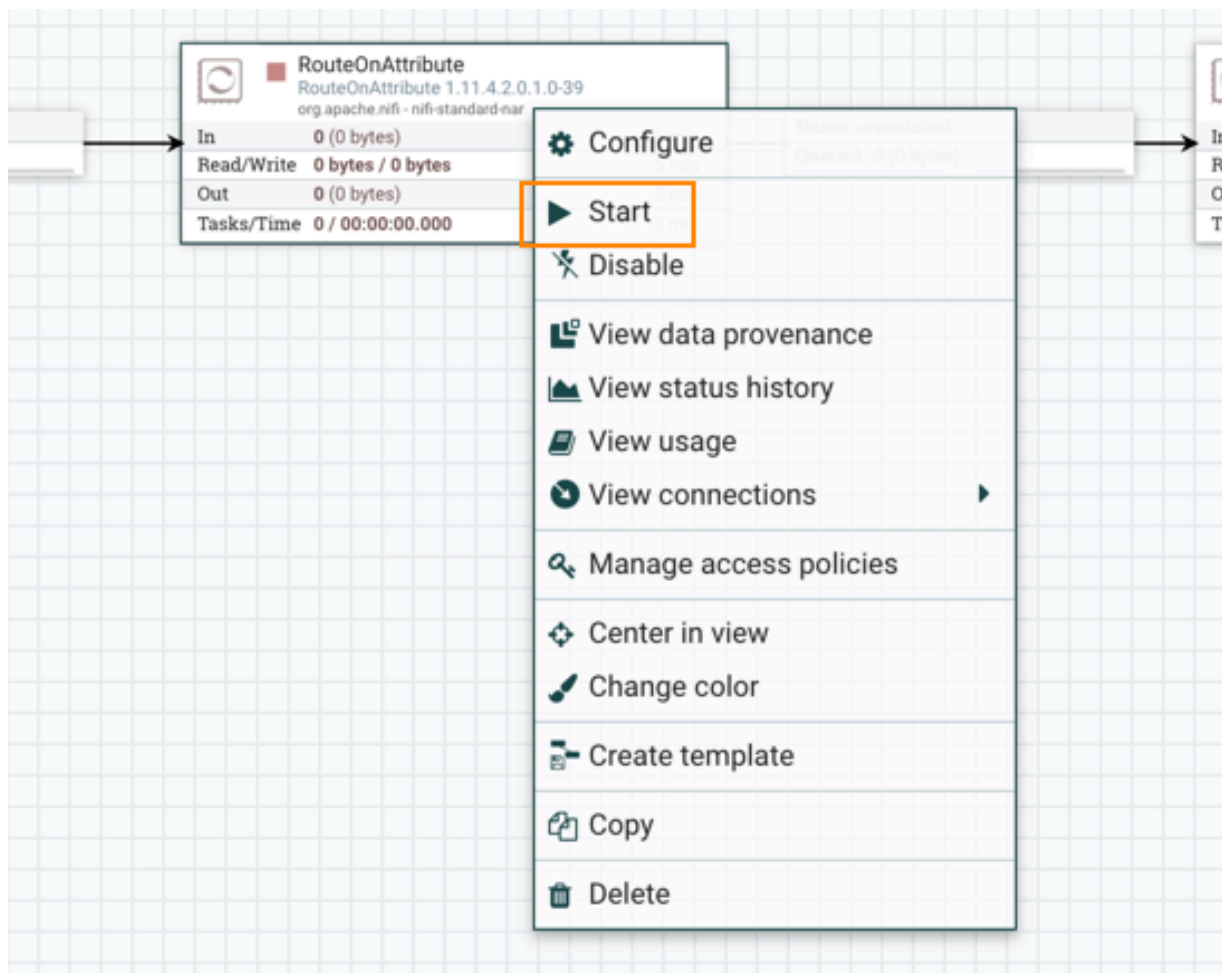
APPLY

3. Connect data flow components.

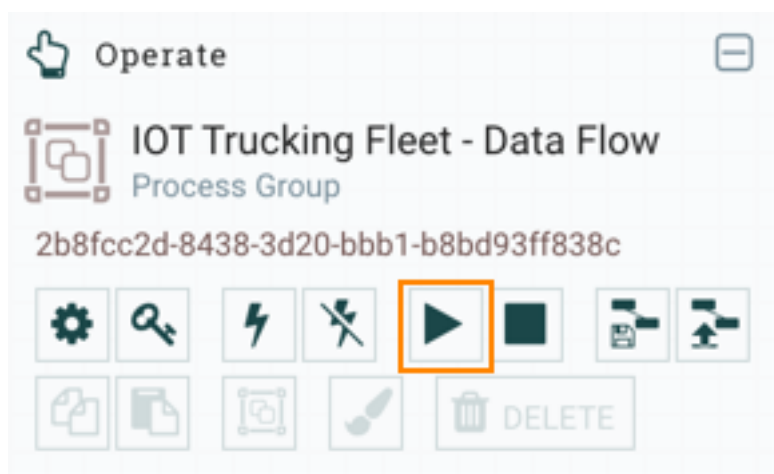
You can connect the data flow components by hovering over a component and dragging the connection icon to another component. You can then choose which relationships you want to include for this connection.

