

Quickstart

Date published: 2021-04-06

Date modified: 2025-09-30

CLOUDERA

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

Cloudera Data Flow Quickstart.....	4
Verify cloud infrastructure prerequisites.....	4
Give administrators access.....	4
Enable Cloudera Data Flow for an environment.....	4
Give users access.....	5
Add the Hello World ReadyFlow definition to the Catalog.....	6
Deploy the Hello World ReadyFlow using the deployment wizard.....	10

Cloudera Data Flow Quickstart

Get started with Cloudera Data Flow quickly by walking through a few simple steps. Ensure that administrators have access to Cloudera on cloud, enable Cloudera Data Flow for an environment, give users access to Cloudera Data Flow, and then add to the Catalog and deploy the Hello World ReadyFlow.

Verify cloud infrastructure prerequisites

As the administrator for your environment, ensure that the environment meets the requirements for Cloudera on cloud and Cloudera Data Flow.

[Deploy Cloudera using Terraform](#)

Give administrators access

To enable Cloudera Data Flow for an environment, users must have the DFAdmin role. Grant the DFAdmin role to a user or group that is planned to be allowed to enable CDF for an environment.

Before you begin

- You must have the Cloudera PowerUser role.

Procedure

1. On the Cloudera Management Console page, click Environments in the left-hand sidebar.
2. Use the search field to find and select the Cloudera on cloud environment for which you want to grant DFAdmin rights.
3. Go to **Actions Manage Access** to display the Environments Access page.
4. Find the user or group to whom you want to grant the DFAdmin role, and click the Update Roles button.
5. On the **Update Resource Roles for** page, select the DFAdmin checkbox from the available roles and click the Update Roles button.
6. Click Synchronize Users.

This propagates the changes you made to the selected environment.

What to do next

When you have finished granting a user or group administrator permissions, they can now proceed by enabling Cloudera Data Flow for an environment.

Related Information

[Cloudera Data Flow Security](#)

Enable Cloudera Data Flow for an environment

Before you can deploy flow definitions, you must enable Cloudera Data Flow for a Cloudera on cloud environment. Enabling Cloudera Data Flow for an environment means that you are preparing an active and healthy Cloudera on cloud environment for use with Cloudera Data Flow.

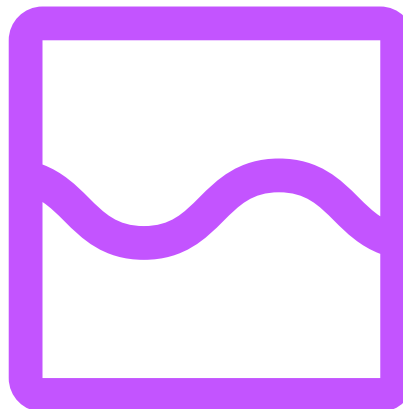
Before you begin

- You must have a cloud provider account and meet the infrastructure and network requirements.
- You must have a healthy Cloudera on cloud environment, with FreeIPA and the data lake running and healthy.

- You must have the DFAdmin role for the Cloudera on cloud environment for which you want to enable Cloudera Data Flow.

Procedure


1.



Navigate to Cloudera Data Flow, by selecting Cloudera on cloud Home Page, or from the navigation pane.

Data Flow from the

2.

Go to  Environments, and click Enable to launch the **Enable Cloudera Data Flow Service** pane for the environment you want to enable.

3. In the **Enable Cloudera Data Flow Service** form, provide the following information:

- Instance Type – From the available instance types, select the one appropriate for your use case. For testing and evaluation, the smallest c5.xlarge instance type with 4 vCPUs, 8 GB RAM is sufficient.
- Cloudera Data Flow Capacity – Specify a minimum and a maximum size for the Kubernetes cluster. You can keep the default settings.
- Networking

4.

Click the  Enable button. Enabling Cloudera Data Flow can take up to one hour.

What to do next

When you have finished enabling Cloudera Data Flow for an environment, proceed by giving users permission to import and deploy flow definitions.

Related Information

[Enabling Cloudera Data Flow for an environment](#)

[Managing Cloudera Data Flow in an environment](#)

Give users access

Cloudera Data Flow restricts who can import flow definitions and deploy them. To get started with Cloudera Data Flow, you must grant users permissions to perform these tasks so that they can import and deploy flow definitions in Cloudera Data Flow.

