

Getting Started with Cloudera Flow Management

Date published: 2019-06-26

Date modified: 2023-11-02

CLOUDERA

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
Set up Cloudera Flow Management	4
Download NiFi and NiFi Registry	4
Install NiFi and NiFi Registry	5
Start NiFi and NiFi Registry	5
Build your first data flow	5
Connect NiFi to NiFi Registry	9
Create a bucket in NiFi Registry	9
Version data flows	10
Start version control	10
Commit local changes	11
Change the version of your data flow	12
Manage versioned data flows	13
Getting started with Apache NiFi	14
Getting started with Apache NiFi Registry	14

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](#)

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 locations](#)

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

[CFM deployment workflow](#)

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](#)

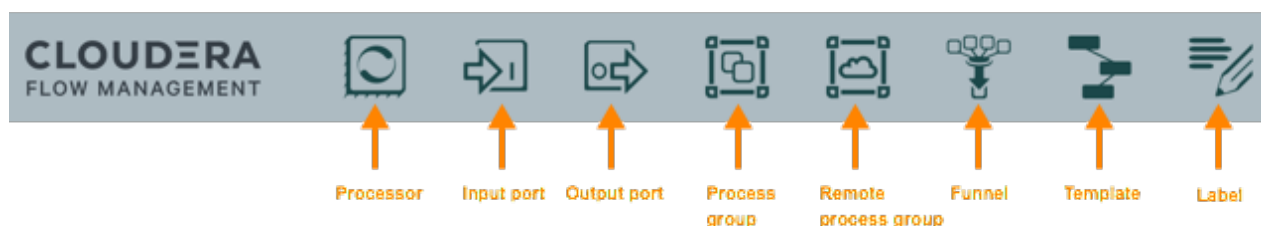
[Apache NiFi Registry User Guide](#)

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

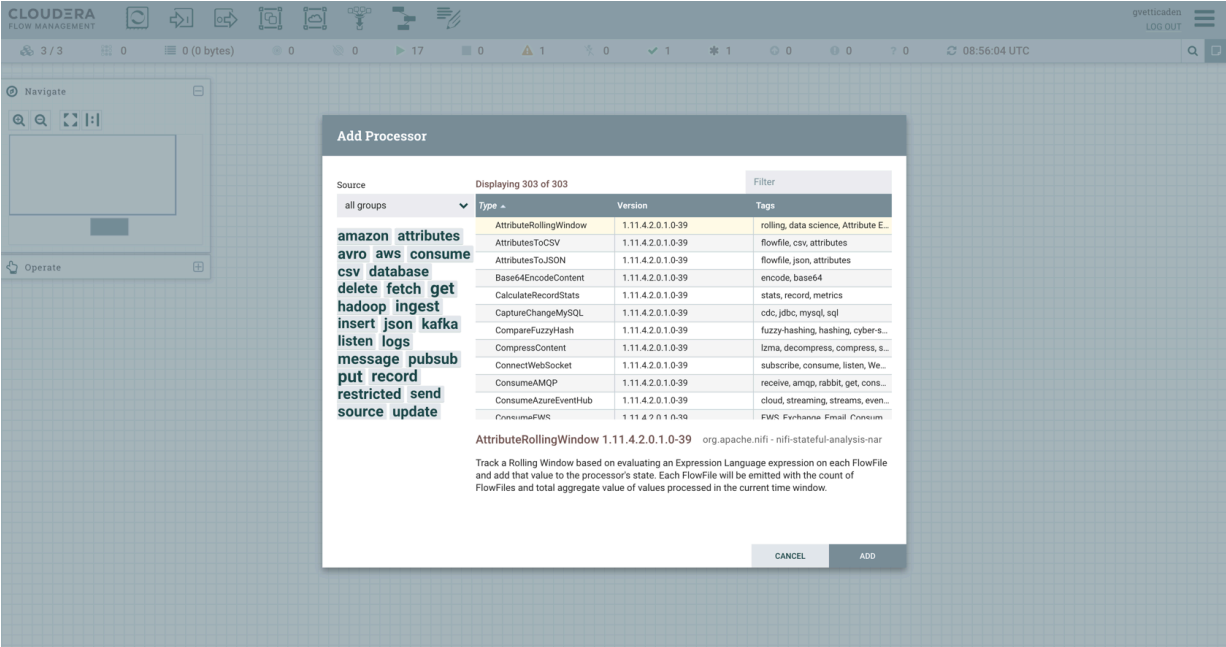
This section walks you through the process of creating NiFi data flows from scratch using basic flow options.

