Cloudera DataFlow

Resources

Date published: 2021-04-06
Date modified: 2024-12-03

CLOUDERA

Legal Notice

Contents

The Resources view

The Resources page is the central place for accessing Cloudera DataFlow resources across environments.

About this task

On the **Resources** you see all workspaces where you have DFFlowUser role. The number of resources is listed by type on each workspace card. To view or manage a resource type in a given workspace, click on the number displayed under the selected resource type.

The following resource types are listed:

- Deployments
- Flow drafts
- Parameter groups
- Custom NAR configurations
- Custom Python configurations

You can filter this view by starting to type the name of a workspace in the search box.

Note: If you do not see the workspaces you expect, check that you have been granted the proper roles for environments.

To manage workspace resources, select a workspace from the list. By clicking the blue number under the resource type, you will be redirected directly to the tab of that resource type on the **Workspace Resoureces** page. By default, the page opens on the **Deployments** tab.

Managing workspace resources

You can access all workspace resources accessible to you through the Resources view. Select a workspace to view, manage, and reassign resources.

Before you begin

You must have the DFAdmin role for the environments where you want to manage resources.

About this task

You can manage the following types of resources:

- Flow deployments
- Flow drafts

- Parameter groups
- Inbound connections
- Custom NAR configurations
- Custom Python configurations

The actions available to you depend on the resource type you select.

Procedure

1. Open Cloudera DataFlow by clicking the DataFlow tile in the Cloudera sidebar.
2. Select Resources.
3. Select the workspace where you want to manage resources.

   Tip: Start typing the resource name in the search bar to filter for a workspace.

4. On the **Workspace Resources** page select the tab with the type of resource you want to manage.

Deployments

On the Deployments tab you can view and manage all deployments within a workspace.

Viewing data flow in NiFi

You can go to the NiFi cluster where your flow is deployed and view or edit the data flow.

About this task

When you access the NiFi cluster, the ability to view or edit the flow is based on your Cloudera DataFlow authorizations. The DFFlowUser role has read-only privileges. The DFFlowAdmin role has full privileges.

Before you begin
You must have deployed a data flow in Cloudera DataFlow.

Procedure

1. Select the **Deployment** that you want to manage.
2. Click ⋮ Options View in NiFi .

   The UI for the NiFi cluster where your flow is deployed opens.
3. View your data flow or edit it based on your NiFi privileges.

   If you edit the flow in NiFi and want the changes to exist in a new deployment, perform the following steps:

   a) Download the flow as a flow definition.

      For more information, see *Downloading a flow definition from NiFi*.
   b) Import the flow definition (as a new flow definition or as a new version of an existing flow definition).

      For more information, see *Importing a flow definition to Cloudera DataFlow*.
   c) Deploy the flow definition.

      For more information, see *Deploy a flow*.

Starting a flow
You can start a stopped flow a Cloudera DataFlow deployment.

About this task

Starting a flow deployment starts all processors of a Cloudera DataFlow deployment.

Before you begin

- You must have a stopped flow deployment in Cloudera DataFlow.
- You must have DFFlowAdmin permission.

Procedure

1. Select the **Deployment** that you want to manage.

2. Click ⋮ Options Start flow .

   The Start [Deployment Name] pop-up appears.

3. Confirm your choice by clicking Start Flow.

Stopping a flow

Stopping the flow of a Cloudera DataFlow deployment temporarily pauses the NiFi flow.

About this task

Stopping a flow results in the following:

- All processors are stopped and no data processing happens within the NiFi flow.
- KPI alerts are stopped. Your KPI alerts are activated again when the flow is restarted.
- Any active KPI alerts are resolved.
- All underlying cloud resources remain allocated for the Cloudera DataFlow deployment.
- You can modify deployment configuration while the flow is stopped.
- Stopped flows are still billable however if auto-scaling is enabled for the flow, a certain amount of cost reduction may occur.

Before you begin

You must have deployed a flow definition in Cloudera DataFlow.

Procedure

1. Select the **Deployment** that you want to manage.

2. Click ⋮ Options Stop flow .

   The Stop [Deployment Name] pop-up appears.

