

Using DataFlow Catalog Registry Client

Date published: 2019-06-26

Date modified: 2023-08-29



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

Creating a Machine User.....	4
Adding DataFlow Catalog as Registry Client.....	6
Checking out a ReadyFlow.....	8
Checking out your flow.....	10
Versioning a flow in the Catalog.....	10

Creating a Machine User

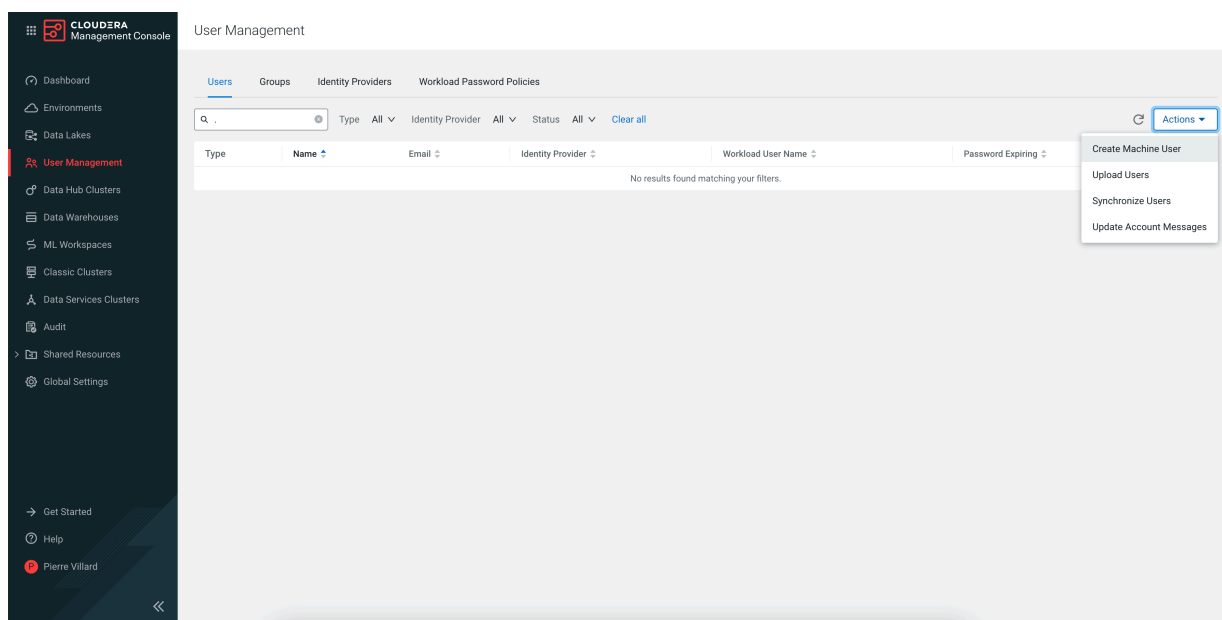
Learn how to create a Machine User with the proper roles for the DataFlow Catalog and generate credentials.

Before you begin

You are a Cloudera customer for CDP Public Cloud.

Procedure

1. Log in to the CDP Public Cloud web interface.
2. Navigate to **Management Console User Management** to create a Machine User that will be used by NiFi to interact with the DataFlow Catalog.
3. Select **Create Machine User** from the Actions menu on the Users tab.



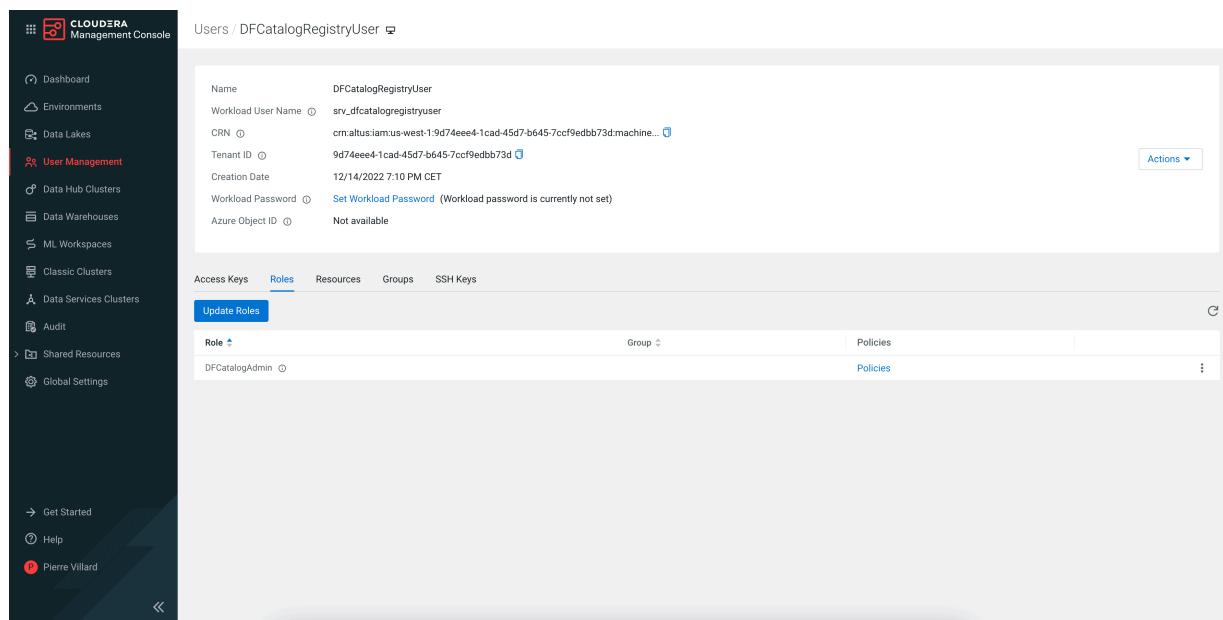
4. Give the user a meaningful name and give it the proper roles:

DFCatalogViewer

It gives read-only access to the DataFlow Catalog. This is enough if you just want to check out existing flows from the Catalog.

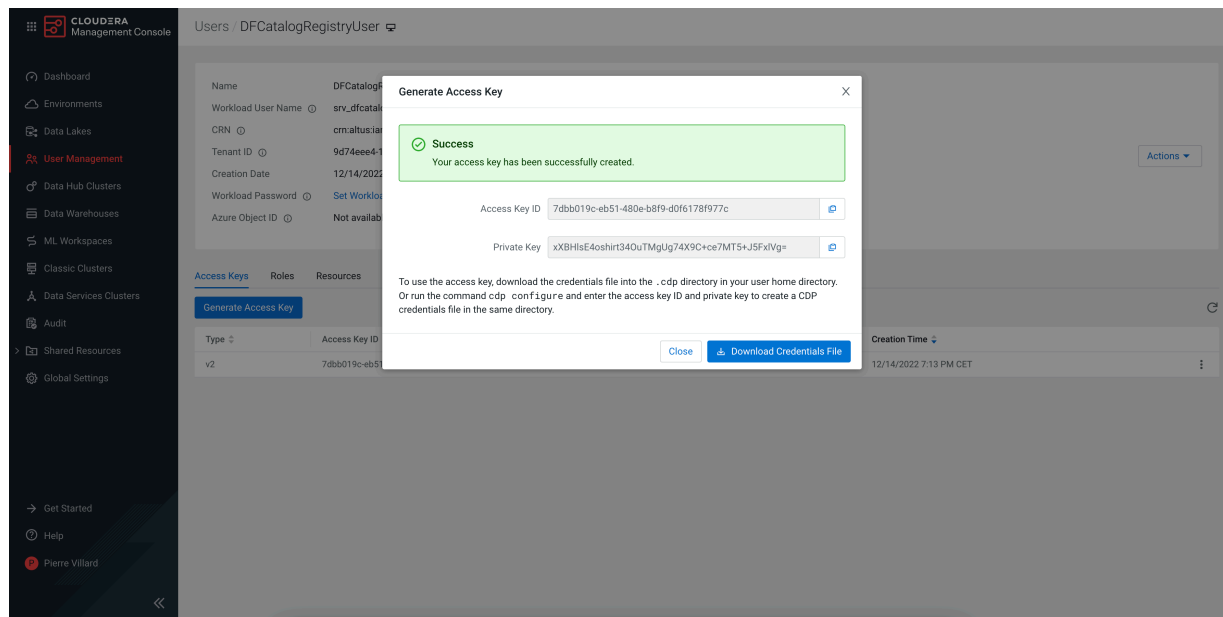
DFCatalogAdmin

It gives read-write access to the DataFlow Catalog. This would allow you to version and control flows from NiFi into the Catalog, and to check out existing flows from the Catalog.



The screenshot shows the Cloudera Management Console interface. On the left is a sidebar with navigation links: Dashboard, Environments, Data Lakes, User Management (highlighted), Data Hub Clusters, Data Warehouses, ML Workspaces, Classic Clusters, Data Services Clusters, Audit, Shared Resources, and Global Settings. At the bottom of the sidebar are links for Get Started, Help, and the user profile (Pierre Villard). The main content area is titled 'Users / DFCatalogRegistryUser'. It shows user details: Name (DFCatalogRegistryUser), Workload User Name (srv_dfccatalogregistryuser), CRN (cm:altusiam:us-west-1:9d74eee4-1cad-45d7-b645-7ccf9edbb73d:machine...), Tenant ID (9d74eee4-1cad-45d7-b645-7ccf9edbb73d), Creation Date (12/14/2022 7:10 PM CET), Workload Password (Set Workload Password), and Azure Object ID (Not available). Below this is a tabbed interface with 'Access Keys', 'Roles', 'Resources', 'Groups', and 'SSH Keys'. The 'Roles' tab is active, showing a table with one role: DFCatalogAdmin. There is a 'Policies' link next to the role name.