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

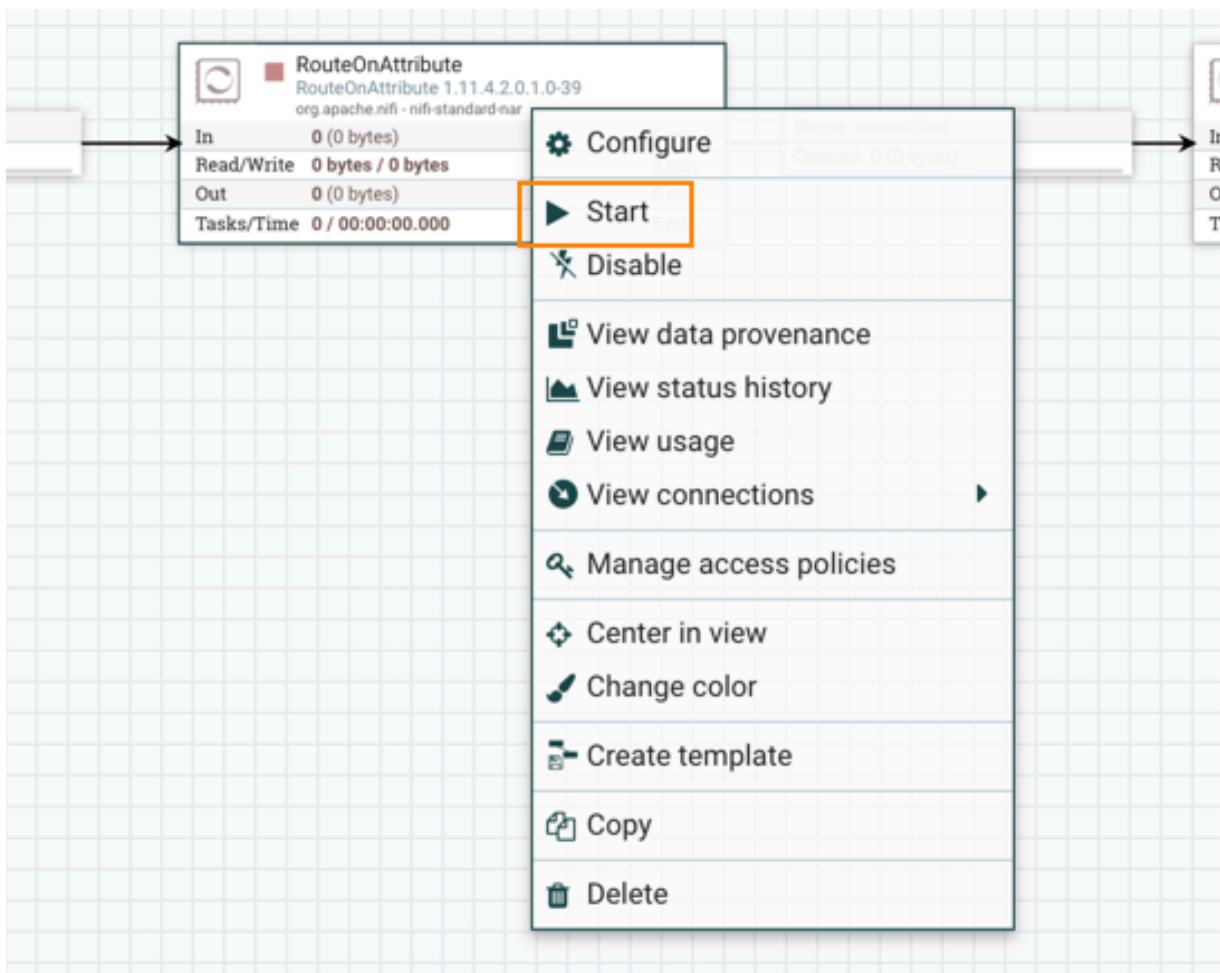
3. Connect data flow components.

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.

