Cloudera Flow Management 2.1.5

Using DataFlow Catalog Registry Client

Date published: 2019-06-26
Date modified: 2023-11-02

CLOUDERA

Legal Notice

Contents

Creating a Machine User

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

About this task

You can connect the Registry Client to the Cloudera DataFlow Catalog, expanding beyond NiFi Registry. This allows you to version and manage flow definitions in a unified location for both on-premises and cloud environments.

Hybrid customers can streamline flow versioning across deployments, while on-premises users with a public cloud tenant gain access to the Cloudera DataFlow ReadyFlow Gallery.

Before you begin

Besides using Cloudera Flow Management on premises, you are a Cloudera customer for Cloudera 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.
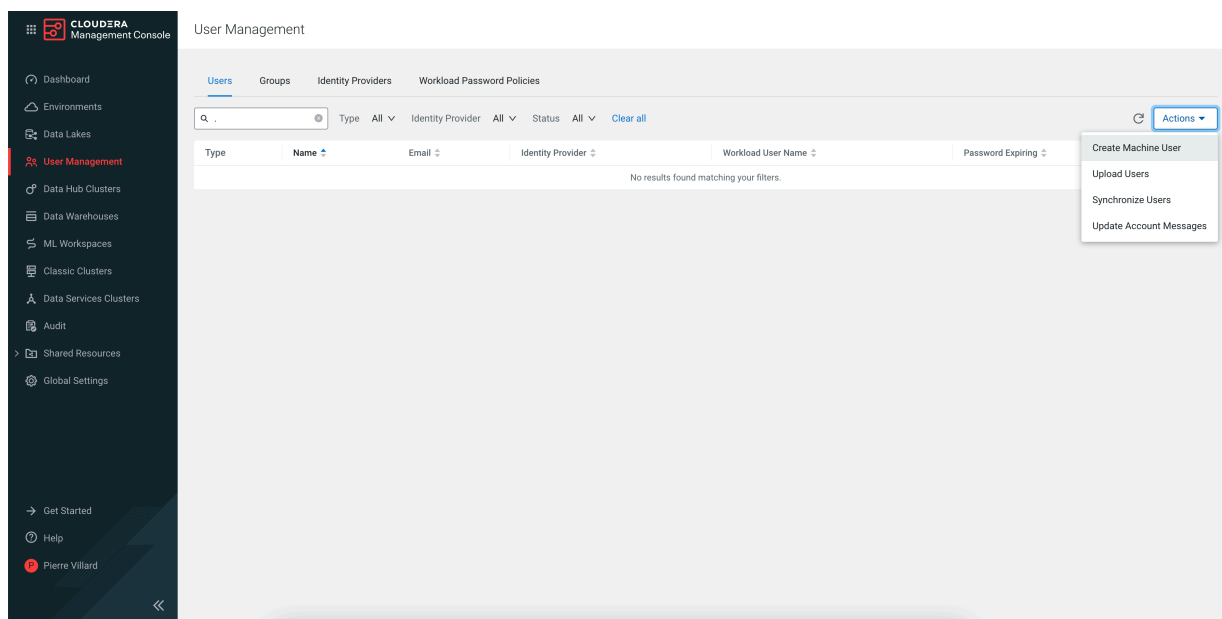
4. Give the user a meaningful name and give it the proper roles:

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.

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

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 a new Registry Client

Learn how to add and configure the DF Catalog client as a new Registry Client.