5. Go to the Access Keys tab and generate credentials for your user.



The screenshot shows the same Cloudera Management Console interface as before, but with the 'Access Keys' tab selected. A 'Generate Access Key' dialog box is open in the center. The dialog has a green success banner that says 'Success: Your access key has been successfully created.' Below this, it displays the 'Access Key ID' (7dbb019c-eb51-480e-b8f9-d0f6178f977c) and the 'Private Key' (xXBHlEt4oshirt34OuTMgUg74X9C+ce7MT5+J5FxlVg=). Below the keys, there is a note: 'To use the access key, download the credentials file into the .cdp directory in your user home directory. Or run the command cdp config and enter the access key ID and private key to create a CDP credentials file in the same directory.' At the bottom of the dialog are 'Close' and 'Download Credentials File' buttons. In the background, the 'Access Keys' table is visible, showing a new entry with Type 'v2' and Access Key ID '7dbb019c-eb51-480e-b8f9-d0f6178f977c'.



Note: Make sure to keep track of the Access Key ID and the Private Key as you will need these credentials to configure the Registry Client in NiFi.

What to do next

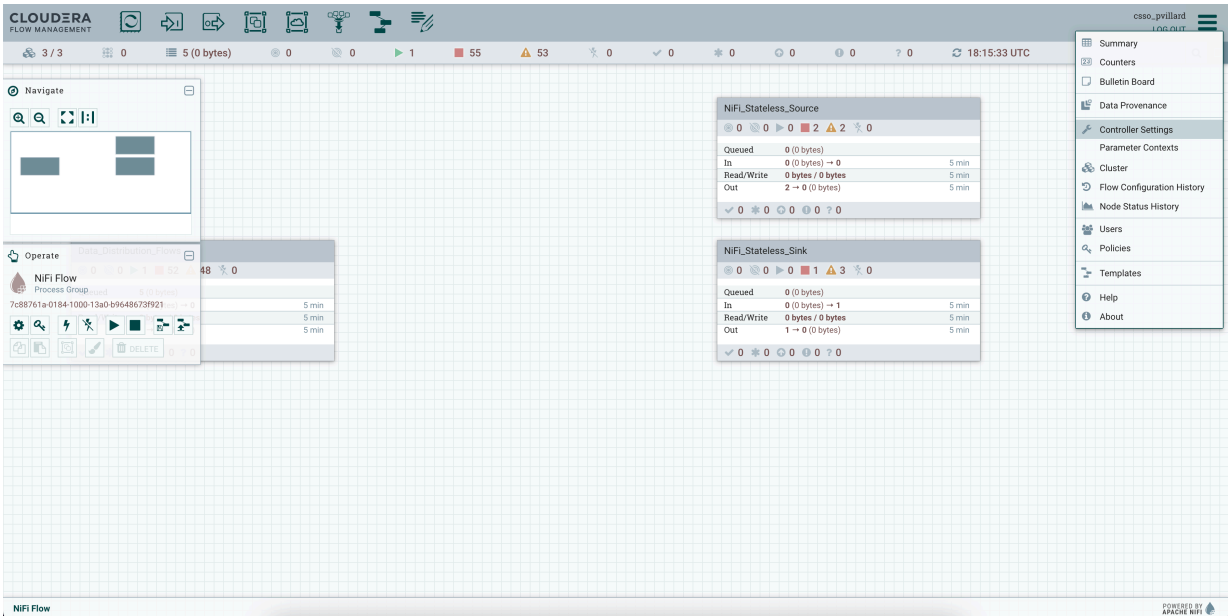
You can now create and configure the DF Catalog client with the available credentials.