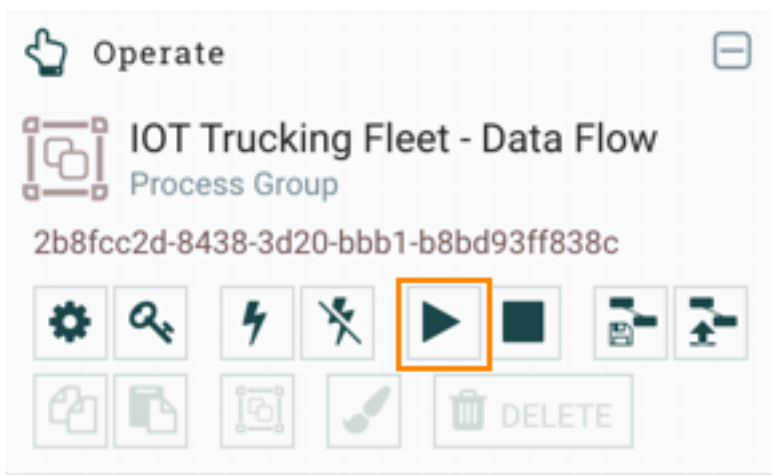
7

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

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

1. From the NiFi UI, select Controller Settings from the Global Menu.
2. Select the Registry Clients tab.
3. Click the Add icon (+) to launch the Add Registry Client dialog.
4. 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

1. Open the Registry UI.
2. Click the wrench icon (Settings) in the top right corner of the screen.
3. Click New Bucket in the Buckets window.
4. Enter a name for your bucket.
5. 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

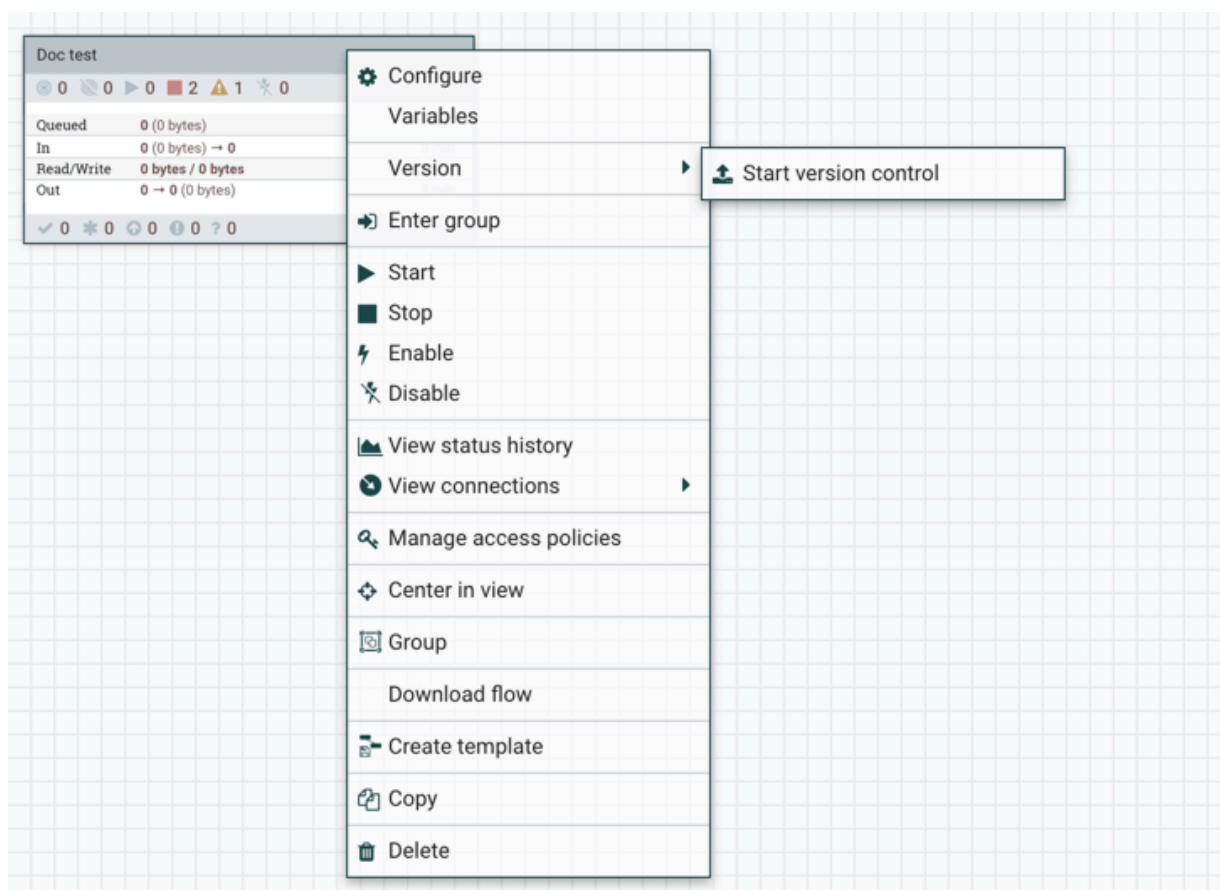
- You must have a process group already created in NiFi.
- You have connected NiFi to NiFi Registry.
- You have created a bucket in Registry for storing and organizing your NiFi data flows.