4. Start your data flow.

When all components are connected, the flow is still not working. To start the processors, click each one individually and then right-click and choose the Start menu item.



Alternatively, you can select all processors and then click the Start icon on the Operate palette.



After you start the data flow, the icon in the top-left corner of the processors changes from a stopped icon to a running icon. You can stop the processors by using the Stop icon on the Operate palette or the Stop context menu item.

Results

By following the above steps, you can choose any processor and other NiFi components to create a flow suitable for your purpose.

Related Information

[Apache NiFi User Guide](#)

Learn about NiFi processors

Reviewing the list of available NiFi processors can help you better understand the capabilities of the out-of-the-box processors.

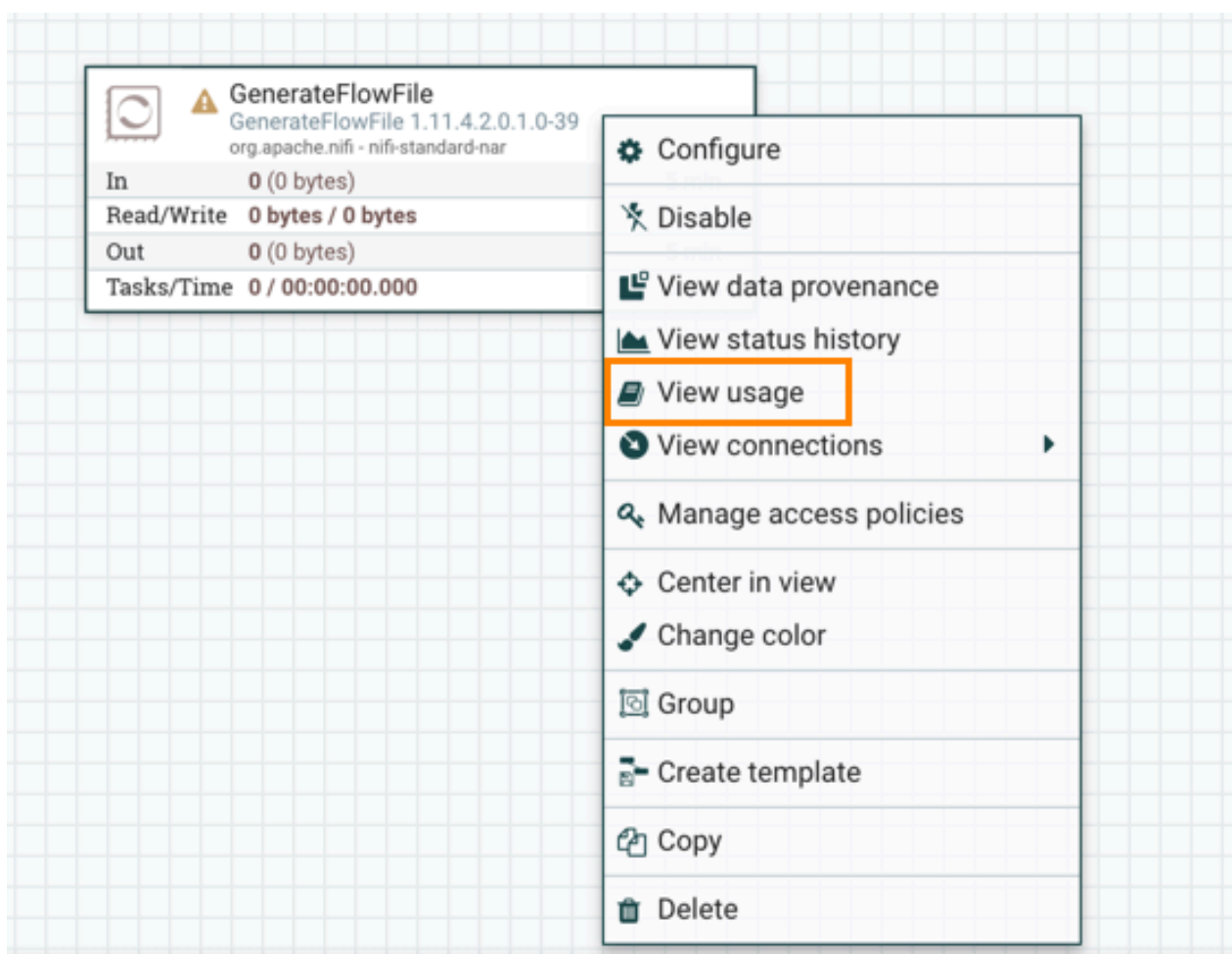
See what processors are available

To create an effective data flow, you must understand what types of processors are available and what they are suited for. NiFi contains many processors that are available out of the box. These provide capabilities to ingest data from different systems, route, transform, process, split, and aggregate data and also distribute data to many systems. The number of processors increases in almost every Apache NiFi release.