Procedure

1. Open the NiFi UI.
2. Click the Global menu and select Controller Settings.
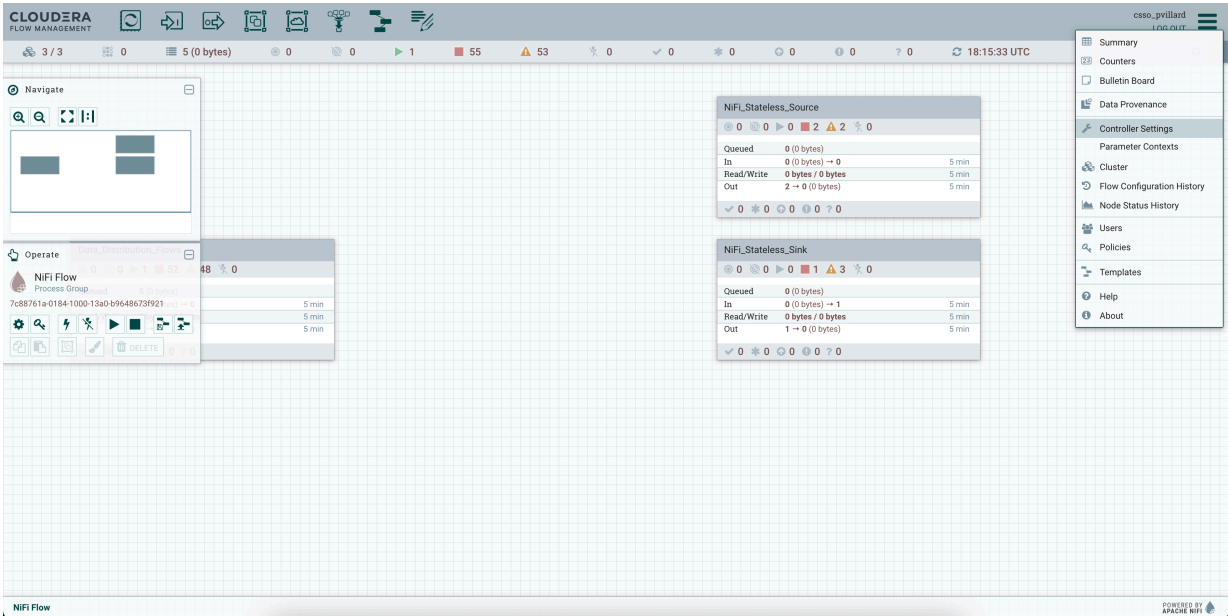
3. Switch to the Registry Clients tab.
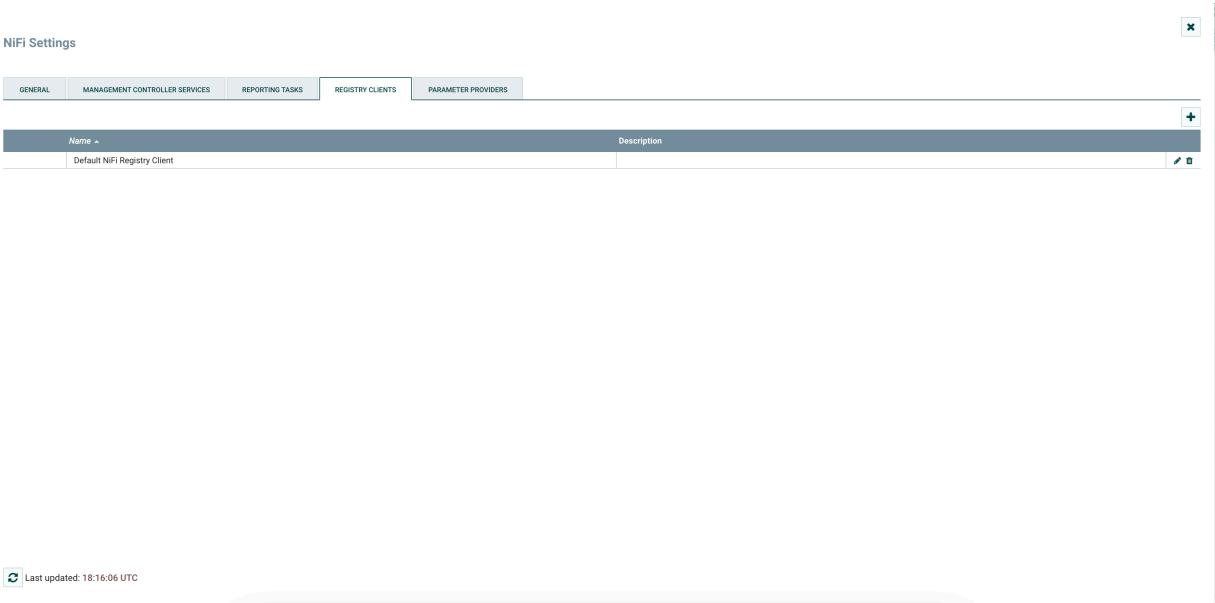
4. Add a new client.