Start version control

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

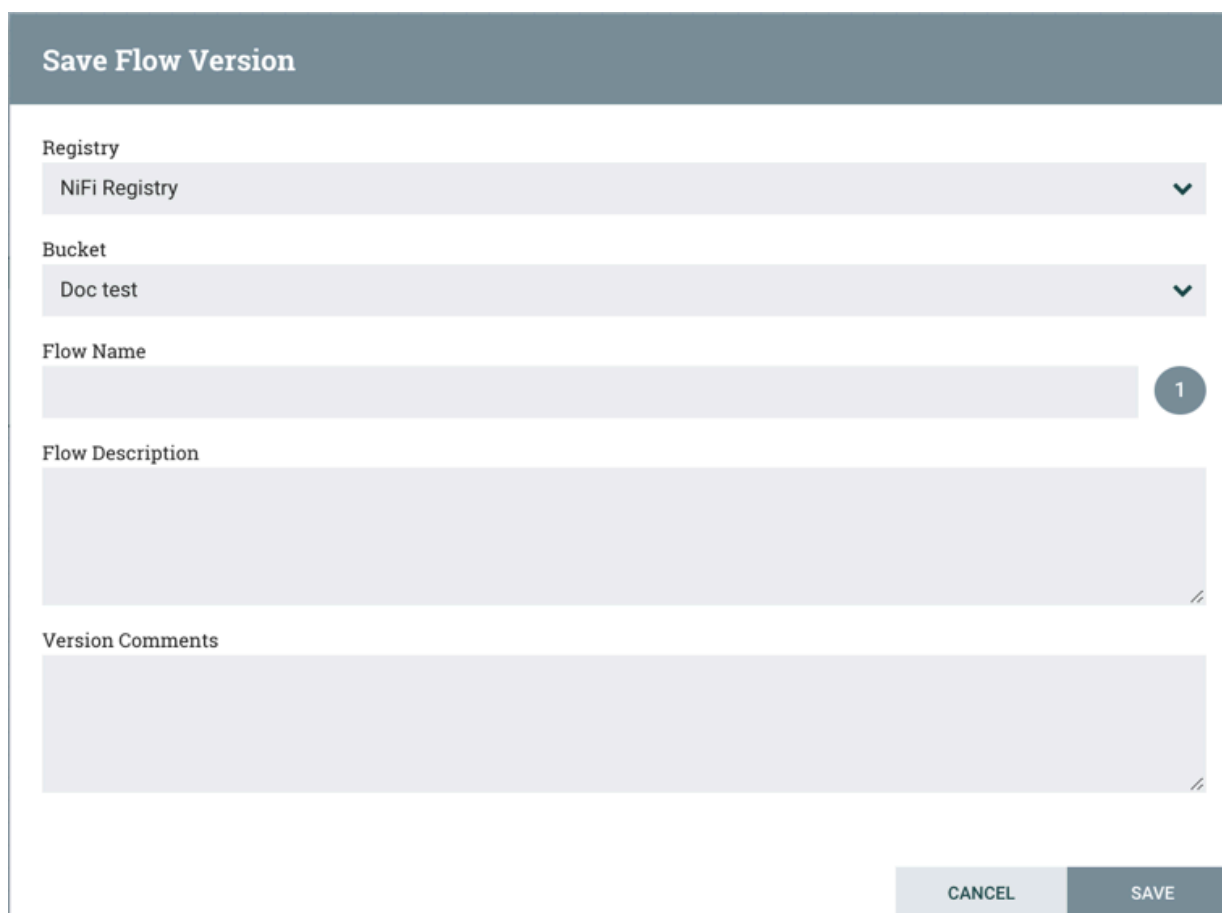
Procedure

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.



