Cloudera Flow Management 4.0.0

Using Cloudera DataFlow Catalog Registry Client

Date published: 2019-06-26 Date modified: 2024-12-11



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

Creating a Machine User	4
Adding a new Registry Client	6
Checking out a ReadyFlow	8
Checking out your flows	9
Versioning a flow in the Catalog	10

Creating a Machine User

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

Before you begin

You are a Cloudera customer for Cloudera Public Cloud.

Procedure

- **1.** Log in to the Cloudera 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 Cloudera DataFlow Catalog.
- 3. Select Create Machine User from the Actions menu on the Users tab.

ELOUDERA Management Console	User Management					
🕜 Dashboard	Users Groups Identity Provide	rs Workload Password P	olicies			
C Environments	Q. O Type All	V Identity Provider All N	✓ Status All ✓ Clear all			C Actions -
🚉 Data Lakes						Create Machine User
옷 User Management	Type Name 🕈	Email ≑	Identity Provider \$ No results found ma	Workload User Name 🗘	Password Expiring ≑	Upload Users
🕈 Data Hub Clusters			No results found ma	toning your niters.		Synchronize Users
🖨 Data Warehouses						Update Account Messages
S ML Workspaces						
昰 Classic Clusters						
A Data Services Clusters						
🕄 Audit						
> 🔄 Shared Resources						
Ø Global Settings						
→ Get Started						
⑦ Help						
Pierre Villard						
*						

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

DFCatalogViewer

It gives read-only access to the Cloudera 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 Cloudera 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.

Hanagement Console	Users / DFCatalogRe	gistryUser 🖵			
 Dashboard 	Name	DFCatalogRegistryUser			
C Environments	Workload User Name 🕥	srv_dfcatalogregistryuser			
😪 Data Lakes	CRN ()	crn:altus:iam:us-west-1:9d74eee4-1ca	d-45d7-b645-7ccf9edbb73d:machine 🗊		
🙊 User Management	Tenant ID 🕥	9d74eee4-1cad-45d7-b645-7ccf9edbb	73d 🕽		Actions -
o ^o Data Hub Clusters	Creation Date	12/14/2022 7:10 PM CET			
	Workload Password ①	Set Workload Password (Workload pa	assword is currently not set)		
🖬 Data Warehouses	Azure Object ID 🔘	Not available			
S ML Workspaces					
문 Classic Clusters	Access Keys Roles F	esources Groups SSH Keys			
A Data Services Clusters	Update Roles				G
😰 Audit					Ŭ
> 🔄 Shared Resources	Role ≑		Group ‡	Policies	
Global Settings	DFCatalogAdmin ①			Policies	:
Çş-					
\rightarrow Get Started					
⑦ Help					
Pierre Villard					
«					

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

CLOUDERA Management Console	Users / DFCatalogR	egistryUser	Ŧ				
 Dashboard Environments 	Name Workload User Name ()		Generate Access Key		×		
운 Data Lakes 옷 User Management ♂ Data Hub Clusters	CRN Tenant ID Creation Date	crn:altus:iar 9d74eee4-1 12/14/2022	Success Your access key has been	successfully created.			Actions 🔻
Data Warehouses ML Workspaces	Workload Password ③ Azure Object ID ①	Set Workloz Not availab	Access Key ID	7dbb019c-eb51-480e-b8f9-d0f6178f977c	e		
드 ML Workspaces 曼 Classic Clusters A Data Services Clusters	Access Keys Roles	Resources		xXBHIsE4oshirt340uTMgUg74X9C+ce7MT5+J5FxIVg= he credentials file into the .cdp directory in your user home di jure and enter the access key ID and private key to create a CI			
😰 Audit	Generate Access Key Type \$	Access Key ID	credentials file in the same directo			Creation Time 💠	G
> 🔄 Shared Resources	v2	7dbb019c-eb51		Close 🛓 Download Creden	tials File	12/14/2022 7:13 PM CET	:
→ Get Started ⑦ Help							
Pierre Villard							



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 Cloudera DataFlow Catalog client with the available credentials.