    a) Select the type and give it a name.
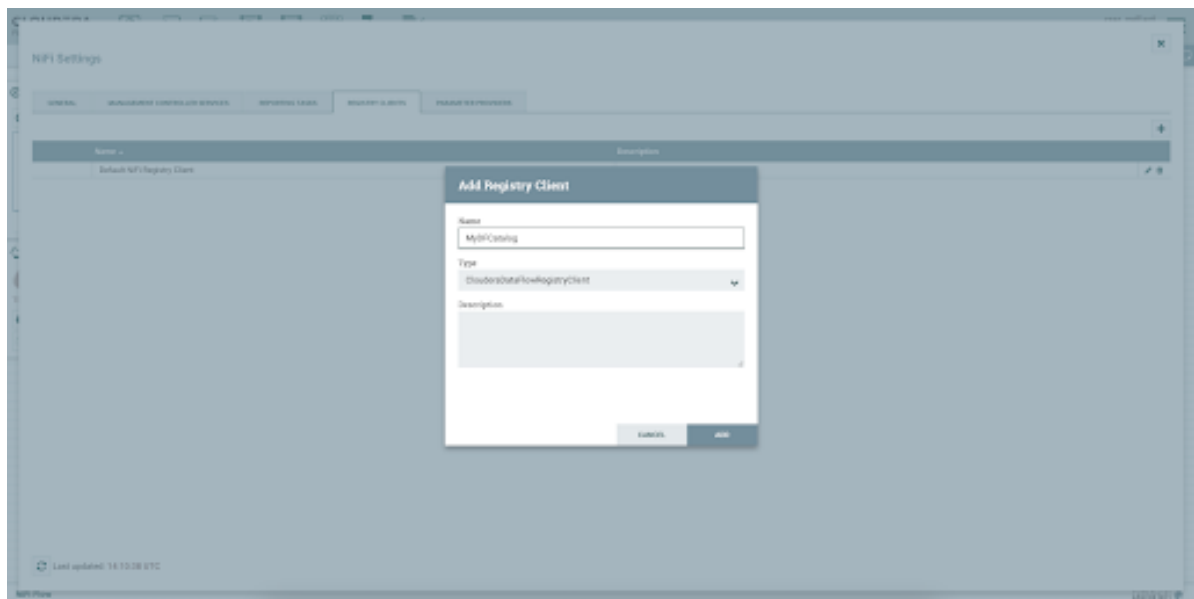
    b) Configure the client by providing the information related to your Machine User.

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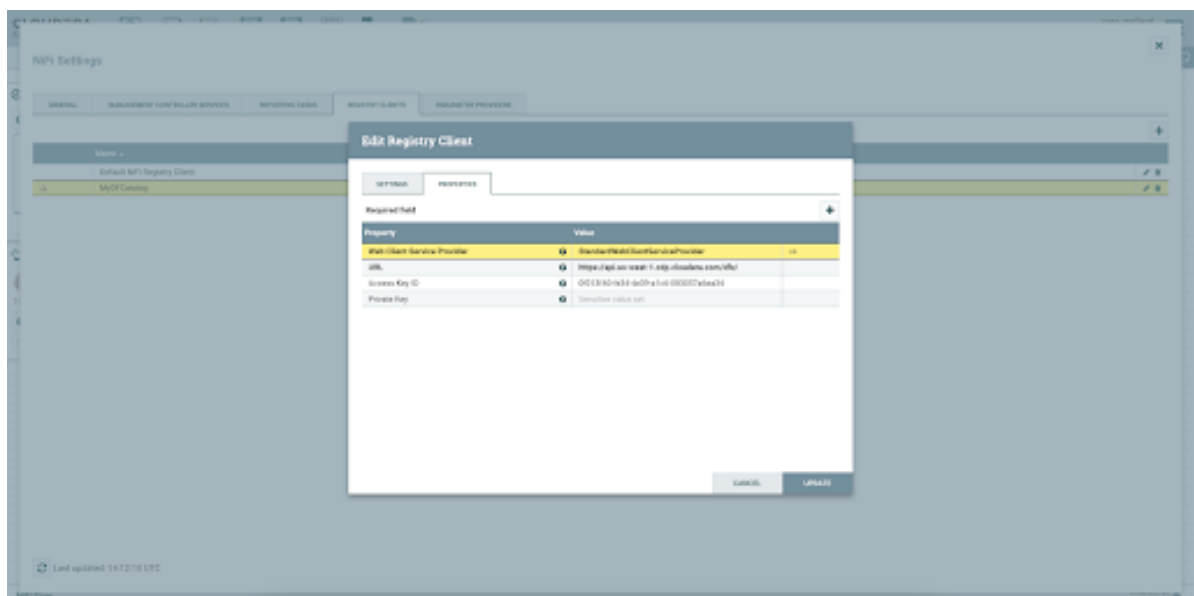
Results

