Cloudera Data Flow

CDF Flow Designer Overview

Date published: 2021-04-06
Date modified: 2025-09-30

CLOUDERA

Legal Notice

Contents

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 Cloudera Data Flow 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 Cloudera Data Flow is enabled for an environment and is tied to the lifecycle of your Cloudera Data Flow service.

Environment

>A logical environment defined with a specific virtual network and region on a customer's cloud provider account. After enabling Cloudera Data Flow 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 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 Cloudera on cloud 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 Cloudera Data Flow 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. Cloudera Data Flow service components run in an environment. In the Cloudera

Data Flow 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 Flow Designer, published to the Catalog and finally deployed through the Deployment Wizard or the CLI as flow deployments.
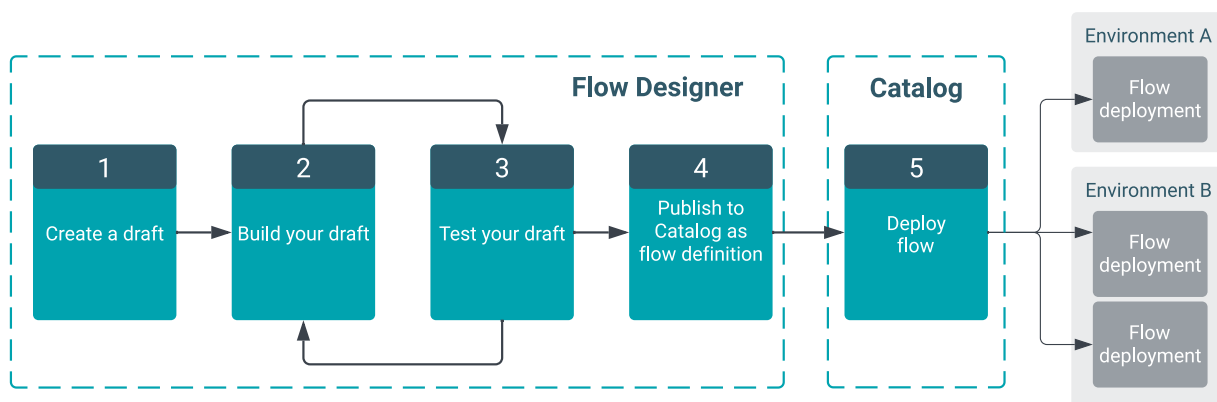
Flows created with Flow Designer have a set lifecycle. You start with creating a draft in Flow Designer. You can start from scratch, but you can also open an existing flow definition from the catalog, or a ReadyFlow from the gallery to use it as a template for your draft. You then build up your flow 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.

From the Cloudera on 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 Designer** Canvas.
    - click ⋮ Actions and select 🖊 Open 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.

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.

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 Apache 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 Cloudera Data Flow Flow Designer.