For the complete list of NiFi processors available in this release, see the list of *Supported NiFi Processors*.

Get more information for a processor

Each processor has multiple properties and relationships, and it can be challenging to remember how all of the different pieces work for the processor. You can review a processor's usage information by right-clicking the processor and choosing the View usage menu item.



Work with attributes

Each FlowFile is created with several attributes, and these attributes change over the life of the FlowFile. The concept of a FlowFile provides three primary benefits:

- It enables you to make routing decisions in the flow so that FlowFiles that meet certain criteria can be handled differently than other FlowFiles. This is done using the RouteOnAttribute and similar processors.
- Attributes are used to configure processors in such a way that the configuration of the processor is dependent on the data itself. For instance, the PutFile processor uses the attributes to know where to store each FlowFile, while the directory and filename attributes may be different for each FlowFile.
- Attributes provide valuable context about the data. This is useful when reviewing the provenance data for a FlowFile. You can search for provenance data that match specific criteria, and it also enables you to view this context when inspecting the details of a Provenance Event. By doing this, you gain valuable insight as to why the data was processed one way or another, by glancing at this context that is carried along with the content.

Each FlowFile has a minimum set of attributes called Common Attributes:

filename

A filename that can be used to store the data to a local or remote file system.

path

The name of a directory that can be used to store the data to a local or remote file system.

uuid

A Universally Unique Identifier that distinguishes the FlowFile from other FlowFiles in the system.

entryDate

The date and time at which the FlowFile entered the system (that is, was created). The value of this attribute is a number that represents the number of milliseconds since midnight, Jan. 1, 1970 (UTC).

lineageStartDate

Any time that a FlowFile is cloned, merged, or split, this results in a 'child' FlowFile being created. As those children are then cloned, merged, or split, a chain of ancestors is built. This value represents the date and time at which the oldest ancestor entered the system. Another way to think about this is that this attribute represents the latency of the FlowFile through the system. The value is a number that represents the number of milliseconds since midnight, Jan. 1, 1970 (UTC).

fileSize

This attribute represents the number of bytes taken up by the FlowFile's content.



Note: The uuid, entryDate, lineageStartDate, and fileSize attributes are system-generated and cannot be changed.

To extract attributes, add user-defined attributes, route FlowFiles based on their attributes, or perform other complex tasks with the help of attributes, see *Working With Attributes* in the *Apache NiFi Getting Started Guide* for details.

Related Information

[Supported NiFi Processors](#)

[Apache NiFi Getting Started Guide](#)

Work with data flow templates

Templates make it easier to reuse and distribute NiFi flows or specific parts of data flows.

About this task

Besides reducing development time, reuse also supports the adoption of standards and patterns in data flow design. Data flow templates can help you to build a library of reusable elements that you can use as pre-built components in your new flows. Templates can be made from any subset of a flow: a single processor, a flattened subflow, or a process group holding a subflow. Alternatively, templates can be made from full flows as well. For more in-depth information about using templates, see the *Apache NiFi User Guide*.

Related Information

[Apache NiFi User Guide](#)

Create a template

Learn how to create data flow templates that you can use as building blocks in another data flow.

Procedure

Follow these steps to create a template from a reusable subflow:

1. Select the components/flow elements you want to include in the template.
2. Click the Create Template icon on the Operate palette.
3. Provide a name and optionally a description for the template.
4. Click Create.

Download a template

You can export templates from Apache NiFi and use them as a starting point for developing other data flows. You can retain or change all configurations of the downloaded template.

About this task

Templates can be exported from NiFi and used as a starting point for developing a data flow. All configurations of the downloaded template are retained by default and can also be changed.

Procedure

To download a template from NiFi:

1. Go to the Templates option in the Global Menu at the top-right hand corner of the NiFi UI.
2. Click the download icon of the template you want to download.

Results

The template XML file is downloaded.

Upload a template

You can import a data flow template to Apache NiFi and use it as a starting point for developing a data flow. When using the template, you can retain or change all of its configurations.

About this task

Templates can be imported to NiFi and used as a starting point for developing a data flow. All configurations of the downloaded template are retained and can also be changed.

Procedure

To upload a template into NiFi:

1. Click the Upload Template icon from the Operate palette.
2. Click the Search icon in the Upload template window and navigate to the XML file you want to upload.
3. Select the file.
4. Click Upload.

Results

The template is uploaded and displayed in the templates table in the NiFi UI. You can now use it to create data flows.

Use a template

Learn how you can build data flows with the help of templates in Cloudera Flow Management (CFM).

About this task

When you have created or imported templates, you can use these as building blocks in various data flows, just as you use a any data flow component.

Procedure

To use a template in a data flow:

1. Click and drag the Template icon from the Component toolbar to the canvas.
The Add template window containing a list of all available templates is displayed.