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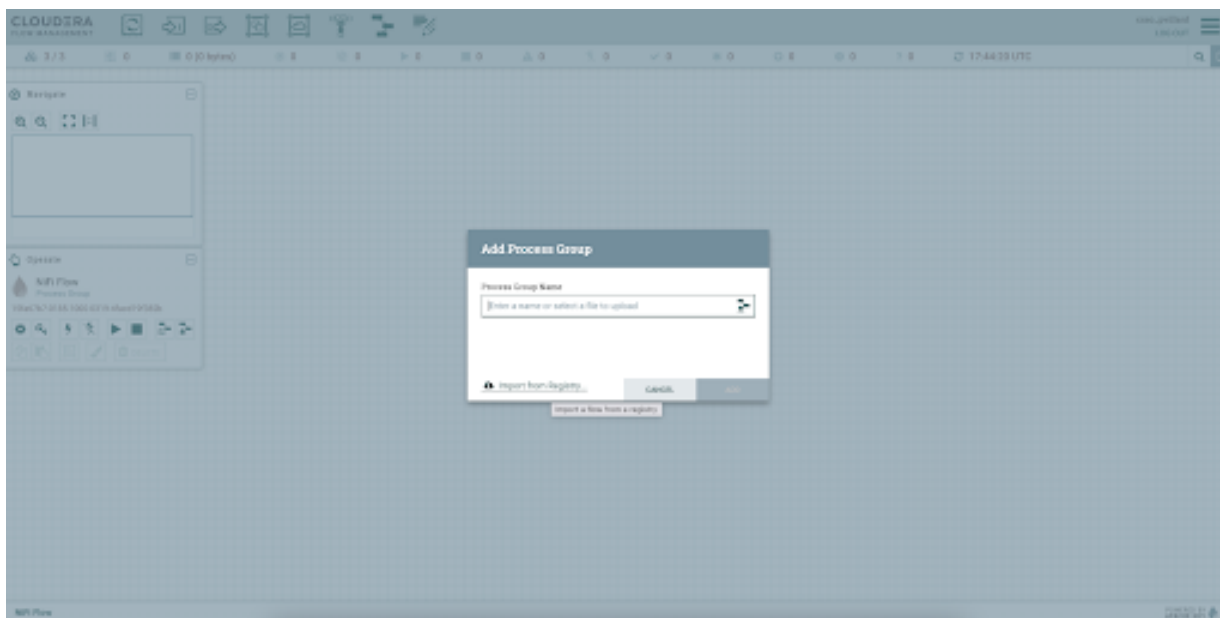
2. In the drop-down list, select the Registry Client configured to access your DataFlow Catalog.
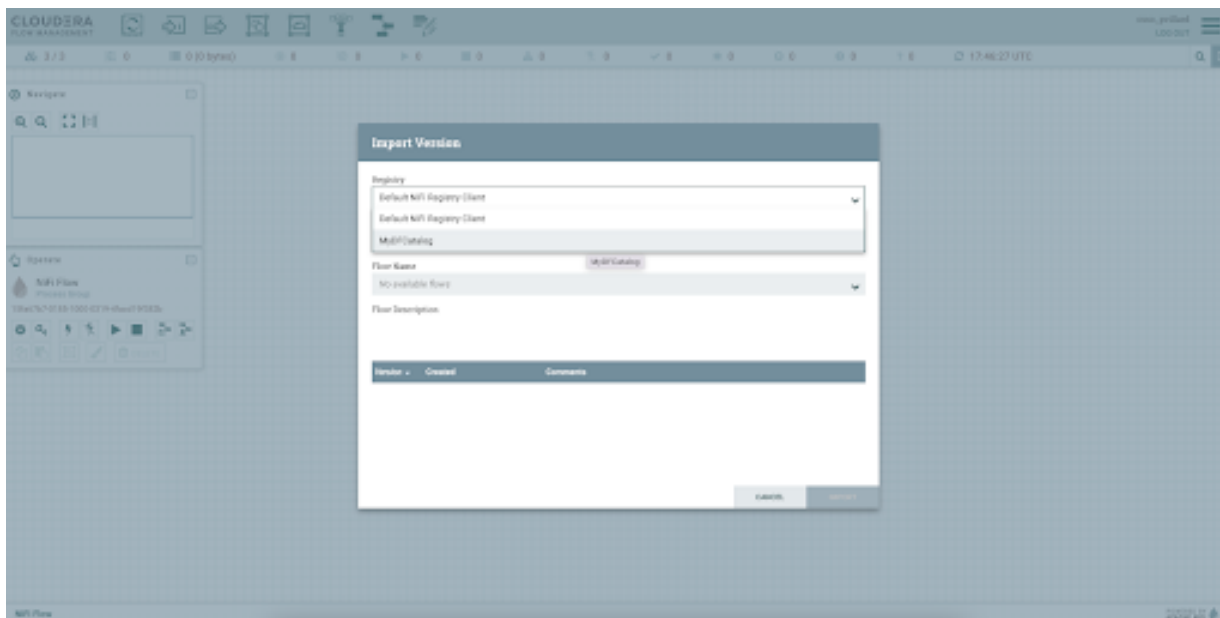
3. Select the ReadyFlows bucket.