Adding DataFlow Catalog as Registry Client

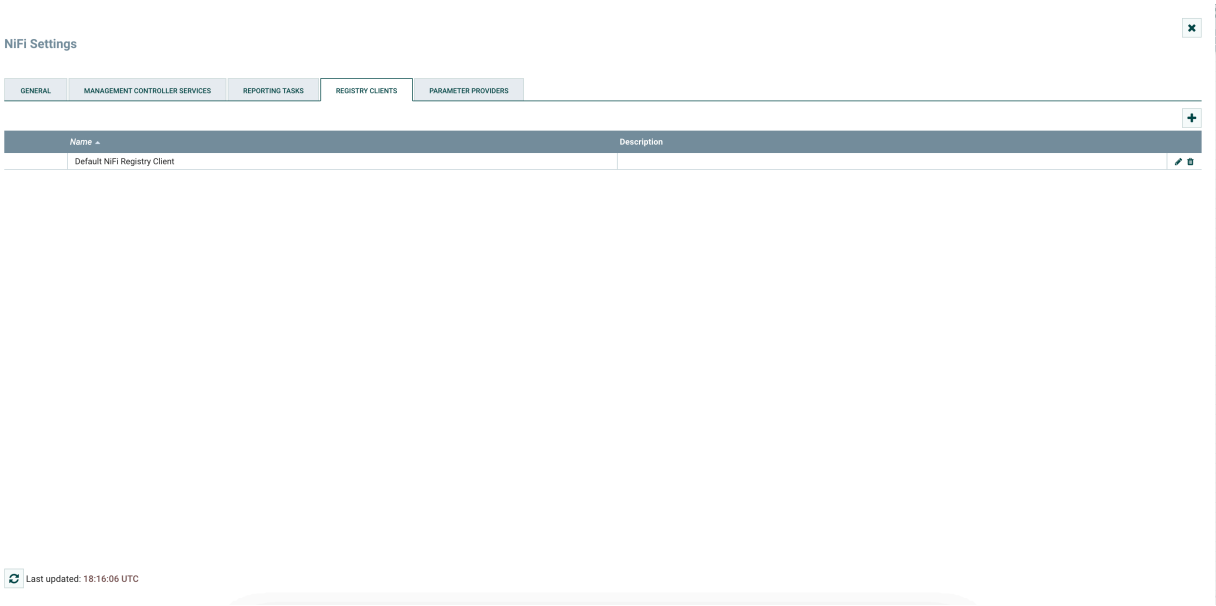
The DF Catalog is where readyflows and custom flow definitions are stored in Cloudera DataFlow for Public Cloud. Learn how to add and configure the DataFlow Catalog client as a new Registry Client.

Procedure

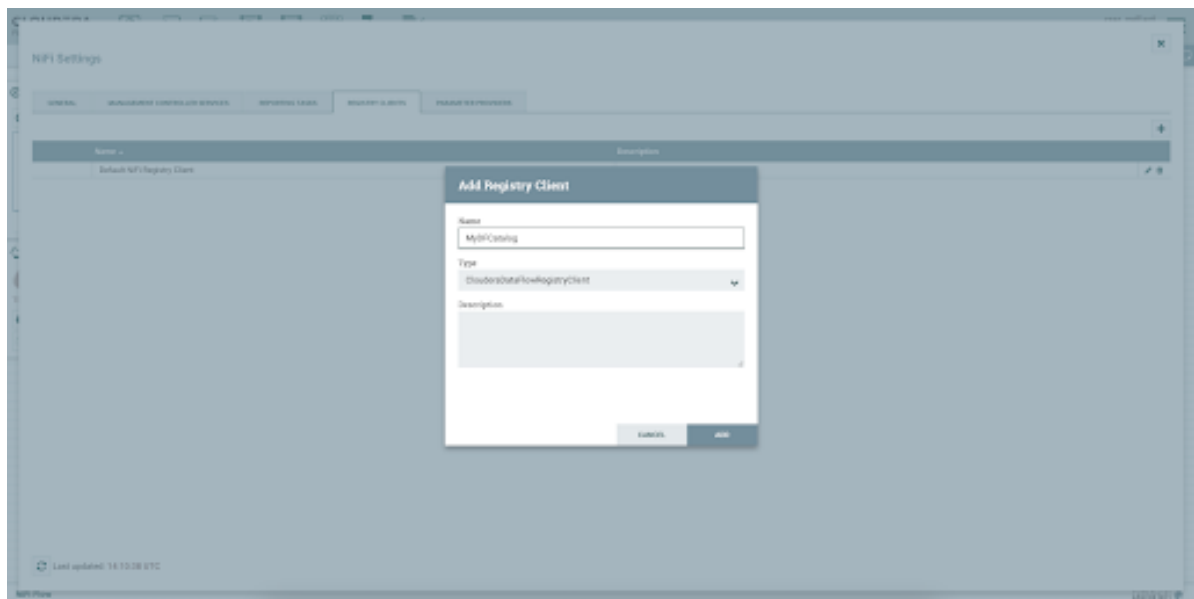
- 1. Open the NiFi UI.
- 2. Click the Global menu and select Controller Settings.



- 3. Switch to the Registry Clients tab.