Adding a new Registry Client

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

Procedure

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

FLOW MANAGEMENT	🛃 🊏 🛅	<i>≡</i> //			csso_pvillard
	◎ 0	1 📕 55 🔺 53	8 % 0 ✔ 0	* 0 ⊙ 0 ● 0 ? 0 😂 18:15:33 UTC	Summary Counters Bulletin Board
Navigate				NiFi_Stateless_Source	Le Data Provenance
Q Q []!				◎ 0 ◎ 0 ▶ 0 ■ 2 ▲ 2 ※ 0	Controller Settings
				Ouesaid 0 (0 bytes) In 0 (0 bytes) → 0 Smin Smin Read/Write 0 bytes / 0 bytes Out 2 → 0 (0 bytes) Smin ✓ 0 ★ 0 ◇ 0 ◎ 0 ? 0	Parameter Contexts
Coperate				NiFi_Stateless_Sink	Q Policies
Operate Data_Distribution_Flows Data_Distribution_Flows Data_Distribution_Flows Distribution_Flows Distributi Distribution_Flows					Templates
	5 min 5 min 5 min			Queued 0 (0 byte) In 0 (0 byte) → 1 5 min Read/Wite 0 bytes / 0 bytes 5 min Out 0 (0 byte) 5 min Vit 0 (0 byte) 5 min Vit 0 (0 byte) 5 min ✓ 0 * 0 © 0 © 0 ? 0 0 ? 0	 Help About
Nifi Flow					POWERED

3. Switch to the Registry Clients tab.

C Last updated: 18:16:06 UTC

NIC Com								×
NiFi Setti	ngs							
GENERAL	MANAGEMENT CONTROLLER SERVICES	REPORTING TASKS	REGISTRY CLIENTS	PARAMETER PROVIDERS				
								+
	Name 🔺				Description			
	Default NiFi Registry Client							/ 0

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

C **	10000 CT CT CT CT CT CT CT		representation and	
а.			×	
				8
8	UNERS, DEPENDENT CONTRACTOR DEPUTES, BETWEEN LEDER, BELLEVILLE, P.	NAME OF PROPERTY.		
4				
	Industr NP1 Reports Chart	Add Registry Client	24	
		nus nepsey canas		
		Name		
		MydPCensing		
ć –		Tipe		
		CloudersbataflowfegatryClast		
-		Description		
		1.000). AND		
	C Last updated 18.10.28 UTC			
-	ten		0000	l

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.

C.	ALIANAN 720 CO CO POR POR OTIO					. total and the second se	EA3
8 -		anium terris, derits					
	Barry (Edit Registry Client				•	
	Entrady KeT's Registry Claims	MITTAGE REPORTED	7			21	
		Required field			+		
		Property		Vilue			
è		Web Olient Service Provider	9	Bander/Balcher/service/sortier	-		
1		28.		https://spi.ac.usat/fi.atjp.cloudes.com/sh/			
4		Jacoms Key (D		OCTIVE-shi dell'a he ilizitzatea hi			
		Private Ray	0	Densitive ratio set			
· Contraction of the local division of				6466	UPIAN		
	C Last spinod 167218 VTC						

Results

When properly configured, you are ready to use the Cloudera 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 Cloudera DataFlow Catalog.

Checking out a ReadyFlow

Learn how to check out a ReadyFlow from the Cloudera 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.

CLOUDERA 🖸 🗟 🗟 🖄 👘		= helling, see
& 3/3 🗄 0 🗰 010 hytes) 🗇 1 🖄 0 > 0		0 I 0 0 7 I 0 17.4433 UTC Q
@ Rongers B		
e.e. () H		
	Add Process Graup Press Grap Kant [Inter a save or which a first to updat] [Inter a save or which a first to updat]	
MR Firm		DEMLE &

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

CLOUDERA D A B D A					Los aut
66 37.3 E 0 ■ 0(03ymm) = 1 = Ø Servicent D		A1 2.4 VI			Q 🖬
Q Q ()	Import Version			-	
O Reveree D Morrow This Conference of the Confer	Ingulary Default Wit Regrays Dant Default Wit Regrays Dant Multi-Station Flow Ranz No available flows Flow Rescription	Water		* *	
	Newlor - Gradial	Garenariis	1.010		
	-			_	
NPI Flew					 51557137 &

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.