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. Select the ReadyFlow that you want to check out in NiFi by clicking the selected version and then clicking Import.

5. ReadyFlows are read-only so you would not be able to change flow. If you want to edit the flow:

   a) Right-click the process group, select Version and click Stop version control.

   b) Make the changes you want.

   c) 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 flows

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

About this task

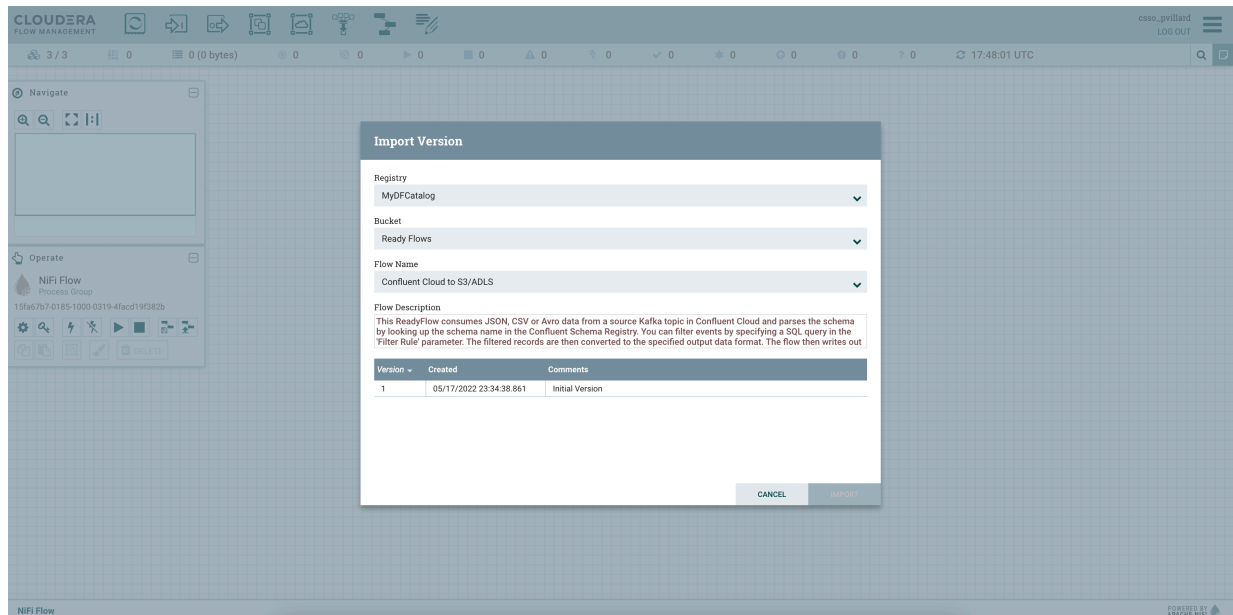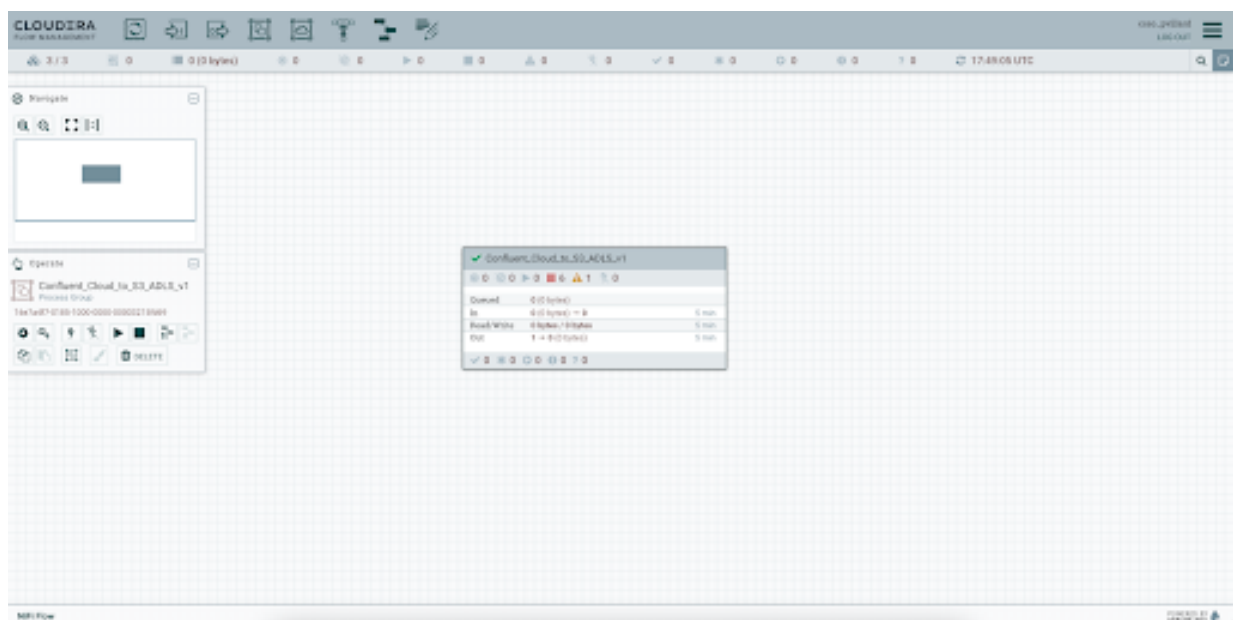If you want check out your own flow from the Catalog, you can follow the same approach as when checking out a ReadyFlow. You just need to select the Custom flows bucket. You can then make changes to your flow and commit the changes to create a new version of the flow in the Catalog.