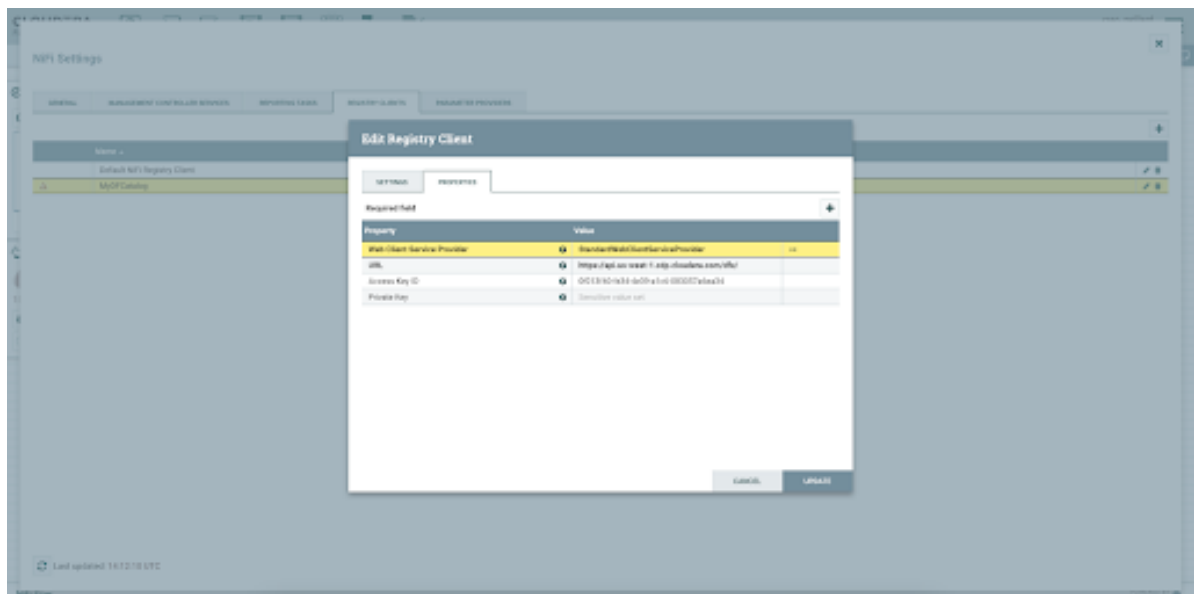
4. Add a new client.
 - a) Select the type and give it a name.



- b) Configure the client by providing the information related to your Machine User.



Note: You will have to create a StandardWebClientServiceProvider controller service and start it with its default configuration.



Results

When properly configured, you are ready to use the DataFlow Catalog to either version control your flows or check out flow definitions from the Catalog into NiFi.

What to do next

You can now start using the new client to version or check out flows from the DataFlow Catalog.

Checking out a ReadyFlow

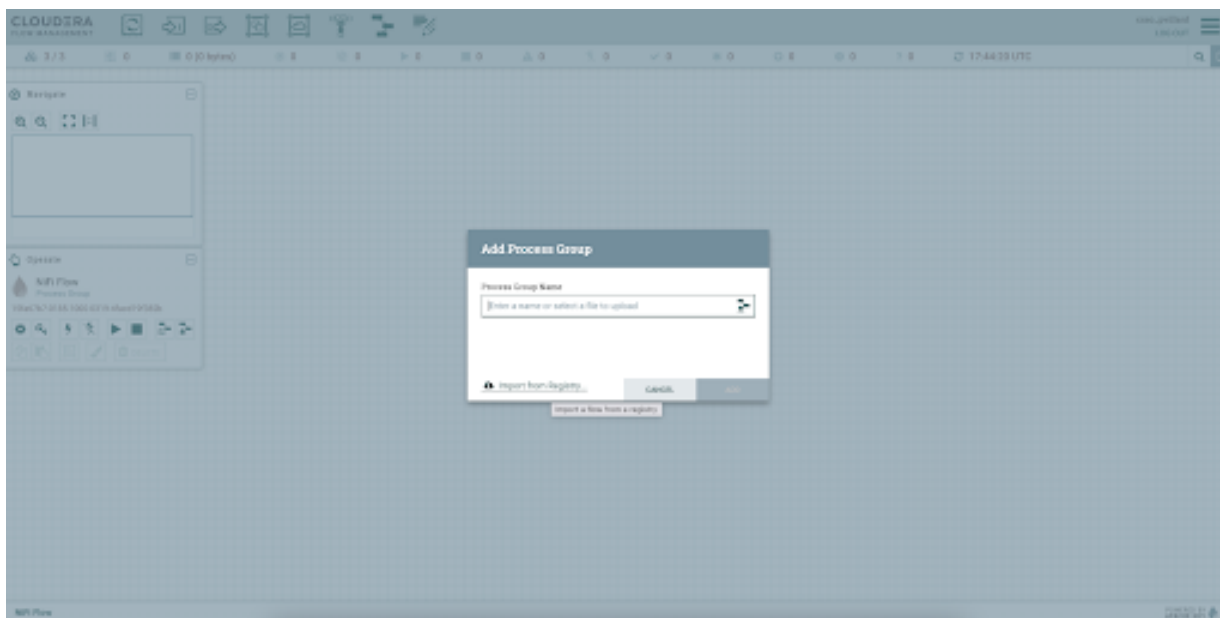
Learn how to check out a ReadyFlow from the DataFlow Catalog.