3. Click Stop Flow to stop the flow deployment.

Changing flow version

Learn how to change the flow definition version of a running flow deployment. Using the 'change flow version' capability eliminates the need to terminate and re-create deployments when you want to deploy a new version of your flow definition.

Before you begin

- You must have DFFlowAdmin permission.
- There is at least one more version for the same flow definition present in the catalog.
- The state of the flow deployment is either Good Health or Stopped.
- You have read the applicable restrictions and version change strategies.

About this task

Important: As every flow is unique, you need to test version change before performing it in a production environment

Restrictions

The following restrictions apply to flow version changes:

- Changing inbound connections is not supported.
- Changing custom resource (custom NARs and custom Python resources) configurations is not supported.
- While you can add, change or remove assets when moving to a new version, you cannot introduce assets (text files, binaries, JARs, or similar) if the currently deployed version does not have any.
- Components where state or provenance and other repositories must be kept between flow versions must keep their flow JSON ids. The id changes if you move the component to a different process group or if you delete and then re-add the component to the same process group. NiFi Identifies components by these ids. If you move a component to a different process group between versions, its id changes and NiFi perceives it as a new component. This results in the original component being deleted during flow version change together with its state and a new, identical processor being created in a different process group. In an extreme case, you could change to an identical flow version with just the component ids changed and it would result in the deletion of the entire NiFi flow and the recreation of an identical one, with all history and data lost.
- Remapping Parameter Group and Parameter Context assignments is not supported as the original assignment is not removed. For example, you have Process Group 1 (PG1) with Parameter Context 1 (PC1) and Process Group 2 (PG2) with Parameter Context 2 (PC2) assigned. If you initiate a flow version change where parameter contexts are flipped, resulting in a PG1-PC2 and PG2-PC1 assignment, NiFi will not re-map the PG to PC assignments.

Version change strategies

Depending on the type of your flow, you may select the flow version change strategy most appropriate to you.

Stop & Process Data

This strategy prioritizes data consistency by stopping source processors and waiting until data is processed before stopping all other components. Once all components have stopped, the flow version is changed and components are started.

Use this strategy when your sources are durable and can handle your source processor being stopped. This generally works well when your source processors are pulling data from sources like Kafka or other messaging queues, databases or file systems.

Should the queued data not be processed within the set time, version change will fail and you can retry the operation with a bigger timeout or you can cancel

Only Restart Affected Components

This strategy prioritizes uptime by identifying and stopping only components that have changed while keeping all others running, replacing and then starting affected components.

Use this strategy when you want to prioritize uptime of unchanged components or you have made only minor processor configuration changes.

This works well for deployments with inbound connections and will keep your source processors running if they have not changed compared to the previous version.

Stop & Empty Queues

This strategy forces a version change by stopping all components, emptying all queues, changing flow version, and then starting all components.

Use this strategy only when you want to force a flow version change without keeping any processors running or attempting to process queued data.

All processors will be stopped and all queues will be emptied as part of this strategy.

Procedure

1. Select the **Deployment** that you want to manage.

2. Click ⋮ Options ⊕ Change Flow Version .

   The **Change Flow Version** modal window opens. It shows the list of available flow versions. The current version is grayed out.

   > **Tip:** You can filter versions by name or by tags.

3. Select the flow version you want to change to and click Continue → .

4. Review a summary of the configuration changes caused by the version change and make any necessary edits from the left pane.

5. Select a flow version change strategy.

   The available options are:

   • Stop & Process Data - If you select this strategy, you can set the maximum wait time for data to be processed and queues to be emptied before the request timed out. The default value is 15 minutes.

   > **Note:** If you initiate flow version change with this strategy on a stopped flow and there is no data in the queues, flow version change instantly happens regardless of the wait time you configured. However if there is data left in the queues of the stopped flow, the change operation instantly fails.

   • Only Restart Affected Components

   > **Note:** If you initiate flow version change with this strategy and there are queued flow files on any connection that is going to be removed in the new flow version, the version change will fail.

   • Stop & Empty Queues - If you select this strategy, you must accept potential data loss by selecting I understand and choose to proceed with the configuration as is.