2. Choose the template that you would like to add to the canvas and click Add.

There are a certain important points to remember when working with templates:

- Sensitive properties (for example, a password configured in a processor) is not added to the template. These sensitive properties have to be populated each time that the template is used.
- If a component that is included in the template references a controller service, the controller service is also added to the template. This means that each time that the template is used in a data flow, it creates a copy of the controller service.

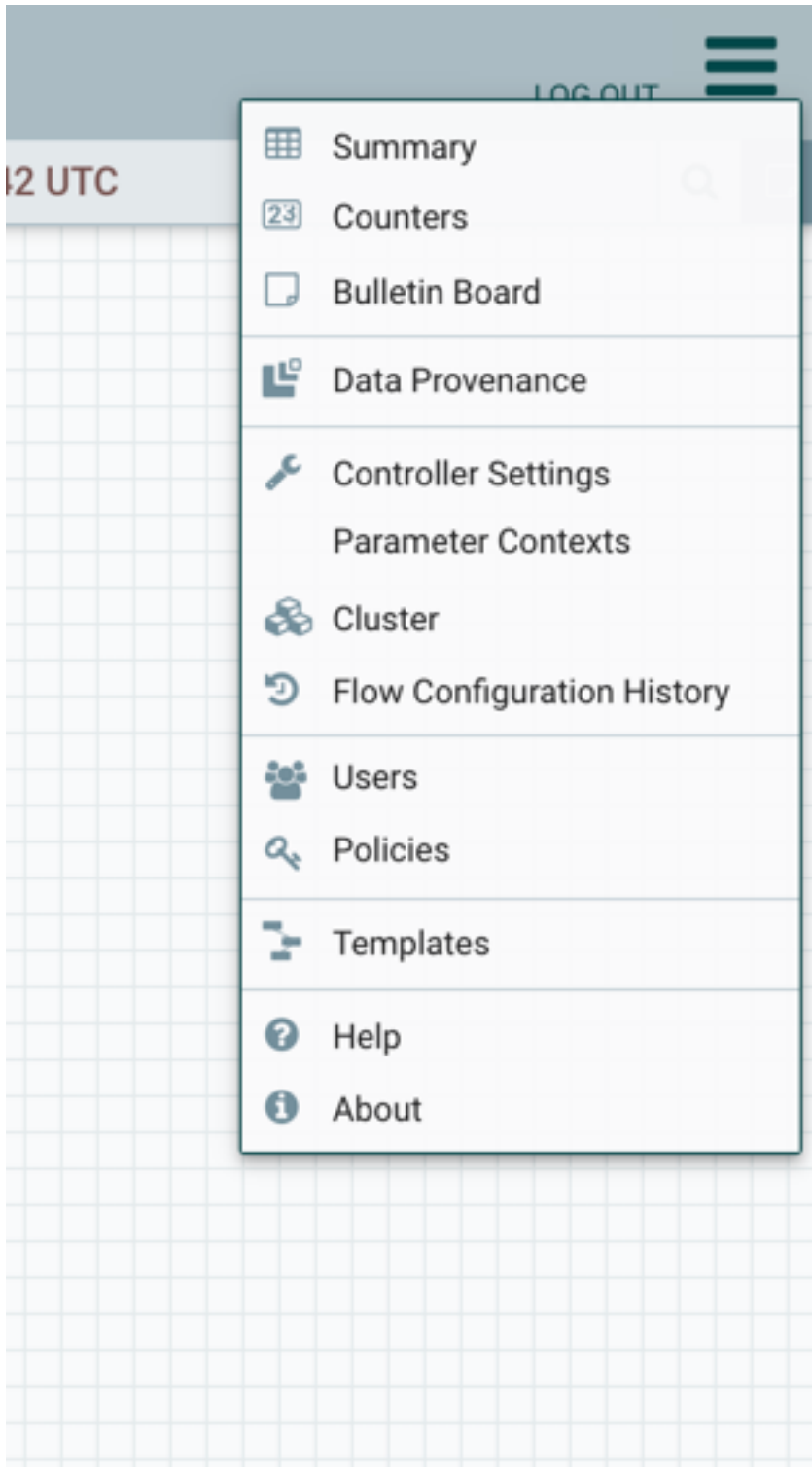
Manage templates

You can access and work with your templates using the NiFi Templates dialog.

Procedure

To open the dialog that displays all available templates with the ability to filter, export, and delete them:

1. Select Templates from the Global Menu.



The NiFi Templates window is displayed.

2. Check the list of available templates. You can filter the list by name or by description to find the template of your interest.

- On the right side of the table you can find the Download icon that you can use to download a template as an XML file.