About this task

To get started with importing and deploying flow definitions, a user must have at least the following Cloudera Data Flow roles:



- DFCatalogPublisher – To import flow definitions to the Cloudera Data Flow Catalog

- DFFlowAdmin – To deploy flow definitions in Cloudera Data Flow

Before you begin

- You must have PowerUser role.
- You must know the user or group name to which you want to grant Cloudera Data Flow user access roles.

Procedure

1. Give a user permission to import flow definitions.
 - a) From Cloudera Management Console, click  User Management.
 - b) Enter the name of the user or group you want to authorize in the Search field.
 - c) Select the user or group from the list that displays.
 - d) Go to Actions Update Roles to display the **Update Roles** modal window.
 - e) Select the DFCatalogPublisher checkbox from the available roles and click the Update button.
2. Give a user or group permission to deploy flow definitions.
 - a) From Cloudera Management Console, click  Environments to display the **Environment / List** page.
 - b) Select the environment where you want the user or group to deploy flow definitions.
 - c) Go to Actions Manage Access to display the **Environments Access** page.
 - d) Find the user or group o whom you want to grant the DFFlowAdmin role and click the Update Roles button.
 - e) On the **Update Resource Roles for** modal window, select the DFFlowAdmin checkbox from the available roles.
 - f) Click the Update Roles button.

What to do next

When you have finished giving user or groups permission to import and deploy flow definitions, proceed by importing a flow definition.

Related Information

[Cloudera Data Flow Security](#)

Add the Hello World ReadyFlow definition to the Catalog

The Hello World ReadyFlow is a flow definition designed to work out of box without any further configuration, to make getting started with Cloudera Data Flow quick and easy. To use it, you must add this ReadyFlow to the Catalog.

Before you begin

- You must have an enabled and healthy Cloudera Data Flow environment.
- You must be assigned the DFCatalogPublisher role granting you access to the Catalog.
- You must be assigned the DFFlowAdmin role for the environment where you want to deploy the flow definition.

About this task



Note: In a new Cloudera Data Flow deployment with an empty Catalog, you can deploy Hello World directly from the **Overview** page.



Get Started with Data Flow


Explore capabilities with ReadyFlows and try out new features.



[Hello World ReadyFlow](#) [Demo Videos](#) [Product Tours](#) [Documentation](#)

Just click the Hello World ReadyFlow option on the **Get Started with Data Flow** card on the **Overview** page. You are forwarded to the **Catalog**, from where you can deploy the ReadyFlow.

If the catalog already contains at least one flow definition, you can add Hello World from the ReadyFlow gallery.

Procedure

1. In the navigation pane, click the  ReadyFlow Gallery item.

 Overview Deployments Catalog **ReadyFlow Gallery** Flow Design Projects Resources

2. On the **ReadyFlow Gallery** pane, filter for and select the Hello World ReadyFlow.



Hello World

Version 4

Gets you started with your first flow deployment without any external dependencies.

Applications

General Connectivity & Protocols

NiFi 2.x

[Add To Catalog](#)[Create New Draft](#)

Tip: Click anywhere on the card to review the ReadyFlow details.

3. Click the Add to Catalog button to add the ReadyFlow to the catalog and make it ready for deployment.
4. Click the Add button to confirm that you want to add Hello World to the catalog.

Results

The Hello World flow definition is added to the catalog and is ready for deployment.

What to do next

You can now proceed with deploying the ReadyFlow.

Deploy the Hello World ReadyFlow using the deployment wizard

Learn about the steps to deploy the Hello World ReadyFlow to get started with Cloudera Data Flow.


About this task

Once you have added the Hello World ReadyFlow into the Catalog, stay in the Catalog and use the Deployment wizard to deploy that flow definition.

Before you begin

- You must have an enabled and healthy Cloudera Data Flow environment.
- You must be assigned the DFCatalogAdmin role granting you access to the Catalog.
- You must have added the Hello World flow definition to the Catalog.
- You must be assigned the DFFlowAdmin role for the environment where you want to deploy the flow definition.

