

CDF Flow Designer Overview

Date published: 2021-04-06

Date modified: 2024-01-09

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

Concepts.....	4
Flow design lifecycle.....	5
Flow Design landing page.....	5
Flow Design canvas.....	6
NiFi component documentation embedded in Flow Designer.....	7

Concepts

Learn about basic concepts of flow design and how those concepts relate to their NiFi counterparts.

Controller Service

Controller Services are extension points that provide information for use by other components (such as processors or other controller services). The idea is that, rather than configure this information in every processor that might need it, the controller service provides it for any processor to use as needed.

Draft

A draft is a flow definition in development, and its lifecycle is tied to the workspace it is created in. You can edit, update, and continuously test drafts by starting a Test Session. When your draft is ready to be deployed in production, you can publish it as a flow definition to the CDF catalog.

Workspace

A workspace is where your drafts are stored. You can manage and delete drafts from this view. The workspace is automatically created when DataFlow is enabled for an environment and is tied to the lifecycle of your DataFlow service.

Environment

A logical environment defined with a specific virtual network and region on a customer's cloud provider account. After enabling DataFlow for an environment, its service components run in an environment.

Parameter group

One default parameter group is auto-created when you create a new draft. You can then add parameters to this one group; you cannot create additional ones. When you initiate a test session, a number of parameters are auto-generated in this parameter group. Depending on the configuration options you chose, these are necessary for your NiFi sandbox deployment, and also make it possible to integrate your NiFi instance with CDP components or external data sources.

This default group is automatically bound to each Process Group you create within your draft, making all parameters available to be used in any process group.

Processor

The Processor is the NiFi component that is used to listen for incoming data; pull data from external sources; publish data to external sources; and route, transform, or extract information from FlowFiles.

For more information, see [Apache NiFi User Guide](#).

Process group

When a data flow becomes complex, it often is beneficial to reason about the data flow at a higher, more abstract level. NiFi allows multiple components, such as Processors, to be grouped together into a Process Group.

In CDF Flow Designer, creating a draft automatically creates a process group with the name you specified when creating the draft, acting as the root process group. You can create child process groups as you see fit to organize your data processing logic.

For more information, see [Apache NiFi User Guide](#).

Service

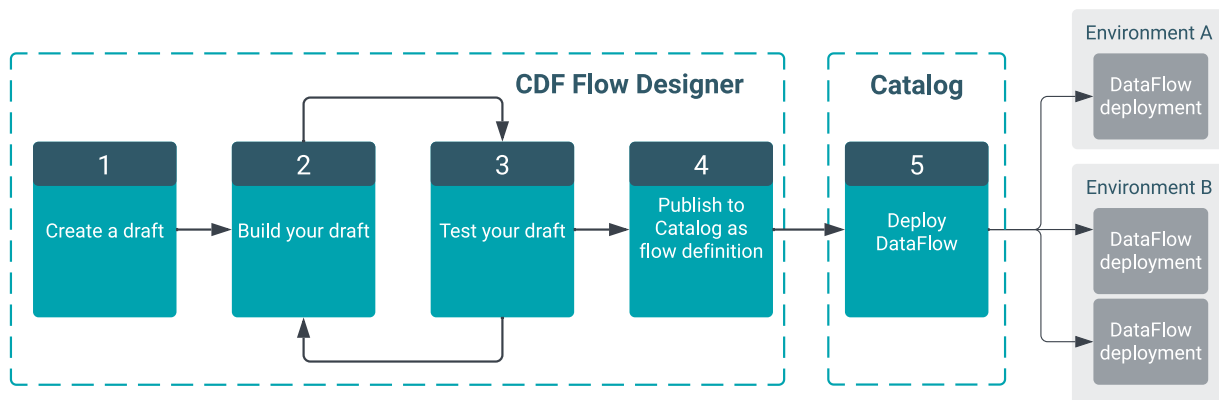
A logical environment defined with a specific virtual network and region on a customer's cloud provider account. CDF service components run in an environment. In the CDF Service, you can provision a workload. This workload allows you to create a number of NiFi Deployments.

Test session

Starting a test session provisions NiFi resources, acting like a development sandbox for a particular draft flow. It allows you to work with live data to validate your data flow logic while updating your draft. You can suspend a test session any time and change the configuration of the NiFi cluster then resume testing with the updated configuration.

Flow design lifecycle

Drafts have a set lifecycle. They are created, built, and tested in CDF Flow Designer, published to the Catalog and finally deployed through the Deployment Wizard or the CLI as flow deployments.



Flows created with CDF Flow Designer have a set lifecycle. You start with creating a draft in Flow Designer. You can start from scratch, but you can also open an existing flow definition from the catalog, or a ReadyFlow from the gallery to use it as a template for your draft. You then build up your flow on the flow design canvas by adding and configuring components, establishing relationships between them, creating services, and so on. Whenever you feel the need, you can start a test session to verify what you have built so far. Starting a test session commissions a NiFi sandbox allowing you to test and simultaneously update your draft. Once you deem your draft to be production ready, you publish it to the catalog as a flow definition. Once published, you can keep updating your draft and publish it as a new version of your flow definition. Should this draft be lost, for example because the workspace holding it got deleted, you can always create a new one from the flow definition you have published the catalog.

Flow Design landing page

The Flow Design landing page provides a read-only view of all drafts you have access to.

The screenshot shows the Cloudera DataFlow Flow Design interface. On the left is a sidebar with navigation options: Dashboard, Catalog, ReadyFlow Gallery, Flow Design (selected), Functions, and Environments. The main area is titled 'Flow Design' and 'All Drafts'. It features a search bar and a 'Create Draft' button. Below is a table listing drafts with columns: Draft Name, Workspace, Test Session, Runtime Version, Base Flow, Last Updated, and Published On. A context menu is open over the 'Hello World CSV updated' draft, showing options: 'Open Design & Edit' and 'View Workspace Manage & Delete Drafts'.