NiFi Templates ✕

Displaying 2 of 2

Filter by name

Date/Time	Name	Description	Process Group Id	
05/27/2020 13:28:37 UTC	CFM-Trucking-Ref-App-Data-SECURE-SASL-SSL-LDAP-CDPDC.xml	Empty string set	52f66ff8-0172-1000-322f-9d9d8ad1adf5	
05/27/2020 03:34:56 UTC	CFM-Trucking-Ref-App-Data-With-Hive-Streaming-Flow-SECURE-SASL-SSL-LDAP-CDPDC.xml	Empty string set	52f66ff8-0172-1000-322f-9d9d8ad1adf5	

Connect NiFi to NiFi Registry

Learn about the steps to configure NiFi to identify the NiFi Registry so that you can support versioned flows in Cloudera Flow Management.

Before you begin

- You have added and configured NiFi and NiFi Registry.
- You have started both NiFi and NiFi Registry.

Procedure

- From the NiFi UI, select Controller Settings from the Global Menu.
- Select the Registry Clients tab.
- Click the Add icon (+) to launch the Add Registry Client dialog.
- Add the name and URL location of the NiFi Registry service.

Results

NiFi Settings

GENERAL	REPORTING TASK CONTROLLER SERVICES	REPORTING TASKS	REGISTRY CLIENTS
Name	Location	Description	
NiFi Registry	https://cfm-node-3.cloudera.com:18433		

Create a bucket in NiFi Registry

You can store and organize versioned data flows in NiFi Registry buckets. You can create several buckets mapping to your categorization and grouping needs. You can use these Registry buckets to organize your versioned flows according to different teams, deployment environments, usage patterns, or different business units.

Procedure

- Open the Registry UI.
- Click the wrench icon (Settings) in the top right corner of the screen.
- Click New Bucket in the Buckets window.
- Enter a name for your bucket.
- Click Create.

Results

You can see that a new bucket is visible in the NiFi Registry.

For information on how to manage Registry buckets, see the *Apache NiFi Registry User Guide*.

Related Information

[Apache NiFi Registry User Guide](#)

Version data flows

When NiFi is connected to a NiFi Registry, data flows can be version controlled at the process group level. Learn how you can work with different flow versions in NiFi and save them in NiFi Registry.

Before you begin

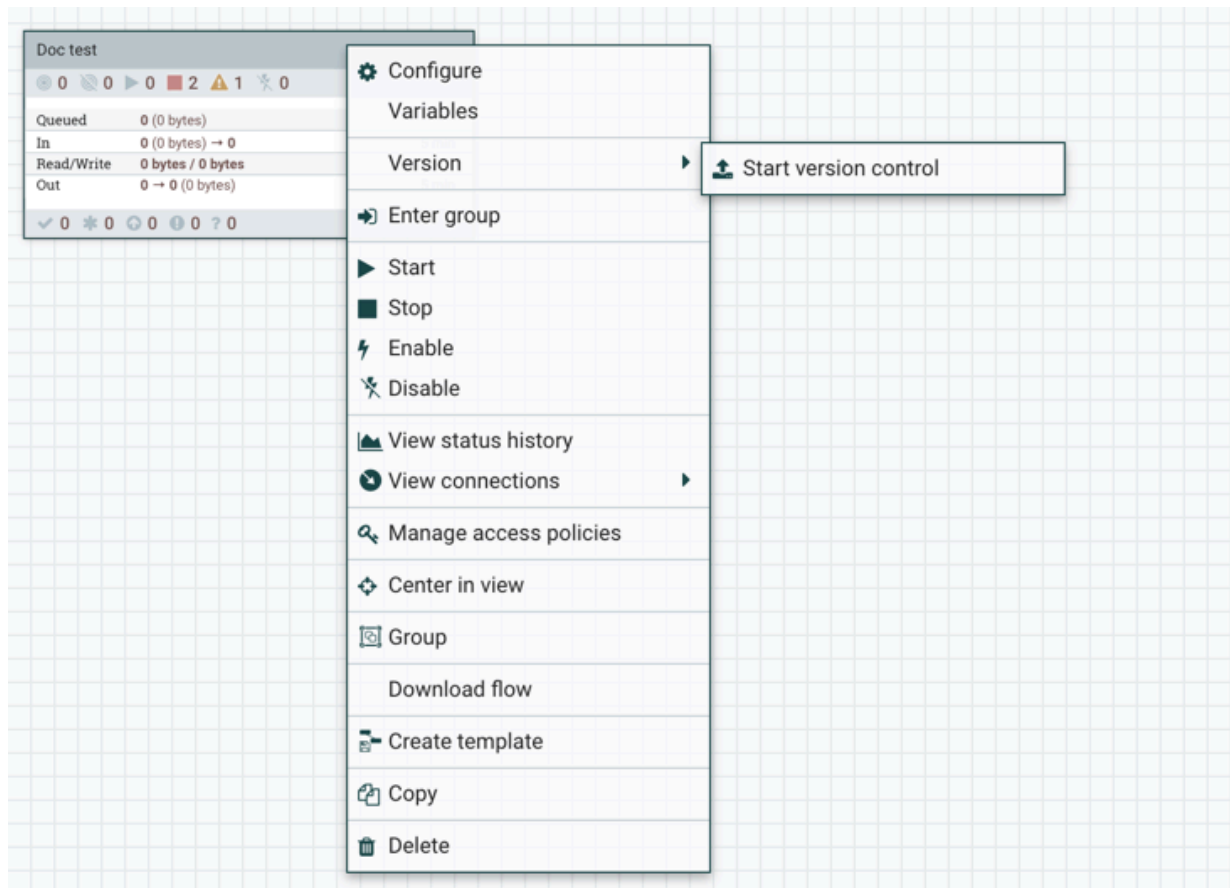
- You must have a process group already created in NiFi.
- You have connected NiFi to NiFi Registry.
- You have created a bucket in Registry for storing and organizing your NiFi data flows.