Procedure

1. In the navigation pane, click the  Catalog item.



CLOUDERA Data Flow



Overview



Deployments



Catalog



ReadyFlow Gallery



Flow Design



Projects

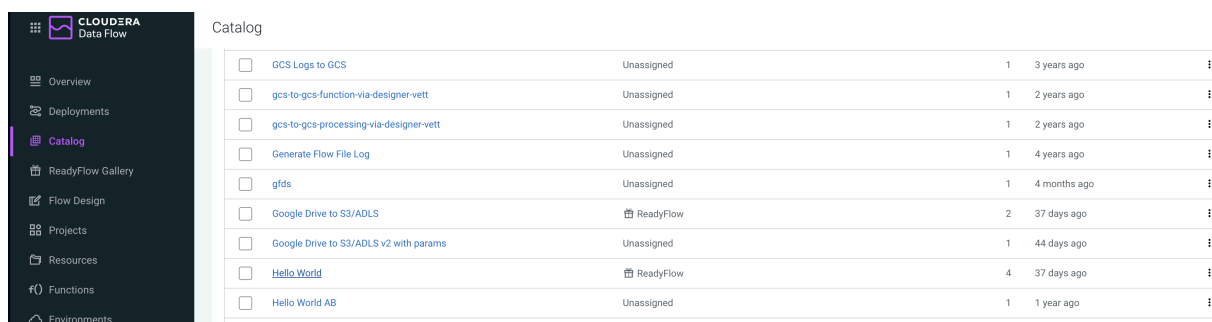


Resources



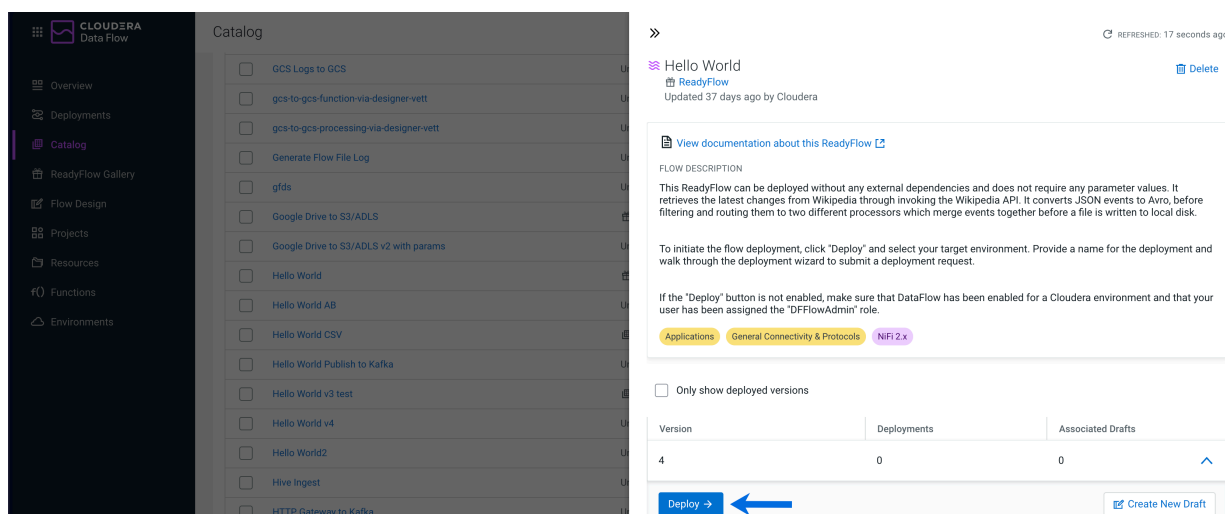
Functions

- To initiate deployment, click the Hello World flow definition on the **Catalog** pane to display the flow definition details.



Flow Definition	Status	Version	Updated	Actions
GCS Logs to GCS	Unassigned	1	3 years ago	
gcs-to-gcs-function-via-designer-vett	Unassigned	1	2 years ago	
gcs-to-gcs-processing-via-designer-vett	Unassigned	1	2 years ago	
Generate Flow File Log	Unassigned	1	4 years ago	
gfds	Unassigned	1	4 months ago	
Google Drive to S3/ADLS	ReadyFlow	2	37 days ago	
Google Drive to S3/ADLS v2 with params	Unassigned	1	44 days ago	
Hello World	ReadyFlow	4	37 days ago	
Hello World AB	Unassigned	1	1 year ago	

- Click the Deploy → button to launch the New Deployment wizard.



» REFRESHED: 17 seconds ago

Hello World
ReadyFlow
Updated 37 days ago by Cloudera

[View documentation about this ReadyFlow](#)

FLOW DESCRIPTION
This ReadyFlow can be deployed without any external dependencies and does not require any parameter values. It retrieves the latest changes from Wikipedia through invoking the Wikipedia API. It converts JSON events to Avro, before filtering and routing them to two different processors which merge events together before a file is written to local disk.