About this task

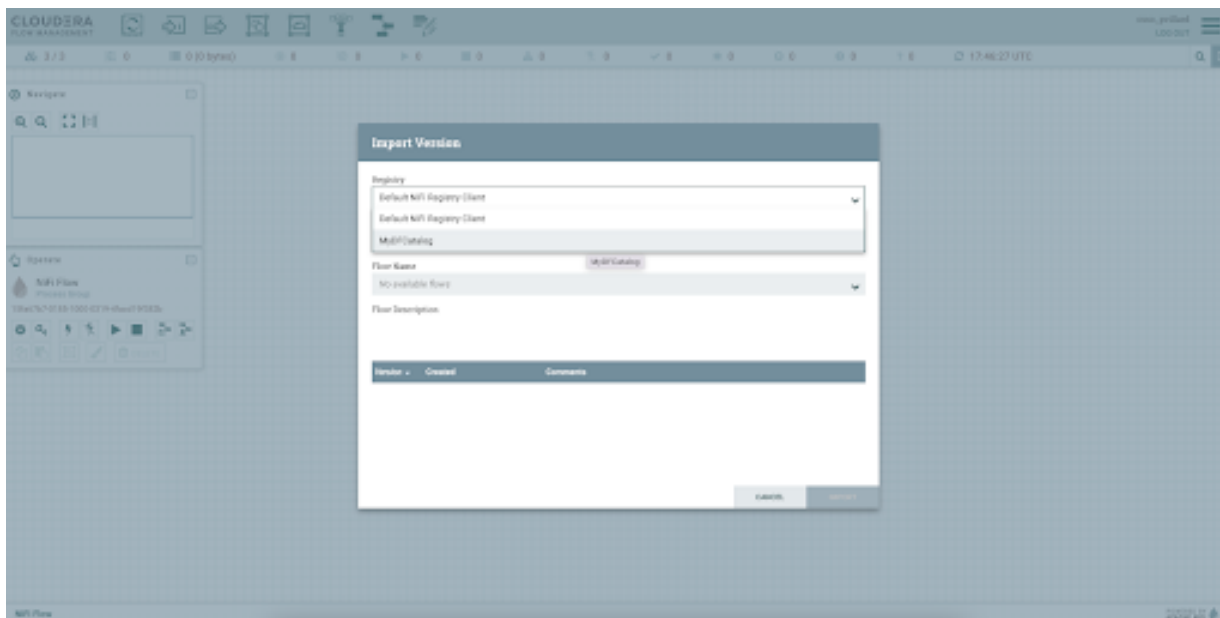
ReadyFlows are predefined, out-of-the-box data flows, which can be deployed by providing a small set of required parameters. They are read-only by default so you would not be able to commit any change made to them. With the Registry Client in NiFi connecting to Cloudera DataFlow Catalog, you can check out a ReadyFlow from the Catalog and edit it as a custom flow.