Start version control

Learn how you can place a process group under version control in NiFi.

Procedure

1. Right-click the process group, then click VersionStart version control.



2. Specify the NiFi Registry bucket where you want to save your versioned data flow.

3. Set the flow name, flow description, and version comments.

Save Flow Version

Registry
NiFi Registry

Bucket
Doc test

Flow Name
1

Flow Description

Version Comments

CANCEL SAVE

4. Click Save.

Results

Your flow version is saved and is ready to be deployed. The process group has a version state icon in the upper-left corner. If it is a green checkmark, it indicates that the process group is saved as a versioned flow in the Registry.

Commit local changes

Learn how you can manage local changes made to your versioned data flow in NiFi.

About this task

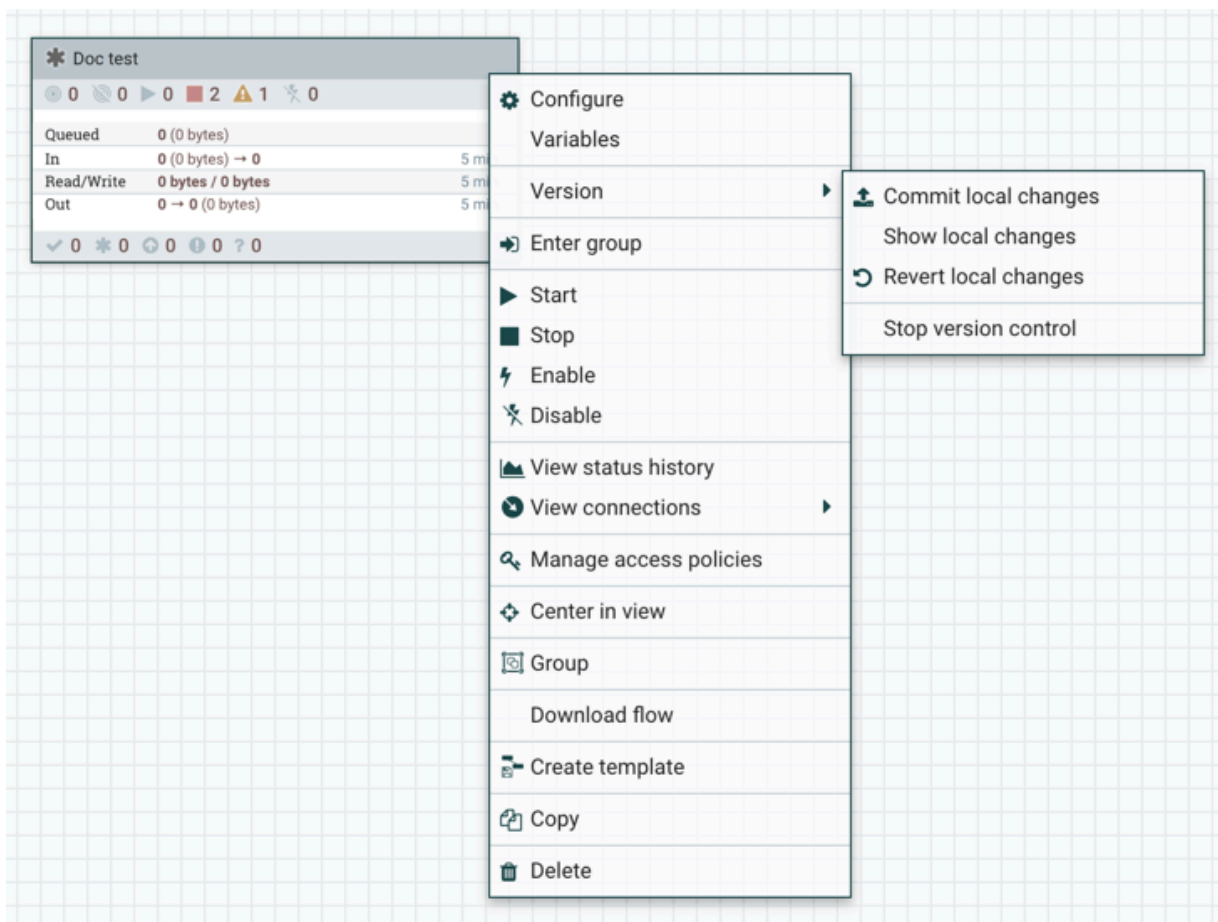
When you make changes in a data flow, the changes are local at first. If you want these changes to be saved as a separate flow version, you have to commit these changes to the NiFi Registry.