The image shows a 'Save Flow Version' dialog box. It has a dark grey header with the title 'Save Flow Version'. Below the header, there are five sections: 'Registry' with a dropdown menu showing 'NiFi Registry'; 'Bucket' with a dropdown menu showing 'Doc test'; 'Flow Name' with a text input field and a circular icon containing the number '1'; 'Flow Description' with a large text area; and 'Version Comments' with a large text area. At the bottom right, there are two buttons: 'CANCEL' and '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

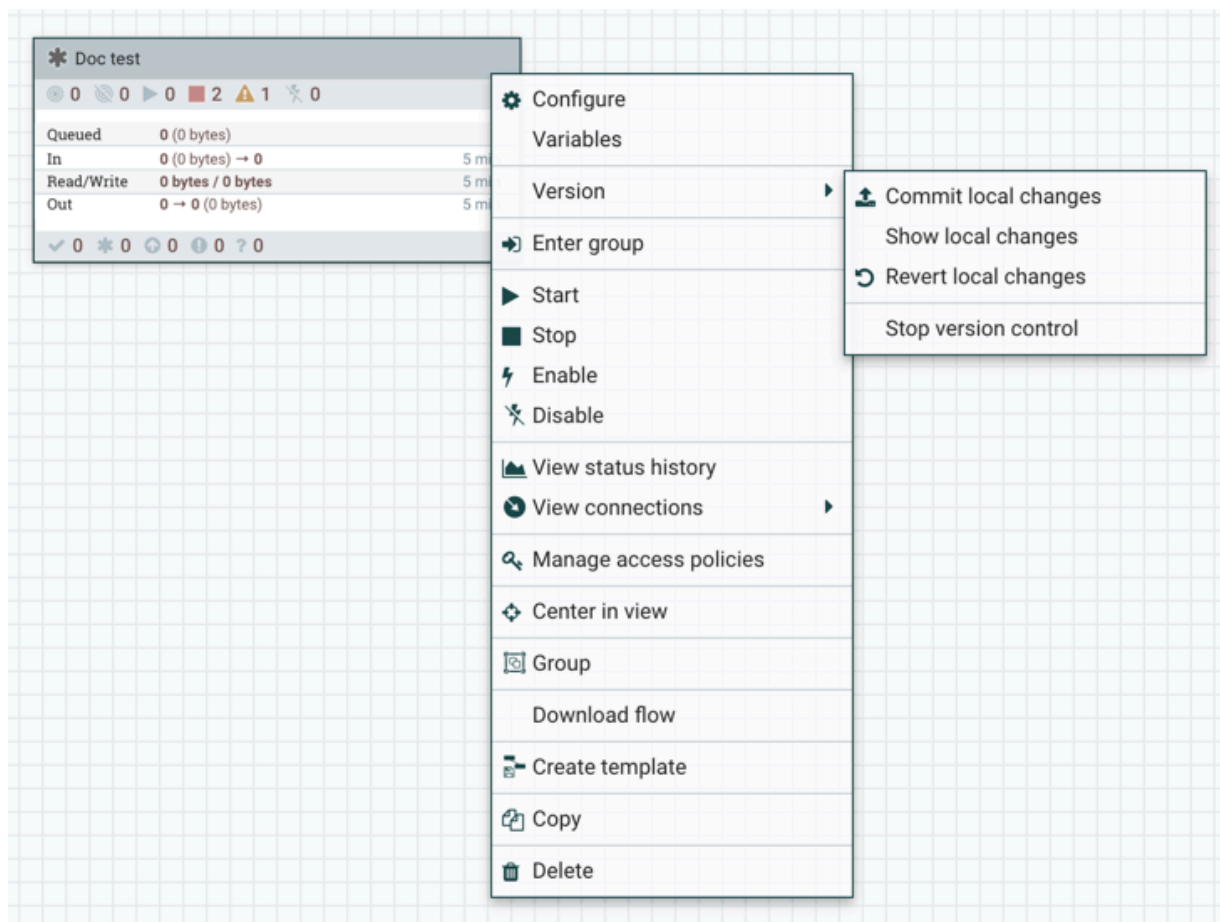
When you make changes in a data flow, the changes are local at first. If you want these changes to be saved as a separate flow version, you have to commit these changes to the NiFi Registry.

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.