To initiate the flow deployment, click "Deploy" and select your target environment. Provide a name for the deployment and walk through the deployment wizard to submit a deployment request.

If the "Deploy" button is not enabled, make sure that DataFlow has been enabled for a Cloudera environment and that your user has been assigned the "DFFlowAdmin" role.

[Applications](#) [General Connectivity & Protocols](#) [NFI 2.x](#)

☐ Only show deployed versions

Version	Deployments	Associated Drafts
4	0	0

Deploy → [Create New Draft](#)

- Select the Target Workspace where you want to deploy the ReadyFlow.


The New Deployment wizard only displays workspaces that meet the following conditions:

- Cloudera Data Flow is enabled for the environment
- The environment is in healthy state
- You are allowed to access the environment


- Click the Continue → button.

- In the Overview, give your flow deployment a unique name.



Note: Flow Deployment names must be unique. The Deployment Wizard indicates whether a name is valid by displaying  a green check below the Deployment Name text box.



Note: As you have created no projects for this quickstart, you can leave your flow deployment  Unassigned.

- Click the Next → button.

- In the NiFi Configuration step, click the Next → button.

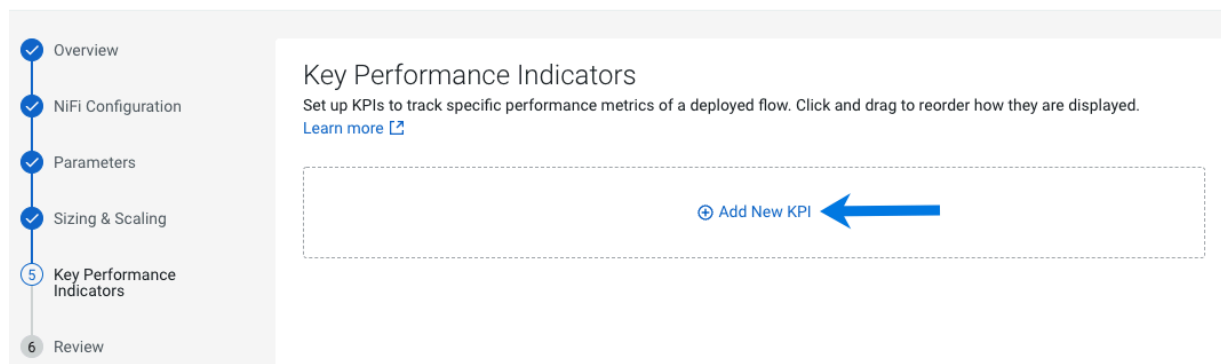
- In the Parameters step, click the Next → button.


- In the Sizing & Scaling step, click the Next → button.

11. To learn about the use of Key Performance Indicators (KPIs), add a KPI that will track how much data the processor is writing to the destination files.

The KPI will raise an alert whenever the value goes below 1MB.

cdf-priv / New Deployment



- a) In the Key Performance Indicators step, click the  Add New KPI button.
- b) Set the following properties:

KPI Scope

Select Processor.

Processor Name

Select Write "Added Content" Events to File.

Metric to Track

Select Bytes Sent

Alerts

Select Trigger alert when metric is less than and set the value to 1 MBytes.

Alert will be triggered when metric is outside the boundary(s) for

Set to 2 Minutes.

Add new KPI



Details

KPI Scope

Processor

Processor Name

Write "Added Content" Events To File

Metric to Track

Bytes Sent

METRIC DESCRIPTION:

Number of bytes sent to an external recipient

Alerts

☐ Trigger alert when metric is greater than

Value

MBytes

☒ Trigger alert when metric is less than

1

MBytes

Alert will be triggered when metric is outside the boundary(s) for

2

Minutes

Cancel

Add

c) Click the Add button to create the KPI.

12. Click the Next button.

13. Review a summary of the information provided. When you are finished, complete flow deployment by clicking the Deploy button.

ResultsOnce you click the Deploy button, you are redirected to the **Active Alerts** tab in the detail view for the deployment where you can track the deployment progress.**Related Information**[Deploying Flow Definitions](#)

[Monitoring and Managing Flow Deployments](#)

[KPI overview](#)

[Working with KPIs](#)