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.
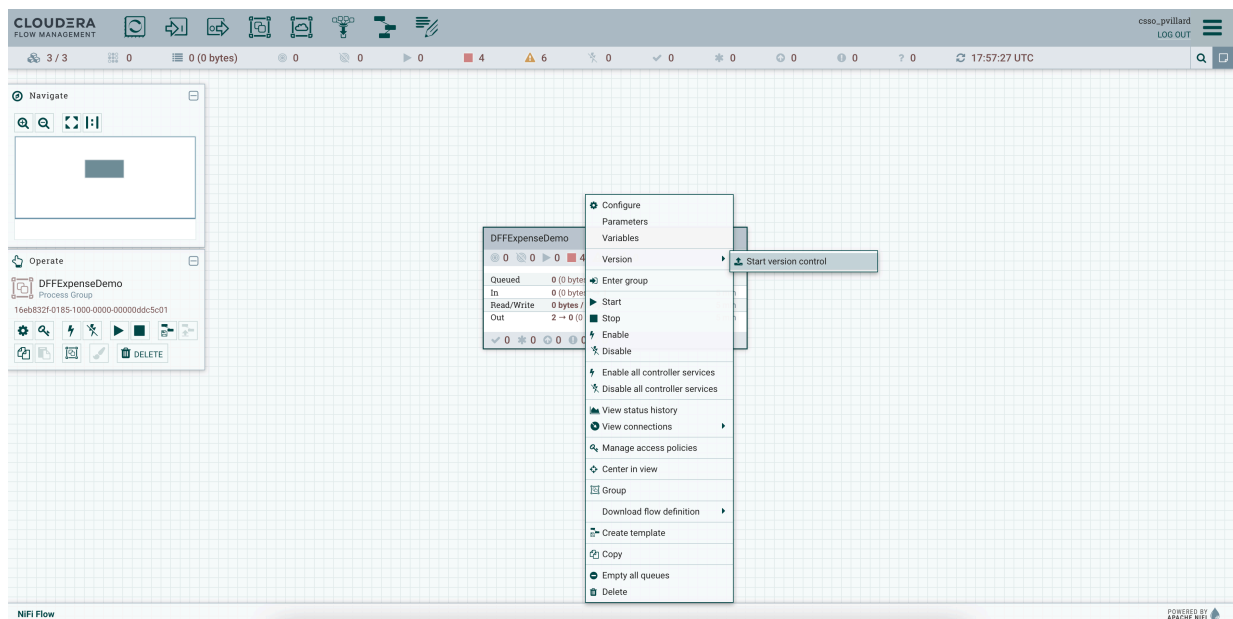
Versioning a flow in the Catalog

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

About this task

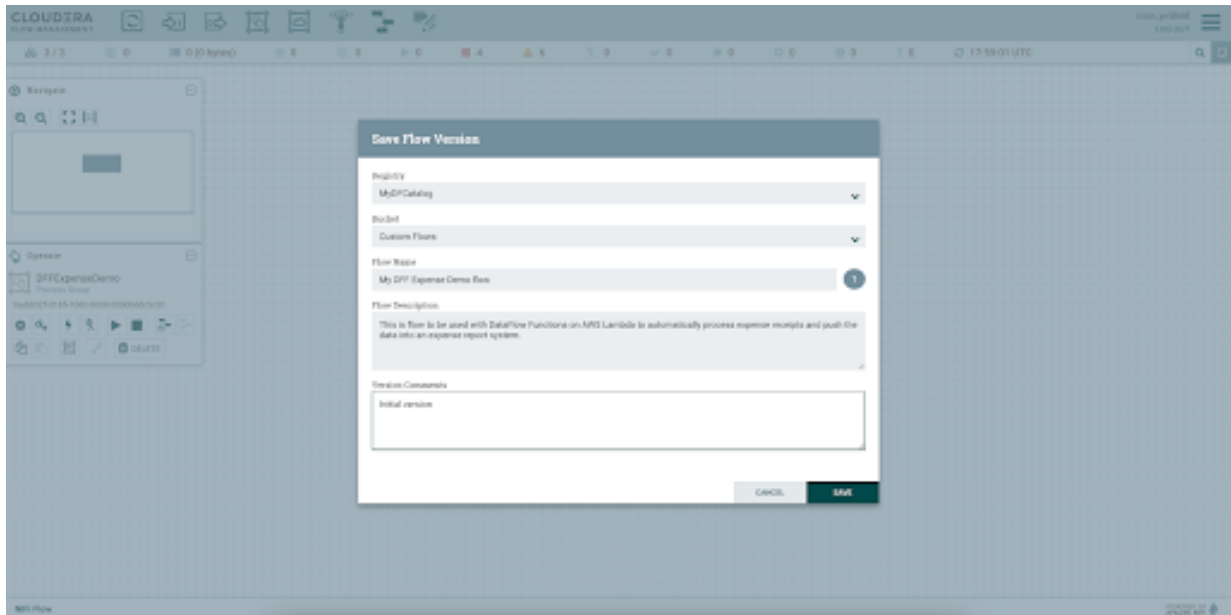Follow the below steps if you have a process group that you want to version as a completely new flow in the Catalog.

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.