If you make changes to a versioned process group, the green checkmark icon changes to a grey asterisk because of the uncommitted local edits.

Procedure

1. Right-click the process group or the canvas in the workflow window and choose Version.

The context menu shows that you can save, review, or revert your changes.



- Select Commit local changes if you want to save the changes as a new version of the data flow.
 - Select Show local changes if you want to see the details of the edits.
 - Select Revert local changes if you want to discard the edits.
2. Select Commit local changes.
The Save Flow Version pop-up window appears, where you can add comments to the version to be saved.
 3. Click Save.

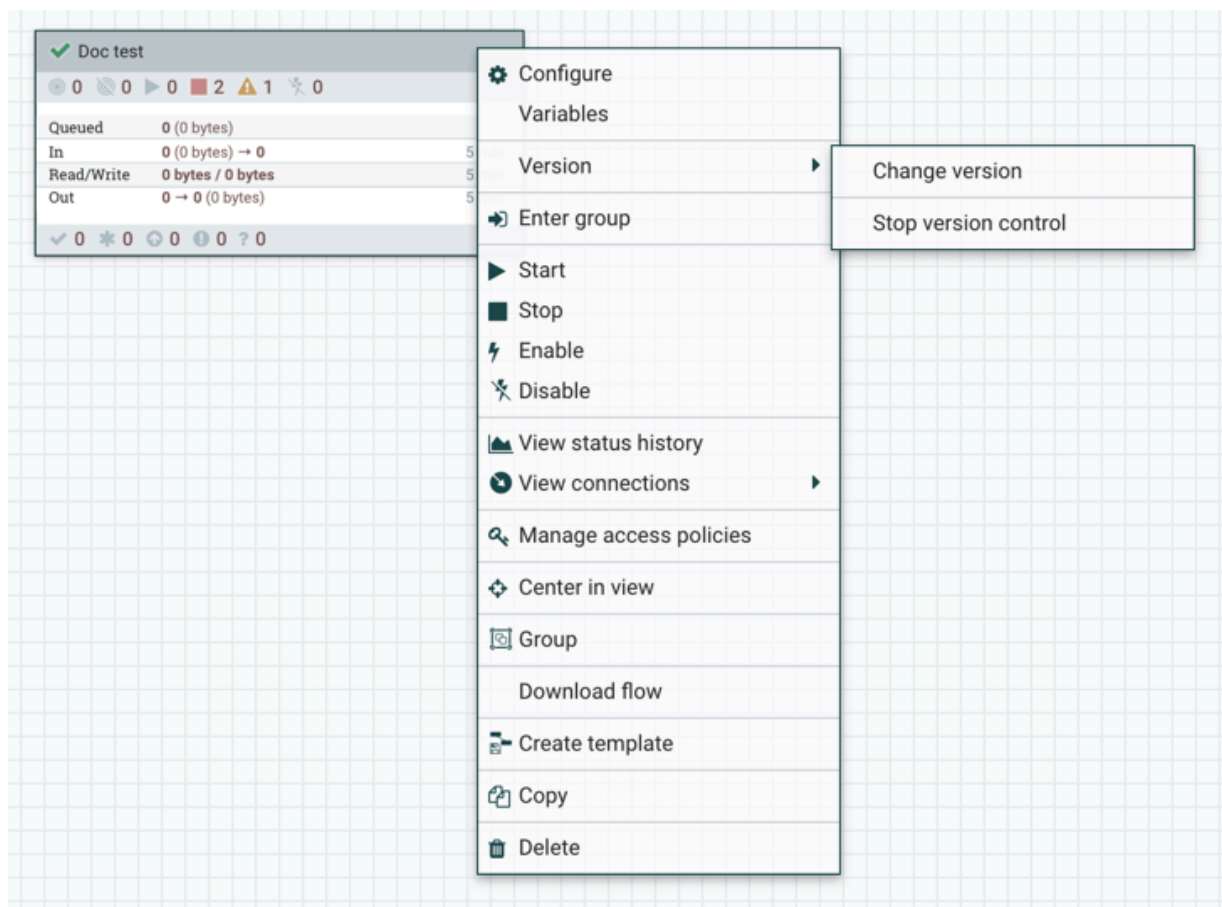
Change the version of your data flow

If you have multiple versions of a data flow saved in NiFi Registry, you can go back to any of the versions. Learn how you can change the versions of your data flows in NiFi.

Procedure

To open a different flow version:

1. In the NiFi UI, right-click the process group and go to VersionChange version to open the list of available versions.



2. Select the version.
3. Click Change.

Results

Your flow version is changed.

Manage versioned data flows

Learn how you can work with the data flows stored in your NiFi Registry buckets.