6. Click ☁ Change Flow Version.

Results

After you click Deploy, you are redirected to the **Alerts** tab in the **Flow Details** view where you can track how the version change progresses.

Downloading NiFi application log

You can download the NiFi application log from the CDF Deployment Manager to use it for troubleshooting.

About this task

This feature allows you to download the NiFi application log that is currently being written. As the log file is rotated and the old file is archived once the file size reaches 10 MB, this is the theoretical maximum you can download using this method. For information on downloading archived log files, see *Diagnostic bundle collection*.

Before you begin

You need DFFlowAdmin permission to perform this action.

Procedure

1. Select the **Deployment** that you want to manage.

2. Click ⋮ Options Download NiFi Log .

   The current NiFi application log is downloaded to your computer in tar.gz format.

Suspending a deployment

Suspending a Cloudera DataFlow deployment terminates cloud resources belonging to a NiFi flow, while maintaining flow persistence.

About this task

Suspending a Cloudera DataFlow deployment results in the following:

- The NiFi flow stops processing data and all underlying cloud resources are terminated. Any unprocessed data in the flow is stored in memory and its processing resumes when you resume the deployment.
- Flow persistence is maintained while a deployment is suspended.
- You cannot modify deployment configuration while the deployment is suspended.
- Suspended deployments are not billable, resulting in reduced costs.

Before you begin

You must have deployed a flow definition in Cloudera DataFlow.

Procedure

1. Select the **Deployment** that you want to manage.

2. Click ⋮ Options Suspend Deployment .

   The Suspend [Deployment Name] modal opens.

3. Select the Finish data processing option and set a maximum wait time in minutes for data to be processed and queues to be emptied before the request times out.

   This option stops source processors first and waits for queued data to be processed before the flow is suspended. Set a wait time between 5 and 60 minutes using the slider.

4. Click Suspend to suspend the Cloudera DataFlow deployment.

Resuming a deployment

You can resume a suspended Cloudera DataFlow deployment.

About this task

Resuming a Cloudera DataFlow deployment reallocates the underlying cloud resources and returns a deployment to the state it was in before being suspended.

Before you begin

You must have a suspended flow deployment in Cloudera DataFlow.

Procedure

1. Select the **Deployment** that you want to manage.

2. Click ⋮ Options Resume Deployment .

   The Resume [Deployment Name] modal opens.

3. Click Resume Deployment to resume the flow deployment, reallocating cloud resources.

Export deployment configuration

You can export a deployment configuration to create additional deployments with a similar configuration in the same or a different environment.

About this task

- Exported configurations may be edited, and you can also modify them after the importing step during flow deployment.
- One deployment can have only one exported configuration. Performing a new export overwrites the existing one.
- Exported deployment configurations are available for every user who can start a new deployment in a given environment, even if the exported deployment was originally created under a specific Project.

Procedure

1. Select the **Deployment** that you want to manage.
2. Click ⋮ Options ⬇ Export Configuration .

   The Export Deployment Configuration modal opens.
3. You can optionally add comments to the exported cofiguration.
4. Confirm your choice by clicking Export in the modal.

Results

The configuration is exported to the {LOG location}/cdf-deployment-backup directory. {LOG location} is configured during the creation of the associated Cloudera Environment. If you want to reuse the exported configuration in a different environment, you can either configure that to use the same {LOG location}, or you can copy the exported .tar.gz and JSON files to the {LOG location}/cdf-deployment-backup directory of the target environment.

What to do next

You can reuse the exported configuration during deployment of the same flow definition to recreate a flow with similar configuration.

Terminating a deployment

You can terminate a deployment to remove it from Cloudera DataFlow.

About this task

If you terminate a deployment, you delete the associated NiFi resources and your flow no longer remains active. The associated flow definition remains in the catalog and is available to be deployed again in a new deployment.

Before you begin

You must have deployed a flow definition in Cloudera DataFlow.

Procedure

1. Select the **Deployment** that you want to manage.
2. Click ⋮ Options Terminate Deployment .

   The Terminate [Deployment Name] modal opens.
3. If you select the Delete assigned endpoint hostname option. If you do not select this option, you can reassign existing, unassigned endpoints during flow deployment.