Procedure

1. Drag and drop a process group onto the canvas and click Import from Registry.

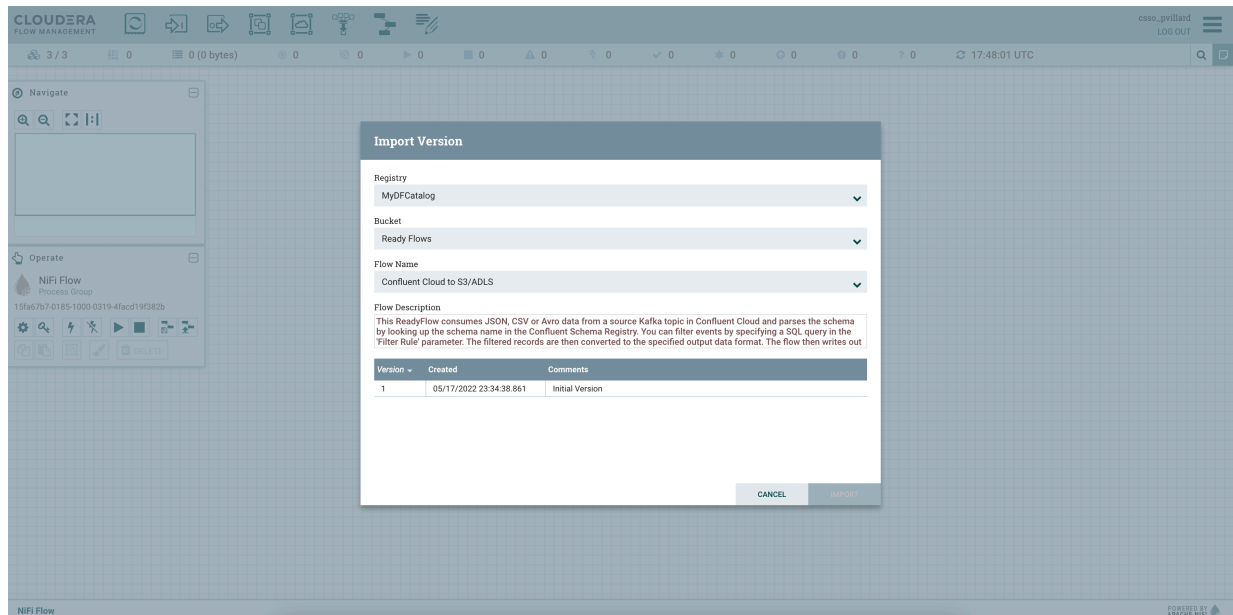


2. In the drop-down list, select the Registry Client configured to access your DataFlow Catalog.

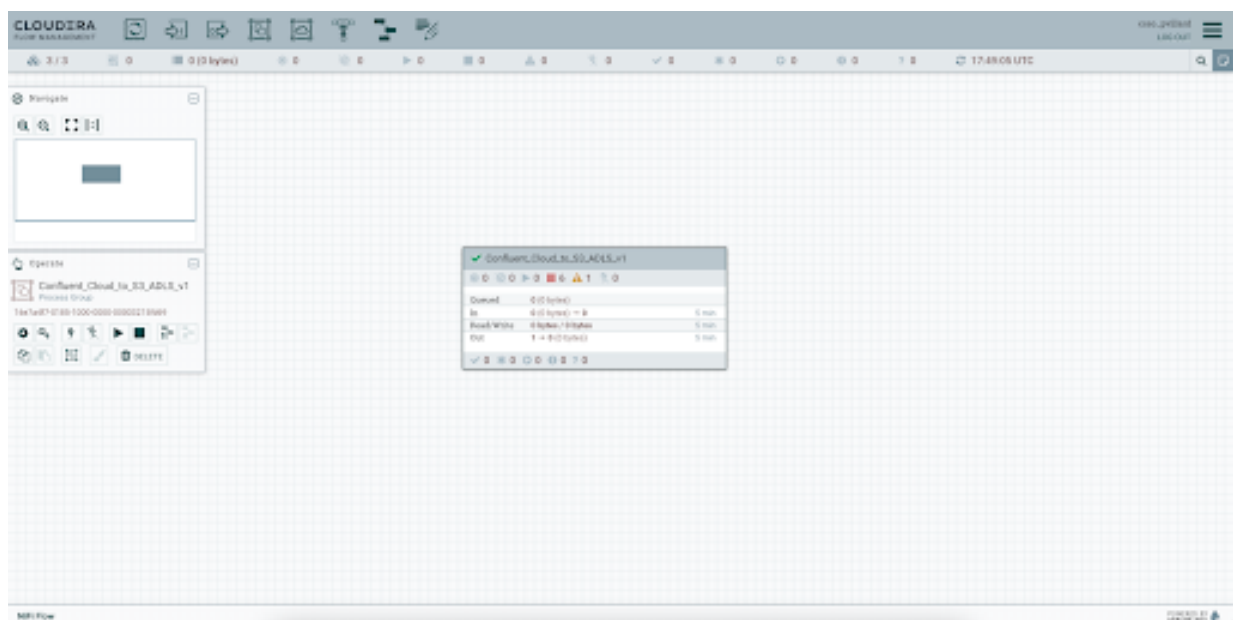


3. Select the ReadyFlows bucket.

You have two buckets - one for the ReadyFlows (which is read-only) and one for your own flows that are stored in the Catalog.



4. Check out the ReadyFlow by clicking the selected version and then clicking Import.



5. ReadyFlows are read-only so you are not able to change the original flow that you have checked-out.

If you want to edit the flow, you can create your own version:

- Right-click the process group, select Version, and click Stop version control.
- Make the changes you want.
- Version the flow as a custom flow in your Catalog: right-click the process group, select Version, and click Start version control.

Checking out your flow

Learn how to check out your own flows from the DataFlow Catalog.

About this task

If you want check out your own flow from the Catalog, you can follow the same approach as when checking out a ReadyFlow. You just need to select the Custom flows bucket. You can then make changes to your flow and commit the changes to create a new version of the flow in the Catalog.

Procedure

1. Drag and drop a process group onto the canvas and click Import from Registry.
2. In the drop-down list, select the Registry Client configured to access your DataFlow Catalog.
3. Select the Custom flows bucket.
4. Edit your flow and commit the changes to create a new flow version in the Catalog.