	1 - W	csso_pvillard
🗞 3 / 3 🕮 0 🗮 0 (0 bytes) 🛞 0 🛞	▶ 0 ■ 0 ▲ 0 % 0 ✓ 0 * 0 ⊙ 0 0 ? 0 C 17/48:01 UTC	Q D
Ø Navigate		
@ Q 【] :	Import Version	
	Registry MyGFCatalog	
	Bucket Ready Flows	
Operate IniFi Flow Process Group	Flow Name Confluent Cloud to S3/ADLS	
15/a67/b7.0185-1000-0319-4facd19/382b	Flow Description This ReadyFlow consumes JSON, GSV or Avro data from a source Kafka topic in Confluent Cloud and parses the schema by looking up the schema name in the Confluent Schema Registry. You can filter events by specifying a SOL query in the "Filter Rule" parameter. The filtered records are then converted to the specified output data format. The flow then writes out	
	Version - Created Comments	
	1 05/17/2022 2334 38.861 Initial Version	
	CANCEL	
NiFi Flow		POWERED BY

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

LOUDIRA 🖸 🖓 🖾 🖾 🖉 🌹 🦆 ಶ		tation and
& 3/3 🗄 0 🖩 0(3 kyles) 8 0 🛞 0 ⊨ 0	■ a ▲ a 3, a ∨ a ≈ a 0 a 0 a 2 a 2 124808UTC	Q.
Ranigati 😑		
u @. 12 -		
Eperan 🖂	✓ Confinem.Bood.m.S0.A015.v1	
	80 00 P0 86 A1 10	
Carflaeti (Goul Ja 33, ADU, v1 Protei Dop	Devel 001ete0	
Carlit? 0180-1000-0000-00002218/WH	in útíl hpre) = k Grain Read-Mithe útapes/ hitsten Sino	
N 1 2 1 1 1 1 1 1 1	Post With Tiples/Tiples S to: DX 1 + 60(tots) 5 tots	
10 H / 0 mm	¥1 80 00 8170	

- 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 Cloudera 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 Cloudera 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

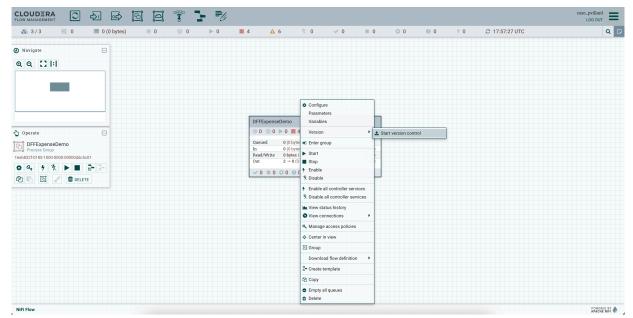
Learn how to use the DataFlow Catalog to version your flows.

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.

	n 12 18	ton.priled
💩 3/3 🗏 0 🔳 0 (0 kytm) 👘 8	E 8 → 6 🗰 4 🔺 8 3 6 √ 8 ≈ 6 0 6 0 6 18 2012.58010TC	Q. 🕑
() Kariyare ()		
a, a, () H	Save Flow Version	
-	Nejsy McDaley	
0 anno 8	Rolet Contro Nors	
Synthe Synthe Synthese S	Plan Rase Wy DY Superat Dense Ress Plan Description	
04.43.20mm 名下型/00mm	This is fism is for and with Batallian functions on APG Lambda is submatically process representements and push the data into an expense report system.	
	Terrilas Canaterila	
	hital cersion	
	Cercit. LM	
NOT Flow		anantar é

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.

III CLOUDERA DataFlow	Flow Catalog	»	C REFRESHED: 14 seconds ago	
 Dashboard 	Q DFF Expense Demo X	My DFF Expense Demo flow Updated 2 minutes ago by DFCatalogRegistryUser	Actions 💌	
III Catalog え ReadyFlow Gallery III Flow Design	Name 1 My DFF Expense Demo flow	FLOW DESCRIPTION This is flow to be used with DataFlow Functions on AWS Lambda to automatically process expense receipts and push the data into an expense report system. GRN #		
f() Functions		Crncdp.df.us-west-1:558bc1d2-8867-4357-8524-311d51259233:flow:My-DFF-Expense-Demo-flow		
		Only show deployed versions Version 1	Deployments	
		Deploy →	土 Download	
		LAST UPDATE 2022:12:15 18:59 CET by DFCatalogRegistryUser 'Initial version'		
		CRN # cmcdp.df.us-west-1:558bc1d2-8867-4357-8524-311d51259233:	flow:My-DFF-Expense-Demo-flow/v.1	
→ Get Started ⑦ Help				
Pierre Villard 2.3.8-b347 《				