4. Enter the name of the deployment to confirm and click Terminate.

> **Tip:** You can also click the ⬚ copy to clipboard icon to copy the deployment name, paste the name of the deployment, and then click Terminate.

Flow drafts

On the Flow drafts tab you can manage the life cycle of flow drafts and associated test sessions.

Opening a draft for editing

You can open a selected flow draft for editing in the Flow Designer.

Before you begin

• Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Flow draft** that you want to manage.

2. Click ⋮ Options Open .

The flow draft opens in the Flow Designer.

Starting, stopping, and restarting a test session

Learn how you can start, stop, or restart a test session for a flow draft.

Before you begin

• Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Flow draft** that you want to manage.

2. Select one of the following options

• To start a test session, click ⋮ Options Start Test Session . For more information on configuring a test session, see Test sessions.

> **Note:** You cannot start a test session if there is already an **Active** or **Failed** test session.

• To stop a test session, click Stop Test Session, then click End to confirm your choice.

• To restart an **Inactive** test session, click ⋮ Options Restart Test Session

Publishing a flow draft

Learn about publishing a draft to the Catalog as a flow definition.

Before you begin

• Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Flow draft** that you want to manage.
2. Click ⋮ Options Publish .
3. Provide a Flow Name and optional Flow Description, Custom Tag, and Version Description.
4. Click Publish.

Reassigning a flow draft to a different project

Learn how to reassign a flow draft to another project.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Flow draft** that you want to reassign.
2. Click Reassign.

   The **Reassign Resource** modal opens.
3. Select a Project and click Reassign.

Deleting a flow draft

Learn about deleting flow drafts.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Flow draft** that you want to delete.
2. Click ⋮ Options 🗑 .
3. Click Delete to confirm your choice.

Parameter groups

You can manage shared parameter groups in the Resources view of Cloudera DataFlow.

Procedure

Parameter Group

- To view the list of flow deployments and flow drafts using this parameter group, select the **Used By** tab.

  It is important to be aware of the flow deployments and flow drafts that are going to be affected by any change you make to a parameter group.
- To view or modify the Parameter Group Name and Description, select the **Details** tab.

Modifying parameters in a parameter group

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Parameter Group** where you want to modify parameters.
2. Click ⋮ Options Open .
3. Click on the parameter card of the parameter you want to modify.

   In the parameter details pane modify the parameter Value and Description as necessary.
4. Click ⬆ Apply Changes.

Deleting a parameter group

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Parameter Group** that you want to delete.
2. Click ⋮ Options Open .

3. Click on the parameter card that you want to delete.

In the parameter details pane click 🗑 to delete the parameter from the group.

If the parameter group is not used by any flow draft, the **Delete [parameter group name]** modal opens. If the parameter group is in use by a flow draft, it cannot be deleted.

4. Click Delete to confirm your choice.

5. Click ⬆ Apply Changes.

Importing a parameter group to a flow draft

Procedure

1. Select the **Parameter Group** that you want to import to a flow draft.

2. Click ⋮ Options 🖉 Import to Flow Draft .

3. In the **Import to Flow Draft** modal select the Flow Draft where you want to import the parameter group.

4. Click Import.

Duplicating a parameter group

Learn how to duplicate parameter groups. Duplicating a parameter group allows you to reassign the newly created group to another project, or to customize it according to your needs without touching the original group.

About this task

Note: Duplicating a parameter group does not duplicate any assets (file-type parameters) associated with it. You need to reupload those to the newly created group.

Procedure

1. Select the **Parameter Group** that you want to import to duplicate.

2. Click ⋮ Options ⊞ Duplicate .

3. In the **Duplicate Parameter Group** modal provide a Parameter Group Name and an optional Description.

4. Click Duplicate.

Reassigning a parameter group to a different project

Learn how to reassign a parameter group to another project.

Before you begin

• Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

About this task

You cannot reassign a parameter group that is used by flow drafts. You have to remove the group from all flow drafts before you can reassign it to a different project.

Note: When you reassign a parameter group to another project, assets associated with it are not moved with it. You need to reupload those to the new project.