Draft Name ↑	Workspace	Test Session	Runtime Version	Base Flow	Last Updated	Published On
CreateOrders	aws cdf-priv	Active	1.20.0.2.3.8.0-31	CreateOrders v1	10 hours ago by csso_m...	10 hours ago
Hello World CSV	aws cdf-priv	Not Defined	N/A	Hello World CSV v2	4 hours ago by csso_mi...	-
Hello World CSV updated	aws cdf-priv	Inactive	1.20.0.2.3.8.0-29	Hello World CSV v2	a day ago by system	-
Iceberg Test - Drew	aws cdf-priv	Active	1.20.0.2.3.8.0-31	Kafka to Iceberg v1	2 minutes ago by csso...	-
inbound	aws cdf-priv	Failed	1.20.0.2.3.8.0-29	N/A	2 days ago by system	-
Kafka to Iceberg	aws cdf-priv	Active	1.20.0.2.3.8.0-31	Kafka to Iceberg - De... v1	5 minutes ago by csso...	10 minutes ago
Kafka to Iceberg - ReadyFlow	aws cdf-priv	Inactive	1.20.0.2.3.8.0-30	Kafka to Iceberg - Re... v1	10 hours ago by system	10 hours ago
Kafka to CDW-Iceberg	aws cdf-priv	Active	1.20.0.2.3.8.0-31	N/A	5 minutes ago by csso...	-
test-flow-design	turcsanyi1-mow-dev-aws	Not Defined	N/A	N/A	39 minutes ago by csso...	-

From the CDP Public Cloud home page, click Cloudera DataFlow, then click Flow Design. You are redirected to the Flow Design landing page. On the landing page you see all drafts that are both located in an environment where you have the DFDeveloper role, and are either unassigned or assigned to a Project where you have the DFProjectMember role.

- To create a new draft, click Create Draft and follow the instructions.
- To interact with an existing draft, you can
 - click the name of the draft to open it on the Flow Design **Canvas**.
 - click Actions and select Open Data Flow to open it on the Flow Designer **Canvas**, or select View Workspace. In the **All Workspace Drafts** view you can manage (edit, publish as a flow definition, delete) all drafts that are located in that given workspace.

Related Information

[Projects](#)

Flow Design canvas


Learn about getting around and performing basic actions on the Flow Design Canvas.

Components sidebar


You can add components to your draft by dragging them from the Components sidebar and dropping them onto the Canvas.

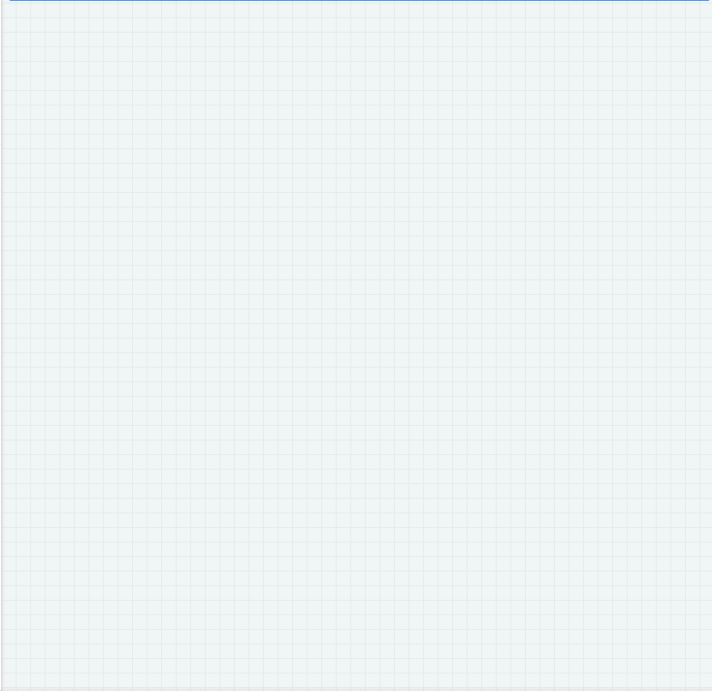
The screenshot shows the Flow Design canvas interface. At the top, it says 'Flow Design / cdf-priv / cdev1' and has a 'Flow Options' button. Below is a components sidebar with icons for various components. A message bar at the top of the canvas says: 'To fully validate your flow and develop with live data, start a test session.' The main area is a large grid for designing the flow.


Configuration pane


To access the Configuration pane, click  Expand in the upper-right corner of the workspace. The Configuration pane is context-sensitive. It always displays the settings belonging to the element selected on the Canvas. When you create a new draft and the Canvas is empty, it defaults to the Process Group that was auto-created with the draft.

Flow Design / cdf-priv / cdev1 Flow Options ▾




 To fully validate your flow and develop with live data, start a test session.





 cdev1
Process Group

[More Details ▾](#)

Settings

Process Group Name

FlowFile Concurrency

Unbounded ▾

Outbound Policy

Stream When Available ▾


Default FlowFile Expiration

Default Back Pressure Object Threshold

Default Back Pressure Data Size Threshold

Comments

Apply Discard Changes


To hide the Configuration pane, click  Hide.



Tip: You can also open the Configuration pane by double-clicking the component that you want to configure on the Canvas.

NiFi component documentation embedded in Flow Designer

The standard NiFi component documentation is directly available in Flow Designer.

To access the component documentation, right-click on a component in the Canvas and select  Documentation.



Note: The embedded component documentation is retrieved from Apache NiFi and may contain information that is not valid in the context of CDF Flow Designer.