Versioning a flow in the Catalog

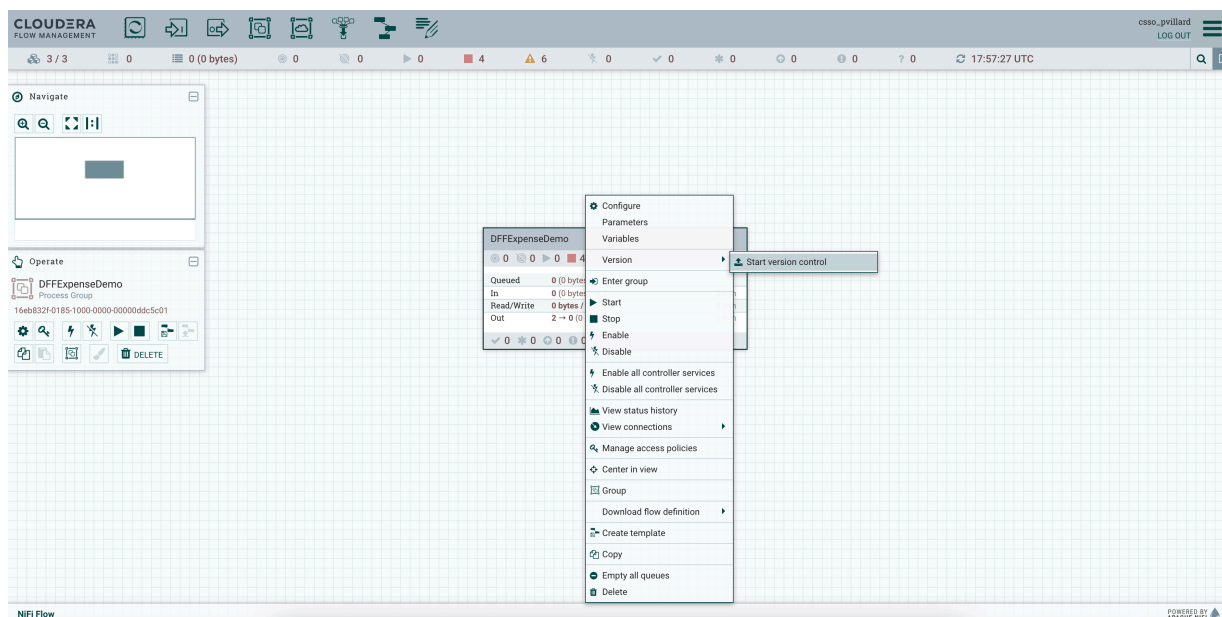
Learn how to use the DataFlow Catalog to version your flows.

About this task

Follow the below steps if you have a process group that you want to version as a completely new flow in the Catalog.

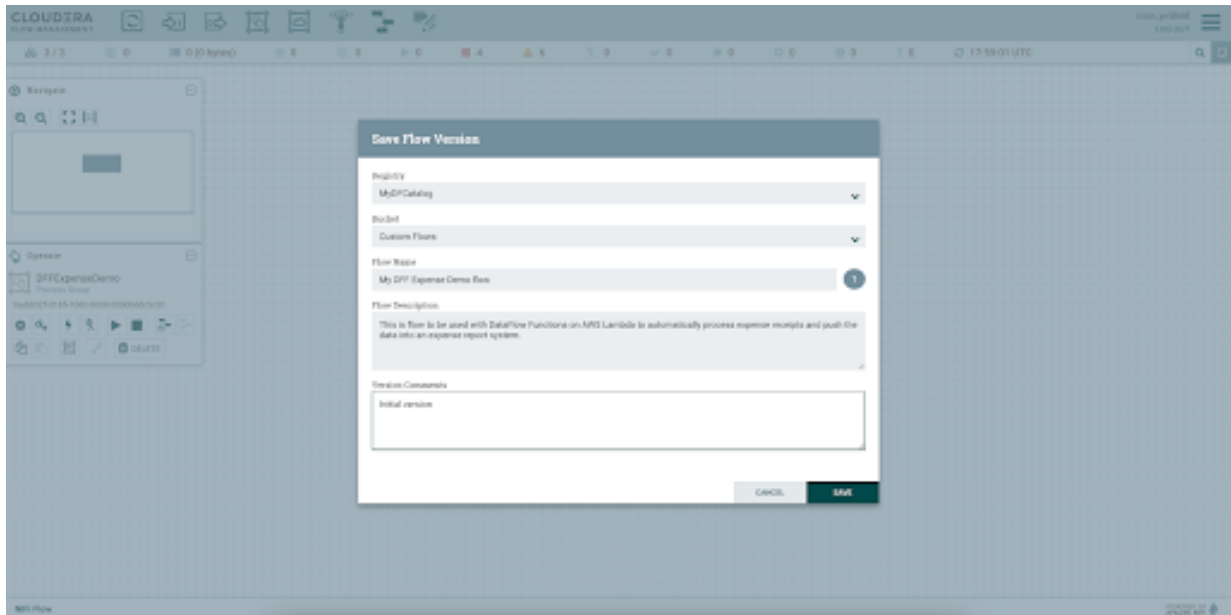
Procedure

1. Right-click the process group and select **Version Start version control**.



2. Select the Registry that you want to use.

3. Select Custom flows for the bucket and give your flow a name, a description, and a comment for the first version.



4. Click Save.

Results

The flow is now available in the DataFlow Catalog.



Note: All commits will be associated with the Machine User you configured when creating the Registry client.