Procedure

1. Select the **Parameter Group** that you want to reassign.

   Note: You can reassign multiple items simultaneously by selecting the checkboxes in front of them.

2. Click Reassign.

   If the parameter group is not used by any flow draft, the **Reassign Resource** modal opens.

3. Select a Project and click Reassign.

4. Click ⬆ Apply Changes.

Inbound connections

On the Inbound Connections tab you can reassign and renew the certifacates of inbound connections.

Renewing the certificate for an inbound connection endpoint

If you need to replace an X.509 certificate for an inbound connection endpoint before it expires, you can do so manually.

Before you begin
You need DFFlowAdmin privilege to perform this action.

Procedure

1. Select the **Inbound Connection** that you want to manage.
2. Click Renew.

   • To renew the server certificate, select NiFi Inbound SSL Context Service.

   Note: Each server certificate is limited to five renewals in a 7 day sliding window.

   • To renew the client certificate, select Client SSL Context.
   • If you leave Revoke previously issued client certificates unchecked, existing client certificates remain valid and existing clients can continue to connect to your deployment using it. By selecting the Revoke previously issued client certificates option, you invalidate all existing certificates and you will need to add the new certificate to existing clients so that they can keep connecting to your Cloudera DataFlow deployment.

3. Click Renew & Restart.
   The UI switches to the **KPIs and Alerts** pane where you can monitor as your deployment restarts and the new certificate or certificates become available.

What to do next
If you have renewed the NiFi Inbound SSL Context Service:

> You have to take no further action.

If you have renewed the Client SSL Context:

> After your Cloudera DataFlow deployment has restarted, you switch to the NiFi Configuration pane to download the Client Certificate and the Client Private Key. You can then add these to your client.

Reassigning an inbound connection endpoint to a different project

Learn how to reassign an inbound connection to another project.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

About this task

You cannot reassign an inbound connection that is currently used by a deployment. You have to terminate the deployment using it making sure that the Delete assigned endpoint hostname option is not selected before you can reassign it to a different project.

Procedure

1. Select the **Inbound Connection** that you want to reassign.

   Note: You can reassign multiple items simultaneously by selecting the checkboxes in front of them.

2. Click Reassign.

   If the inbound connection is not used by any deployment, the **Reassign Resource** modal opens.

3. Select a Project and click Reassign.

4. Click ⬆ Apply Changes.

Custom NAR configs

On the Custom NAR Configs tab you can reassign and validate Custom NAR configs.

Validating a custom NAR config

You can validate custom NAR files and their locations in the Resources view.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Custom NAR** that you want to validate.

2. Click ⋮ Options ✓ Validate .

   Cloudera DataFlow displays a message about the validity of the NAR config and its storage location.

Reassigning a custom NAR config to a different project

Learn how to reassign a custom NAR config to another project.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

About this task

You cannot reassign an inbound connection that is currently used by a deployment. You have to teminate the deployment using it making sure that the Delete assigned endpoint hostname option is not selected before you can reassign it to a different project.

Procedure

1. Select the **Custom NAR** that you want to reassign.

    Note: You can reassign multiple items simultaneously by selecting the checkboxes in front of them.

2. Click Reassign.

    If the custom NAR config is not used by any deployment, the **Reassign Resource** modal opens.

3. Select a Project and click Reassign.

4. Click ⬆ Apply Changes.

Custom Python configs

On the Custom Python Configs tab you can reassign and validate Custom Python configs.

Validating a custom Python config

You can validate custom Python configs and their storage locations in the Resources view.

Before you begin

• Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Custom Python Config** that you want to validate.

2. Click ⋮ Options ✓ Validate .

    Cloudera DataFlow displays a message about the validity of the Python config and its storage location.

Reassigning a custom Python config to a different project

Learn how to reassign a custom Python config to another project.

Before you begin

• Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see Cloudera DataFlow Authorization.

Procedure

1. Select the **Custom Python Config** that you want to reassign.

    Note: You can reassign multiple items simultaneously by selecting the checkboxes in front of them.

2. Click Reassign.

    If the custom Python config is not used by any deployment, the **Reassign Resource** modal opens.

3. Select a Project and click Reassign.

4. Click ⬆ Apply Changes.