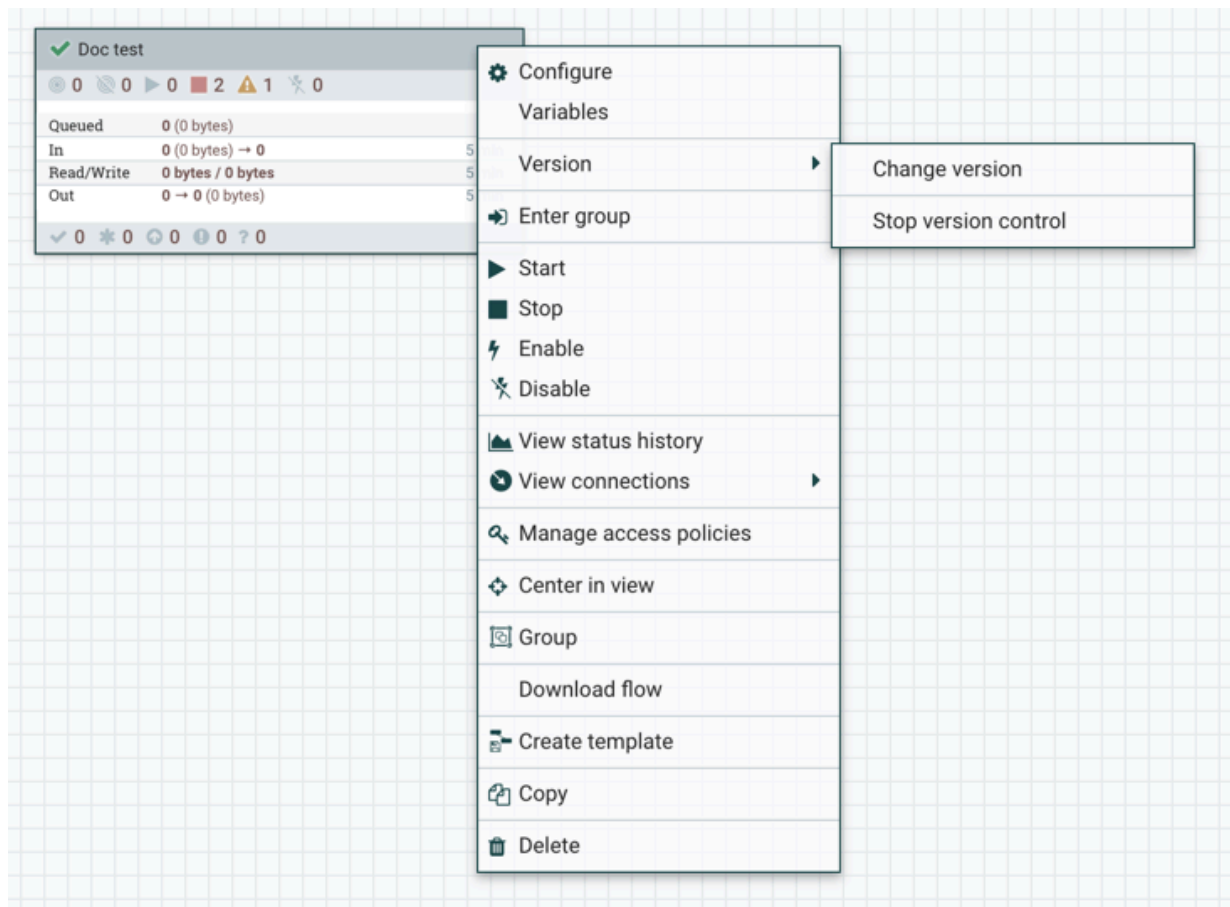
Change the version of your data flow

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](#)

Getting started with Apache NiFi

See the Apache NiFi documentation to access the *Apache NiFi Getting Started Guide*.



Important: The Apache NiFi documentation is available both online and in the NiFi UI. The online documentation always shows the latest version of the NiFi documentation. There may be differences in the content compared to the version of NiFi you are using. You can access the documentation corresponding to the NiFi version you are using directly from the NiFi UI.

1. Launch the NiFi UI.
2. From the Global Menu, select **Help** **Documentation** .

Related Information

[Apache NiFi Getting Started Guide](#)

Getting started with Apache NiFi Registry

See the Apache NiFi Registry documentation to access the *Apache NiFi Registry Getting Started Guide*.



Important: The Apache NiFi Registry documentation is available both online and in the NiFi Registry UI. The online documentation always shows the latest version of the NiFi Registry documentation. There may be differences in the content compared to the version of NiFi Registry you are using. You can access the documentation corresponding to the NiFi Registry version you are using directly from the NiFi Registry UI.

1. Launch the NiFi Registry UI.
2. From the Global Menu, select **Help** **Documentation** .

Related Information

[Apache NiFi Registry Getting Started Guide](#)