Procedure

- You can view your versioned flows in NiFi registry in a single list or by bucket.
- You can sort and filter the flows according to different criteria.
- You can also delete a versioned flow from the registry.
- You can check the changelog of each flow for details on its versions. You can get information on who made a change and when the change was made. If You can also take some actions on the version flow.

For detailed instructions on how to manage flows, see the *Apache NiFi Registry User Guide*.

Related Information

[Apache NiFi Registry User Guide](#)

Monitor data flows

As data travels through your data flows in the Cloudera Flow Management system, it is important to understand how these data flows are performing. Monitoring information can help you to assess if a data flow requires more resources. Learn how you can check and observe your flows in Cloudera Flow Management.

There are multiple ways to monitor the NiFi flows to check on errors, memory usage, CPU usage, data flow statistics, and so on.

Status bar

The Status bar provides information about the overall health of the system. You can find it at the top of the NiFi UI, under the component toolbar. It contains important statistics about the current health of NiFi. The number of active threads can indicate how hard NiFi is working, and the queued statistics indicate how many FlowFiles are queued across the entire flow, and the total size of those FlowFiles.

If the NiFi instance is in a cluster, we will also see an indicator here telling us how many nodes are in the cluster and how many are currently connected. In this case, the number of active threads and the queue size are indicative of the sum of all nodes that are currently connected.

Component statistics

Each processor, process group, and remote process group on the canvas provides fine-grained statistics about their operations. Connections and process groups provide information about the amount of data in their queues. Processors provide information on how much data they processed in the last five minutes. This is a rolling window and allows you to see things like the number of FlowFiles that have been consumed by a processor, as well as the number of FlowFiles that have been emitted by the processor. The connections between processors also expose the number of items that are currently queued.

To see historical values for these metrics, right-click on a component and choose the View Status History menu item. This shows you a graph that spans the time since NiFi was started, or up to 24 hours, whichever is less. You can extend or reduce the amount of time that is shown by changing the configuration in the properties file.

Queue interaction

You can view the FlowFiles queued in a connection when necessary. Right-click a connection, and click List queue in the context menu to open the queue listing. You can check the list even when the source and destination processors are actively running.

Additionally, details for a Flowfile in the listing can be viewed by clicking Details in the left-most column. From here, the FlowFile details and attributes are available along with the buttons for downloading or viewing the content. If the source or destination of the connection is actively running, there is a chance that the desired FlowFile is no longer in the active queue.

You can delete FlowFiles queued in a connection. Right-click the connection, and click Empty queue in the context menu to initiate the removal of the FlowFiles. You can remove FlowFiles even if the source and destination processors are actively running.

Bulletins

If a processor logs anything as a WARNING or ERROR, you see a Bulletin Indicator displayed in the top-right corner of the processor. This indicator resembles a sticky note and is displayed for five minutes after the event occurs. Hovering over the bulletin provides information about the incident so that you do not have to go through log messages to find it. You can also change the log level at which bulletins occur in the Settings tab of the Configure dialog for a processor.

When the framework generates a bulletin, a bulletin indicator is highlighted at the top-right of the screen. The bulletin board shows the latest WARNING and ERROR messages that are generated by NiFi processors in real time. To

access the bulletin board, click Bulletin Board in the Global Menu at the top-right corner of the NiFi UI. Clicking this option takes you to the bulletin board where you can see all bulletins that occur across the NiFi instance and can filter based on the message, name, ID, and group ID. It refreshes automatically and you can also disable it. You can navigate to the component with a bulletin by clicking the component UUID.

NiFi summary page

NiFi Summary provides information about all of the components on the canvas in a tabular format. It also provides system diagnostics that include disk usage, CPU utilization, Java heap, and garbage collection information. To access the summary page, click Summary in the Global Menu at the top-right corner of the NiFi UI. Clicking this option takes you to the summary page, where you can get information about all the components of that particular NiFi instance or cluster. You can monitor the following components in the NiFi summary UI:

- Processors
- Input ports
- Output ports
- Remote process groups
- Connections
- Process groups

Depending on the component, you can filter by name, type, or UIR.

For more in-depth monitoring information, see the *Apache NiFi User Guide*.

Related Information

[Apache NiFi User Guide](#)

Trace Data Lifecycle

You can track down the life of data, its origin, transformation, and destination using data provenance. You can use this information to troubleshoot and evaluate data flow compliance and optimization in real time.

NiFi keeps a very granular level of detail about each piece of data that it ingests. As the data objects are processed through the system, NiFi records and indexes data provenance details, and this information is stored in NiFi's Provenance Repository. By default, NiFi updates this information every five minutes, but that is configurable.

To search and view this information, select Data Provenance from the Global Menu. This enables you to see the most recent data provenance information, search the information for specific items, and filter the search results. It is also possible to open additional dialog windows to see event details, replay data at any point within the dataflow, and see a graphical representation of the data's lineage, or path through the flow.

For more information about data provenance, see the *Apache NiFi User Guide*.

Related Information

[Apache NiFi User Guide](#)